Population value decomposition, a framework for the analysis of image populations

C.M. Crainiceanu B.S. Caffo S. Luo V. Zipunnikov N.M. Punjabi

Abstract

Images, often stored in multidimensional arrays are fast becoming ubiquitous in medical and public health research. Analyzing populations of images is a statistical problem that raises a host of daunting challenges. The most severe challenge is that data sets incorporating images recorded for hundreds or thousands of subjects at multiple visits are massive. We introduce the population value decomposition (PVD), a general method for simultaneous dimensionality reduction of large populations of massive images. We show how PVD can seamlessly be incorporated into statistical modeling and lead to a new, transparent and fast inferential framework. Our methodology was motivated by and applied to the Sleep Heart Health Study, the largest community-based cohort study of sleep containing more than 85 billion observations on thousands of subjects at two visits.

Keywords: signal extraction, population value decomposition, EEG

1 Introduction

We start by considering the following thought experiment using data displayed in Figure 1. Inspect the plot for a minute and try to remember it as well as possible; ignore the meaning of the data and try to answer the question: "How many features (patterns) from this plot do you remember?" Now, consider the case when you are flipping through thousands of similar images and try to answer the slightly modified

question: "How many common features from all these plots do you remember?" Regardless of who is answering either question, the answer for this data set seems to be invariably between 3 and 25.

To mathematically represent this experiment we introduce the population value decomposition (PVD) of a sample of matrices. Here we focus on providing the intuition, while the formal definition is introduced in Section 3. Consider a sample $\boldsymbol{Y}_i$, $i = 1, \ldots, n$, of matrices of size $F \times T$, where F or T or both are very large. Suppose that the following approximate decomposition holds

$$\boldsymbol{Y}_i \simeq \boldsymbol{P}\boldsymbol{V}_i\boldsymbol{D} \tag{1}$$

where $\boldsymbol{P}$ and $\boldsymbol{D}$ are population specific matrices of size $F \times A$ and $B \times T$, respectively. If A or B are much smaller than F and T then equation (1) provides a useful representation of a sample of images. Indeed, the "subject-level features" of the image are coded in the low dimensional matrix $\boldsymbol{V}_i$, while the "population frame of reference" is coded in the matrices $\boldsymbol{P}$ and $\boldsymbol{D}$. Important differences between PVD and the singular value decomposition (SVD) are: 1) PVD applies to a sample of images not just one image; 2) the matrices $\boldsymbol{P}$ and $\boldsymbol{D}$ are population-, not subject-, specific; 3) the matrix $\boldsymbol{V}_i$ is not necessarily diagonal.

With this new perspective we can revisit Figure 1 to provide a reasonable explanation for how our vision and memory might work. First, the image can be decomposed using a partition of frequencies and time in several subintervals. A checkerboard-like partition of the image is then obtained by building the 2-D partitions from the 1-D partitions. The size of the partitions is then mentally adjusted to match the observed complexity in the image. When decomposing a sample of images the thought process is similar, except that some adjustments will be made on the fly to ensure maxi-

mum encoding of information with minimum amount of memory. Some smoothing across subjects further improves efficiency by taking advantage of observed patterns across subjects. A mathematical representation of this process would be to consider subject-specific matrices, $\boldsymbol{P}$ and $\boldsymbol{D}$, with columns and rows corresponding to the 1-D partitions. The matrix $\boldsymbol{V}_i$ is then constructed by taking the average of the image in the induced 2-D subpartition. Our methods transfer this empirical reasoning into a statistical framework. This process is crucial because

1. Reducing massive images to a manageable set of coefficients that are comparable across subjects is of primary importance. Note that Figure 1 displays $57,000$ observations, only a fraction of the total of $228,160$ observations of the original, uncut, image. The matrix $\boldsymbol{V}_i$ typically contains less than 100 entries.

2. Statistical inference on samples of images is typically hard. For example, the Sleep Heart Health Study (SHHS) described in Section 2 contains one image for each of two visits for more than $3,000$ subjects. The total number of observations used in the analysis presented in Section 5 is more than $450,000,000$. In contrast, replacing $\boldsymbol{Y}_i$ by $\boldsymbol{V}_i$ reduces the data set to $600,000$ observations.

3. Obtaining the coefficient matrix $\boldsymbol{V}_i$ is easy once $\boldsymbol{P}$ and $\boldsymbol{D}$ are known. Using the entries of $\boldsymbol{V}_i$ as predictors in a regression context is then straightforward; this strategy was employed by [1] for predicting Alzheimer risk using functional Magnetic Resonance Imaging (fMRI).

4. Modeling of the coefficients $\boldsymbol{V}_i$ can replace modeling of the images $\boldsymbol{Y}_i$. In Section 3 we show that the Karhunen-Loève (KL) decomposition [16, 18] of a sample of images can be approximated by using a computationally tractable algorithm based on the coefficients $\boldsymbol{V}_i$. This avoids the intractable problem of calculating and diagonalizing very large covariance operators.

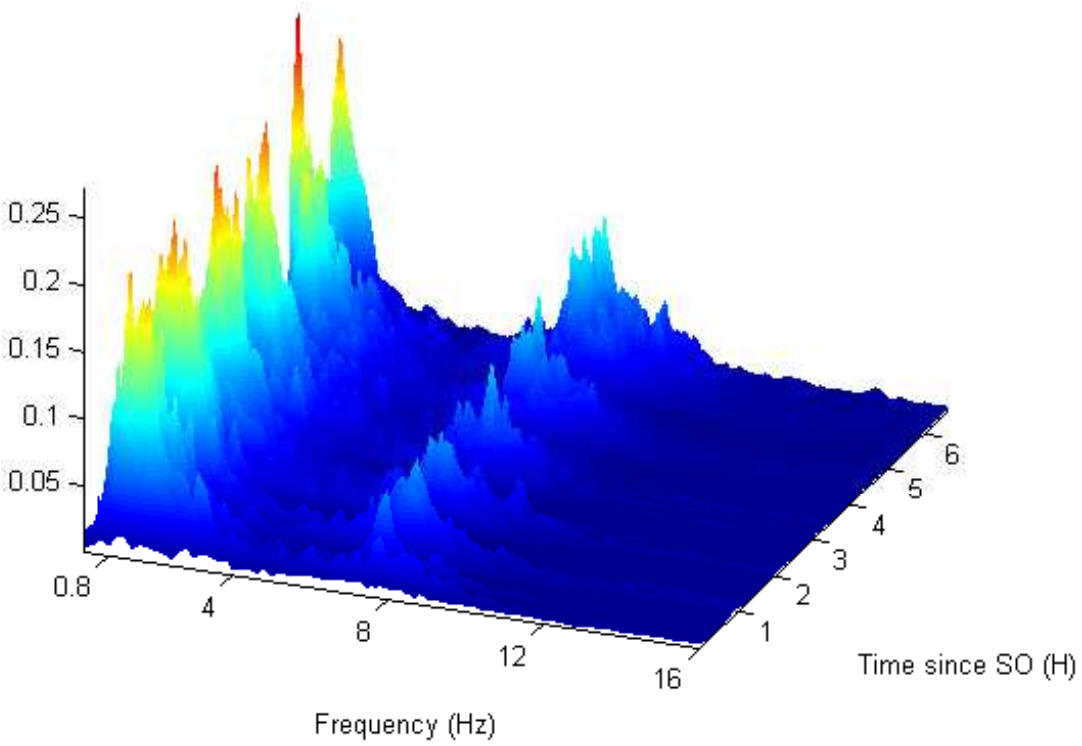

Figure 1: Frequency by time percent power for the sleep-EEG data for one subject. The X-axis is time in hours since sleep onset, where each row corresponds to a 30-second interval. The Y-axis is the frequency from 0.2Hz to 16Hz. The other frequencies were not shown because they are "quiet", that is the proportion of power in those frequencies is very small.

The paper is organized as follows. In Section 2 we introduce the SHHS and the associated methodological challenges. In Section 3 we introduce the PVD and its application to the analysis of samples of images. Section 4 provides simulations while Section 5 provides extensive results for the analysis of the SHHS data set. Some of the unresolved methodological and applied problems are discussed in Section 6.

2 The Case Study

The SHHS is a landmark study of sleep and its impacts on health outcomes. A detailed description of the SHHS can be found in [10, 11, 20]. The SHHS is a multi-center cohort study that utilized the resources of existing epidemiologic cohorts, and conducted further data collection, including measurements of sleep and breathing. Between 1995 and 1997, in-home polysomnogram (PSG) data were collected from a sample of $6,441$ participants. A PSG is a quasi-continuous multi-channel recording of physiological signals acquired during sleep that include two surface electroencephalograms (EEG). After the baseline visit, a second SHHS follow-up visit was undertaken between 1999 and 2003 and included a repeat PSG. A total of $4,361$ participants completed a repeat in-home PSG. The main goals of the SHHS are to quantify the natural variability of complex measurements of sleep in a large community cohort, identify potential biomarkers of cardiovascular and respiratory disease, and study the association between these biomarkers and various health outcomes including sleep apnea, cardiovascular disease and mortality.

The focus on sleep EEG is due to the expectation that spectral analysis of electroneural data will provide a set of reliable, reproducible and easy to calculate biomarkers. Current quantification of sleep in most research settings is based on a visually-based counting process that attempts to identify brief fluctuations in the EEG (i.e., arousals) and classify time-varying electrical phenomena into discrete sleep stages. While metrics of sleep based on visual scoring have been shown to have clinically meaningful associations, they are subject to several limitations. First, interpretation of scoring criteria and lack of experience can increase error variance in the derived measures of sleep. For example, even with the most rigorous training and certification requirements, technicians in the large multi-center Sleep Heart Health Study

were noted to have an intra-class correlation coefficient of 0.54 for scoring arousals [26]. Second, there is a paucity of definitions for classifying EEG patterns in disease states as the criteria were developed primarily for normal sleep. Third, many of the criteria do not have a biological basis. For example, an amplitude criterion of 75 μV is used for the identification of slow waves [22] and a shift in EEG frequency for at least 3-seconds is required for identifying an arousal. Neither of these criteria is evidence based. Fourth, visually-scored data is described with summary statistics of different sleep stages resulting in complete loss of temporal information. Finally, visual assessment of overt changes in the EEG provides a limited view of sleep neurobiology. In the setting of SDB, a disorder characterized by repetitive arousals, visual characterization of sleep structure cannot capture common EEG transients. Thus, it is not surprising that previous studies have found weak correlations between conventional sleep stage distributions, arousal frequency, and clinical symptoms [4, 12, 17, 19]. Power spectral analysis provides an alternate and automatic means for the studying of the dynamics of the sleep EEG often revealing global trends of EEG power density during the night. While methods for quantitative analysis of the EEG have been employed in sleep medicine, its use has focused on characterizing EEG activity during sleep in disease states or in experimental conditions. There is a limited number of studies that have undertaken analyses of the EEG for the entire night to delineate the role of disturbed sleep structure in cognitive performance and daytime alertness. Such studies are often based on samples of less than 50 subjects and are thus not generalizable to the general population. Finally, there are only isolated reports using quantitative techniques to characterize EEG during sleep as a function of age and gender with the largest study consisting of only 100 subjects.

To address these problems we focus on the statistical modeling of the time-varying spectral representation of the subject-specific raw EEG signal. The main components

of this strategy are described below.

C1. RAW SIGNAL $\mapsto$ IMAGE (FFT)

C2. FREQUENCY $\times$ TIME IMAGE $\mapsto$ IMAGE CHARACTERISTICS (PVD)

C3. ANALYZE IMAGE CHARACTERISTICS (FPCA and MFPCA)

Component C1. is a well established data transformation and compression technique at the subject level. Even though we make no methodological contributions in C1., its presentation is necessary to understand the application. The technical details for C1. are provided in Sections 2.1 and 2.2. Component C2. is our main contribution and is a second level of compression at the population level. This is an essential component when images are massive, but could be avoided when images are small. Methods for C2. are presented in Section 3. Component C3. is our second contribution that generalizes Multilevel Functional Principal Component Analysis (MFPCA) [11] to multilevel samples of images. Technical details for C3. are presented in Sections 3.2.1 and 3.2.2.

2.1 Fourier transformations and local spectra

In the SHHS EEG is sampled at frequency 125Hz (125 observations per second) and an 8 hour sleep interval will contain $U = 125\text{Hz} \times 60'' \times 60' \times 8\text{h} = 3,600,000$ observations. A standard data reduction step for EEG is to partition the entire time series into adjacent five-second intervals. The five-second intervals are further aggregated into adjacent groups of 6 intervals for a total time of 30 seconds. These adjacent 30 second intervals are called epochs. Thus, for an 8 hour sleep interval the number of five-second intervals is $U/625 = 5,760$ and the number of epochs is $T = U/(625 \times 6) = 960$. In general, U and T are subject and visit-specific because the length of sleep is.

Consider now the partitioned data and denote by $x_{th}(n)$ the nth observation of the raw EEG signal, $n = 1, \ldots, N = 625$, in the hth five-second interval, $h = 1, \ldots, H = 6$, of the tth 30 second epoch, $t = 1, \ldots, T$. In each five-second window data are first centered around their mean. We continue to denote the centered data by $x_{th}(n)$. A Hann weighting window is then applied to the data, which replaces the $x_{th}(n)$ with $w(n)x_{th}(n)$, where $w(n) = 0.5 - 0.5\cos\{2\pi n/(N-1)\}$. To these data we apply a Fourier transform and obtain $X_{th}(k) = \sum_{n=0}^{N-1} w(n)x_{th}(n)e^{-2\pi kn\sqrt{-1}/N}$ for $k = 0, \ldots, N-1$. Here $X_{th}(k)$ are the Fourier coefficients corresponding to the hth five-second interval of the tth epoch and frequency $f = k/5$. For each each frequency, $f = k/5$, and 30 second epoch, t, we calculate $P(f,t) = \frac{1}{H}\sum_{h=1}^{H} |X_{th}(k)|^2$ the average over the $H = 6$ five-second intervals of the square of the Fourier coefficients. More precisely, $P(f,t) = \frac{1}{H}\sum_{h=1}^{H} |\sum_{n=0}^{N-1} w(n)x_{th}(n)e^{-2\pi kn\sqrt{-1}/N}|^2$. Total power in a spectral window could be calculated as $\mathrm{PS}_b(t) = \sum_{f\in D_b} P(f,t)$, where D_b denotes the spectral window (collection of frequencies) indexed by b.

In this paper we focus on $P(f,t)$ and treat it as a bivariate function of frequency f (expressed in Hz) and time t (expressed in epochs). The power in a spectral window, $\mathrm{PS}_b(t)$, was analyzed in [9] and [11]. Here we concentrate on methods that generalize the spirit of the methods in [11], while focusing on solutions to the much more ambitious problem of population level analysis of images. Before describing our methods we provide more insights into the interpretation of the frequency-time analysis.

2.2 Insights into the Discrete Fourier Transform

First, note that the inverse Fourier transform is $w(n)x_{th}(n) = \frac{1}{N}\sum_{k=0}^{N-1} X_{th}(k)e^{2\pi kn\sqrt{-1}/N}$ and the Fourier coefficients are the projections of the data on the orthonormal basis $e^{2\pi kn\sqrt{-1}/N}$, $k = 0, \ldots, N-1$. Thus, a larger, in absolute value, $X_{th}(k)$ corresponds to

a larger contribution of the frequency $k/5$ to explaining the raw signal. Parseval's theorem provides the following equality $\sum_{n=0}^{N-1} |w(n)x_{th}(n)|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |X_{th}(k)|^2$. The left hand side of the equation is the total observed variance of the raw signal and the right hand side provides an ANOVA-like decomposition of the variance as a sum of $|X_{th}(k)|^2$. This is the reason why $|X_{th}(k)|^2$ is interpreted as the part of variability explained by frequency $f = k/5$. In signal processing $|X_{th}(k)|^2$ is called the power of the signal in frequency $f = k/5$.

We finish our pre-processing of the data by normalizing the observed power as $Y(f,t) = P(f,t)/\sum_f P(f,t)$, which is the "proportion" of observed variability of the EEG signal that is attributable to frequency f in epoch t. In practice, for surface EEG frequencies above 32Hz have a negligible contribution to the total power and we define $Y(f,t) = P(f,t)/\sum_{f\leq 32} P(f,t)$. We will call $Y(f,t)$ the normalized power. The true signal measured by $Y(f,t)$ will be called the frequency by time image of the EEG time series.

Figure 1 provides the frequency by time plot of $Y(f,t)$ for one subject who slept for more than 6 hours. The X-axis is the frequency from 0.2Hz to 16Hz. The other frequencies were not shown because they are "quiet", that is, the proportion of power in those frequencies is very small. The Y-axis is time in hours since sleep onset, where each row corresponds to a 30-second interval. Note that a large proportion of the observed variability is in the low frequency range, say between $[0.8 - 4.0\text{Hz}]$. This range is known as the δ-power band and is traditionally analyzed in sleep research by averaging the frequency values across all frequencies in the range. Another interesting range of frequencies is roughly between 5 and 10Hz with the proportion of power quickly converging to zero beyond 12-14Hz. The $[5.0 - 10.0\text{Hz}]$ range is not standard in EEG research. Instead, research tends to focus on the θ $[4.1 - 8.0\text{Hz}]$ and α $[8.1 - 13.0\text{Hz}]$ bands. A careful inspection of the plot will reveal that in the δ, θ and

α frequency ranges the proportion of power tends to show cycles across time (note the wavy pattern of the data as time progresses from sleep onset). While it may be less clear from Figure 1, the δ-band behavior tends to be negatively correlated with θ and α-bands. This happens because there is a natural trade-off between slow and fast neuronal firing.

3 Population Value Decomposition

In this section we introduce a population level data compression that allows the coefficients of each image to be comparable and interpretable across images. If $\boldsymbol{Y}_i$, $i = 1, \ldots, n$, is a sample of $F \times T$ dimensional images then a Population Value Decomposition (PVD) is

$$\boldsymbol{Y}_i = \boldsymbol{P}\boldsymbol{V}_i\boldsymbol{D} + \boldsymbol{E}_i, \tag{2}$$

where $\boldsymbol{P}$ and $\boldsymbol{D}$ are population-specific matrices of size $F \times A$ and $B \times T$, respectively, $\boldsymbol{V}_i$ is a $A \times B$ dimensional matrix of subject-specific coefficients, and $\boldsymbol{E}_i$ is an $F \times T$ dimensional matrix of residuals. Many different decompositions of type (2) exist. Consider, for example, any two full-rank matrices $\boldsymbol{P}$ and $\boldsymbol{D}$, where $A < F$ and $B < T$. Equation (2) can be written in vector format as follows. Denote by $\boldsymbol{y}_i = \text{vec}(\boldsymbol{Y}_i^T)$, $\boldsymbol{v}_i = \text{vec}(\boldsymbol{V}_i^T)$, $\boldsymbol{\epsilon}_i = \text{vec}(\boldsymbol{e}^T)$ the column vectors obtained by stacking the row vectors of $\boldsymbol{Y}_i$, $\boldsymbol{V}_i$, and $\boldsymbol{E}_i$, respectively. If $\boldsymbol{X} = \boldsymbol{P} \otimes \boldsymbol{D}^T$ is the $FT \times AB$ Kronecker product of matrices $\boldsymbol{P}$ and $\boldsymbol{D}$ then equation (2) becomes the following standard regression $\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{v}_i + \boldsymbol{\epsilon}_i$. Thus, a least squares estimator of $\boldsymbol{v}_i$ is $\widehat{\boldsymbol{v}}_i = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_i$. This provides a simple recipe for obtaining the subject-specific scores, $\boldsymbol{v}_i$ or, equivalently, $\boldsymbol{V}_i$, once the matrices $\boldsymbol{P}$ and $\boldsymbol{D}$ are fixed. The scores can be used in standard statistical models either for prediction or associations studies. Note that $\boldsymbol{X}'\boldsymbol{X}$ is a small dimensional matrix that is easy to invert. Moreover, all calculations can be done

even on very large images by partitioning files into sub-files and using block-matrix computations.

3.1 Default Population Value Decomposition

There are many types of PVDs and definitions can and will change in particular applications. In this section we introduce our default procedure, which is inspired by the subject-specific SVD and by the thought experiment described in Section 1. Consider the case when for every subject-specific image one can obtain the SVD. This can be done in all applications we are aware of including the SHHS and fMRI studies; see [1] for an example.

For every subject let $\boldsymbol{Y}_i = \boldsymbol{U}_i \boldsymbol{\Sigma}_i \boldsymbol{V}_i^T$ be the SVD of the image. If $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ were the same across all subjects then the SVD would be the default PVD. However, in practice $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ will tend to vary from person to person. Mimicking the thought process described in Section 1 we try to find the common features across subjects among the column vectors of the $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$ matrices.

We start by considering the $F \times L_i$ dimensional matrix $\boldsymbol{U}_{L_i}$ consisting of the first L_i columns of the matrix $\boldsymbol{U}_i$ and the $T \times R_i$ dimensional matrix consisting of the first R_i columns of the matrix $\boldsymbol{V}_i$. The choice of L_i and R_i could be based on various criteria including variance explained, signal-to-noise ratios, or practical considerations. This is not a major concern in this paper.

We focus on $\boldsymbol{U}_{L_i}$ as a similar procedure is applied to $\boldsymbol{V}_{R_i}$. Consider the $F \times L$ dimensional matrix $\boldsymbol{U} = [\boldsymbol{U}_{L_1} | \ldots | \boldsymbol{U}_{L_n}]$, where $L = (\sum_{i=1}^{n} L_i)$, obtained by horizontally binding the $\boldsymbol{U}_{L_i}$ matrices across subjects. The space spanned by the columns of $\boldsymbol{U}$ is a subspace of $\mathbb{R}^F$ and contains subject-specific left eigenvectors that explain most of the observed variability. While these vectors are not identical, they will be similar if images share common features. Thus, we propose to apply principal component

analysis (PCA) to the matrix $\boldsymbol{U}\boldsymbol{U}^T$ to obtain the main directions of variation in the column space of $\boldsymbol{U}$. Let $\boldsymbol{P}$ be the $F \times A$ dimensional matrix formed with the first A eigenvectors of $\boldsymbol{U}\boldsymbol{U}^T$ as columns, where A is chosen to ensure that a certain percentage of variability is explained. Then the matrix $\boldsymbol{U}$ is approximated via the projection equation $\boldsymbol{U} \approx \boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{U})$. At the subject level one obtains $\boldsymbol{U}_{L_i} \approx \boldsymbol{P}(\boldsymbol{P}^T\boldsymbol{U}_{L_i})$. This approximation becomes a tautological equality if $A = F$, that is, if we use the entire eigenbasis. Similar approximations can be obtained using any orthonormal basis; we prefer the eigenbasis for our default procedure because it is parsimonious. Similarly we obtain $\boldsymbol{D}^T$ a $T \times B$ dimensional matrix of the first eigenvectors of the matrix $\boldsymbol{V}\boldsymbol{V}^T$ where $\boldsymbol{V} = [\boldsymbol{V}_{R_1}|\ldots|\boldsymbol{V}_{R_n}]$. We have the similar approximation $\boldsymbol{V} \approx \boldsymbol{D}(\boldsymbol{D}^T\boldsymbol{V})$. At the subject level one obtains $\boldsymbol{V}_{R_i} \approx \boldsymbol{D}^T(\boldsymbol{D}\boldsymbol{V}_{R_i})$. We conclude that PVD is a two-step approximation process for all images that can be summarized as follows

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{V}_i^T \approx \boldsymbol{U}_{L_i}\boldsymbol{\Sigma}_{L_i,R_i}\boldsymbol{V}_{R_i}^T \\ &\approx \boldsymbol{P}\{(\boldsymbol{P}^T\boldsymbol{U}_{L_i})\boldsymbol{\Sigma}_{L_i,R_i}(\boldsymbol{V}_{R_i}^T\boldsymbol{D}^T)\}\boldsymbol{D}, \end{aligned} \tag{3}$$

where $\boldsymbol{U}_{L_i}$ and $\boldsymbol{V}_{R_i}$ are obtained by retaining the first L_i and R_i columns from the matrices $\boldsymbol{U}_i$ and $\boldsymbol{V}_i$, respectively, and $\boldsymbol{\Sigma}_{L_i,R_i}$ is obtained by retaining the first L_i rows and R_i columns from the matrix $\boldsymbol{\Sigma}_i$. The first approximation of the image $\boldsymbol{Y}_i$, first row in equation (3), is obtained by retaining the left and right eigenvectors that explain most of the observed variability at the subject level. The second approximation, second row in equation (3), is obtained by projecting the subject-specific left and right eigenvectors on the corresponding population-specific eigenvectors.

If we denote by $\boldsymbol{V}_i = (\boldsymbol{P}^T\boldsymbol{U}_{L_i})\boldsymbol{\Sigma}_{L_i,R_i}(\boldsymbol{V}_{R_i}^T\boldsymbol{D}^T)$ then we obtain the PVD equation (2). This formula reveals that $\boldsymbol{V}_i$ will not, in general, be a diagonal matrix even though $\boldsymbol{\Sigma}_{L_i,R_i}$ is. This is one of the fundamental differences between SVD and PVD. Note that all approximations can be trivially transformed into equalities. For exam-

ple, choosing $L_i = F$ and $R_i = T$ will ensure equality in the first approximation, while choosing $A = F$ and $B = T$ will ensure equality in the second equation. From a practical perspective these cases are not of scientific importance as no data compression would be achieved. However, our focus is on parsimony and not perfection of the approximation. The choice of L_i, R_i, A and B could be based on various criteria including variance explained, signal-to-noise ratios, or practical considerations. In this paper we use thresholds for the percent variance explained.

Calculations in this section are possible due to the following matrix algebra trick. We summarize this trick that allows calculations of SVD for very large matrices as long as one of the dimensions is not much larger than a few thousands.

Suppose that $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$ *is the SVD decomposition of an* $F \times T$ *dimensional matrix where, say,* F *is very large and* T *is moderate. Then* $\boldsymbol{D}$ *and* $\boldsymbol{V}$ *can be obtained from the spectral decomposition of the* $T \times T$ *dimensional matrix* $\boldsymbol{Y}^T\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{D}^2V^T$. *The* $\boldsymbol{U}$ *matrix can then be obtained from* $\boldsymbol{U} = \boldsymbol{Y}\boldsymbol{V}\boldsymbol{D}^{-1}$.

3.2 Functional statistical modeling

An immediate application of PVD is to use the entries subject-specific matrix $\boldsymbol{V}_i$ as predictors. For this purpose one can use a range of strategies from using one entry at a time to using groups of entries or selection or averaging algorithms based on prediction performance. The first example of such an approach is [1] who found empirical evidence of alternative connectivity in clinically asymptomatic at-risk of Alzheimer subjects when compared to controls. The authors used PVD with a 5×5 dimensional $\boldsymbol{V}_i$ and boosting to identify important predictors.

Here we focus on how PVD can be used to conduct nonparametric analysis of the images themselves. Specifically, we are interested in approximating the Karhunen-Loève (KL) decomposition [16, 18] of a sample of images. More precisely, if $\boldsymbol{y}_i =$

vec($\boldsymbol{Y}_i^T$) is the vector obtained by stacking the rows of the matrix $\boldsymbol{Y}_i$ we would like to obtain a decomposition of the type $\boldsymbol{y}_i = \sum_{k=1}^{K} \xi_{ik}\boldsymbol{\Phi}_k + \boldsymbol{e}_i$, where $\boldsymbol{\Phi}_k$ are the orthonormal eigenfunctions of the covariance operator, $\boldsymbol{K}_y$, of the process $\boldsymbol{y}$ and ξ_{ik} are the random uncorrelated scores of subject i on eigenfunction k, and $\boldsymbol{e}_i$ is an error process that could be, but typically is not, zero. A direct, or brute-force, functional approach to this problem would require the calculation, diagonalization, and smoothing of $\widehat{K}_y$, which is a $FT \times FT$ dimensional matrix. This can be done relatively easily when FT is small, but becomes computationally prohibitive as FT increases. For example, in the SHHS one could deal with data for all frequencies in the δ-band ($F = 17$) and one hour of sleep ($T = 120$), as computational complexity increases sharply both with respect to F and T. Indeed, computational complexity is $O(F^3T^3)$ and storage requirements are $O(F^2T^2)$. Table 1 displays the computing time required by the direct functional approach using a personal computer with dual core processors 3GHz CPU and 8Gb RAM. Computing time increases steeply with T and F making the approach impractical when both exceed about 100. Thus, it is essential to develop methods that accelerate the analysis. The PVD offers one solution.

Table 1: The comparison of computing time (in minutes) for functional data analysis of samples of images for various number of grid points in the time and frequency domains

N_{freq}	N_{time}					
	20	40	60	80	100	120
8	0.1	0.3	0.7	1.3	2.1	3.1
16	0.3	1.4	3.0	5.7	8.7	13.2
32	1.3	5.5	12.9	19.4	32.9	49.8
64	4.7	20.5	51.8	97.3	176.0	496.5
128	21.5	100.7	467.0	681.0	1195.6	2097.1

3.2.1 Functional principal component analysis of samples of images

To avoid the brute-force approach we propose to first obtain the spectral decomposition of the vectors $\boldsymbol{v}_i$, or, equivalently, of the corresponding matrix $\boldsymbol{V}_i$. As discussed, we expect that in most applications the matrix $\boldsymbol{V}_i$ will have far fewer than 500 entries; thus, obtaining a decomposition for $\boldsymbol{v}_i$ instead of $\boldsymbol{y}_i$ is not only achievable, but very fast. The KL expansion for the $\boldsymbol{v}_i$ process can be obtained easily; see, for example [21, 30]. The expansion can be written directly in matrix format as

$$\boldsymbol{V}_i = \sum_{k=1}^{K} \xi_{ik} \boldsymbol{\phi}_k + \boldsymbol{\eta}_i, \tag{4}$$

where $\boldsymbol{\phi}_k$ are the eigenvectors of the process $\boldsymbol{v}$ written as an $A \times B$ matrix, $\boldsymbol{\eta}_i$ is a noise process, and ξ_{ik} are mutually uncorrelated random coefficients. Here all vector to matrix transformations follow the same rules of the transformations $\boldsymbol{v}_i \leftrightarrow \boldsymbol{V}_i$. By left and right multiplying in equation (4) with the $\boldsymbol{P}$ and $\boldsymbol{D}$ matrices, respectively, we obtain the following decomposition of the sample of images

$$\begin{aligned} \boldsymbol{Y}_i &= \textstyle\sum_{k=1}^{K} \xi_{ik} \boldsymbol{P}\boldsymbol{\phi}_k\boldsymbol{D} + \boldsymbol{P}\boldsymbol{\eta}_i\boldsymbol{D} + \boldsymbol{E}_i \\ &= \textstyle\sum_{k=1}^{K} \xi_{ik} \boldsymbol{\Phi}_k + \boldsymbol{e}_i, \end{aligned} \tag{5}$$

where $\boldsymbol{\Phi}_k = \boldsymbol{P}\boldsymbol{\phi}_k\boldsymbol{D}$ is an $F \times T$ dimensional image, and $\boldsymbol{e}_i = \boldsymbol{P}\boldsymbol{\eta}_i\boldsymbol{D} + \boldsymbol{E}_i$ is an $F \times T$ noise process. These results provide a constructive recipe for image decomposition with the following simple steps: 1) obtain $\boldsymbol{P}$, $\boldsymbol{D}$ and $\boldsymbol{V}_i$ matrices as described in Section 3.1; 2) obtain the eigenfunctions $\boldsymbol{\phi}_k$ of the covariance operator of $\boldsymbol{V}_i$; 3) obtain the scores ξ_{ik} from the mixed effect model (4); and 4) obtain the basis for the image expansion $\boldsymbol{\Phi}_k = \boldsymbol{P}\boldsymbol{\phi}_k\boldsymbol{D}$. The following results provide the theoretical insights supporting this procedure.

Theorem 1 *Suppose that* $\boldsymbol{P}$ *is a matrix obtained by column binding* A *orthonormal eigenvectors of size* $F \times 1$ *and* D *is a matrix obtained by row binding* B *orthonormal eigenvectors of size* $1 \times T$*. Then the following results hold: i) the vector version of the eigenimages* $\boldsymbol{\Phi}_k = \boldsymbol{P}\boldsymbol{\phi}_k\boldsymbol{D}$ *are orthonormal in* $\mathbb{R}^{FT}$*; and ii) the scores* ξ_{ik} *are exactly the same in equations (4) and (5).*

3.2.2 Multilevel functional principal component analysis of samples of images

There are many studies, including our own SHHS, where images have a natural multilevel structure. This happens, for example, when image data are clustered within the subjects or data are observed at multiple visits within the same subject. PVD provides a natural way of working with the data in this context. Suppose that $\boldsymbol{Y}_{ij}$ are images observed on subject i at time j and assume that $\boldsymbol{Y}_{ij} = \boldsymbol{P}\boldsymbol{V}_{ij}\boldsymbol{D} + \boldsymbol{E}_{ij}$ is the default PVD for the entire collection of images. Using the MFPCA methodology introduced by [11] and further developed in [10] we can decompose the $\boldsymbol{V}$ process into subject- and subject/visit-specific components. More precisely,

$$\boldsymbol{V}_{ij} = \sum_{k=1}^{K} \xi_{ik}\boldsymbol{\phi}_k^{(1)} + \sum_{l=1}^{L} \zeta_{ijl}\boldsymbol{\phi}_l^{(2)} + \boldsymbol{\eta}_i, \tag{6}$$

where $\boldsymbol{\phi}_k^{(1)}$ are mutually orthonormal subject-specific (or level 1) eigenvectors, where $\boldsymbol{\phi}_k^{(2)}$ are mutually orthonormal subject/visit-specific (or level 2) eigenvectors, and $\boldsymbol{\eta}_i$ is a noise process. The level 1 and 2 eigenvectors are required to be orthonormal within level not across levels. The subject-specific scores, ξ_{ik}, and the subject/visit specific scores, ζ_{ijl}, are assumed to be mutually uncorrelated random coefficients. Just as in the case of a cross-sectional sample of images we can multiply the equation (6) with the matrix $\boldsymbol{P}$ at the left and $\boldsymbol{D}$ at the right. We obtain the following model for a

sample of images with a multilevel structure

$$\begin{aligned} \boldsymbol{Y}_{ij} &= \sum_{k=1}^{K} \xi_{ik} \boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D} + \sum_{l=1}^{L} \zeta_{ijl} \boldsymbol{P}\boldsymbol{\phi}_l^{(2)}\boldsymbol{D} + \boldsymbol{P}\boldsymbol{\eta}_i\boldsymbol{D} + \boldsymbol{E}_i \\ &= \sum_{k=1}^{K} \xi_{ik} \boldsymbol{\Phi}_k^{(1)} + \sum_{l=1}^{L} \zeta_{ijl} \boldsymbol{\Phi}_l^{(2)} + \boldsymbol{e}_i, \end{aligned} \tag{7}$$

where $\boldsymbol{\Phi}_k^{(1)} = \boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D}$ is a subject-specific $F \times T$ dimensional image, $\boldsymbol{\Phi}_k^{(2)} = \boldsymbol{P}\boldsymbol{\phi}_k^{(2)}\boldsymbol{D}$ is a subject/visit-specific $F \times T$ dimensional image, and $\boldsymbol{e}_i = \boldsymbol{P}\boldsymbol{\eta}_i\boldsymbol{D} + \boldsymbol{E}_i$ is an $F \times T$ noise process. The following theorem shows that it is enough to conduct MFPCA on the simple model (6) instead of the intractable model (7).

Theorem 2 *Suppose that $\boldsymbol{P}$ is a matrix obtained by column binding A orthonormal eigenvectors of size $F \times 1$ and D is a matrix obtained by row binding B orthonormal eigenvectors of size $1 \times T$. Then the following results hold: i) the vector version of the subject-specific eigenimages $\boldsymbol{\Phi}_k^{(1)} = \boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D}$ are orthonormal in $\mathbb{R}^{FT}$; ii) the vector version of the subject/visit-specific eigenimages $\boldsymbol{\Phi}_l^{(2)} = \boldsymbol{P}\boldsymbol{\phi}_l^{(2)}\boldsymbol{D}$ are orthonormal in $\mathbb{R}^{FT}$; iii) the vector version of $\boldsymbol{\Phi}_k^{(1)}$ and $\boldsymbol{\Phi}_l^{(2)}$ are not necessarily orthogonal; and iv) the scores ξ_{ik} and ζ_{ijl} are exactly the same in equations (6) and (7).*

We contend that Theorems 1 and 2 provide simple ways of obtaining ANOVA-like decompositions of very large images based on computable algorithms even for massive images, such as those obtained from brain fMRI. While other methods may appear in the future, we consider that PVD provides one of the most exciting opportunities for the longitudinal analysis of images while using all, or almost all, available information. Proofs can be found in the web supplement.

4 Simulation Studies

In this section, we generate the frequency by time image $\boldsymbol{Y}_{ij}$ for subject i and visit j from the following model

$$\boldsymbol{Y}_{ij}(f,t) = \sum_{k=1}^{4} \xi_{ik}\phi_k^{(1)}(f,t) + \sum_{l=1}^{4} \zeta_{ijl}\phi_l^{(2)}(f,t) + \epsilon_{ij}(f,t) \text{ for } i = 1,\ldots,I,\ j = 1,\ldots,J \quad (8)$$

where $\xi_{ik} \sim N\{0, \lambda_k^{(1)}\}$ for $k = 1,\ldots,4$, $\zeta_{ijl} \sim N\{0, \lambda_l^{(2)}\}$ for $l = 1,\ldots,4$, $\epsilon_{ij}(f,t) \sim N(0,\sigma^2)$, $\{f = 0.2f \text{ Hz} : f = 1,\ldots,F\}$, where F is the number of frequencies, and $\{t = \frac{t}{T} : m = 1,2,\ldots,T\}$, where T is the number of epochs. We consider $F = 128$ and $T = 120$ in the simulation below. We simulate $I = 200$ subjects (clusters) with $J = 2$ visits per subject (measurement per cluster). The true eigenvalues are $\lambda_k^{(1)} = 0.5^{k-1}$, $k = 1,2,3,4$, and $\lambda_l^{(2)} = 0.5^{l-1}$, $l = 1,2,3,4$. We consider multiple scenarios corresponding to different noise magnitudes: $\sigma = 0$ (no noise), $\sigma = 2$ (moderate), $\sigma = 4$ (large). We conduct 100 simulations for each scenario. The frequency-time eigenfunctions $\phi_k^{(1)}(f,t)$ and $\phi_k^{(2)}(f,t)$ are generated from bases in frequency and time domains as illustrated below. The bases in the frequency domain are derived from the Haar family of functions defined as $\psi_{pq}(f) = 2^{p/2}/\sqrt{N}$ for $(q-1)/2^p \le (f - f_{min})/(f_{\max} - f_{\min}) < (q-0.5)/2^p$, $\psi_{pq}(f) = -2^{p/2}/\sqrt{N}$ for $(q-0.5)/2^p \le (f - f_{min}/(f_{\max} - f_{\min}) < q/2^p$ and $\psi_{pq}(f) = 0$ otherwise. Here N is the number of frequencies and $f_{\min}$ and $f_{\max}$ are the minimum and maximum frequencies under consideration, respectively. In particular, we let the level 1 eigenfunctions be $h_1^{(1)}(f) = \psi_{11}(f)$, $h_2^{(1)}(f) = \psi_{12}(f)$ and level 2 eigenfunctions be $h_1^{(2)}(f) = \psi_{21}(f)$, $h_2^{(2)}(f) = \psi_{22}(f)$. For example, if $f_{\min} = 0.2$Hz, $f_{\max} = 1.6$Hz, and frequency increments by 0.2Hz, then $N = 8$. The eigenfunctions in this case are $h_1^{(1)}(f) = (0.5, 0.5, -0.5, -0.5, 0, 0, 0, 0)$, $h_2^{(1)}(f) = (0, 0, 0, 0, 0.5, 0.5, -0.5, -0.5)$ and $h_1^{(2)}(f) = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, 0, 0, 0, 0, 0)$, $h_2^{(2)}(f) = (0, 0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 0, 0, 0, 0)$. For the time

domain we consider the following two choices.

Case 1. Mutually orthogonal bases. Level 1: $g_1^{(1)}(t) = \sqrt{2}\sin(2\pi t),\ g_2^{(1)}(t) = \sqrt{2}\cos(2\pi t)$. Level 2: $g_1^{(2)}(t) = \sqrt{2}\sin(6\pi t),\ g_2^{(2)}(t) = \sqrt{2}\cos(6\pi t)$.

Case 2. Mutually non-orthogonal bases. Level 1: same as in Case 1. Level 2: $g_1^{(2)}(t) = 1,\ g_2^{(2)}(t) = \sqrt{3}(2t-1)$.

In the following, we present only results for Case 2, as similar results were obtained for Case 1. The frequency-time eigenfunctions are generated by multiplying each component of the bases in frequency and time domains, i.e., $\phi_k^{(1)}(f,t) = h_{k_f}^{(1)}(f)^T g_{k_t}^{(1)}(t)$ where $k = k_f + 2(k_t - 1)$ for $k_f, k_t = 1, 2$ and $\phi_k^{(2)}(f,t) = h_{l_f}^{(2)}(f)^T g_{l_t}^{(2)}(t)$ where $l = l_f + 2(l_t - 1)$ for $l_f, l_t = 1, 2$. The first figure in the web supplement displays simulated data from model (8) for one subject at two visits with different magnitudes of noise. It shows that as the magnitude of noise increases, the patterns become harder to delineate. For clarity, in this plot we used $F = 16$ and $T = 20$.

4.1 Eigenvalues and eigenfunctions

Figure 2 shows estimated level 1 and 2 eigenvalues for the different magnitudes of noise using the PVD method described in Section 3. Note that the potential measurement error is not accounted for in this figure. In the case of no noise ($\sigma = 0$), the eigenvalues can be generally recovered without bias, although some small bias is present in the estimation of the first eigenvalue at level 2. The bias does not seem to increase substantially with the noise level. Overall, the quality of the estimation procedure is quite remarkable.

Figure 3 shows estimated eigenfunctions at four randomly selected frequencies from 20 simulated datasets. The simulated data have no measurement error, i.e., $\sigma = 0$. We conclude that PVD successfully separates level 1 and 2 variation and correctly captures the shape of each individual eigenfunction.

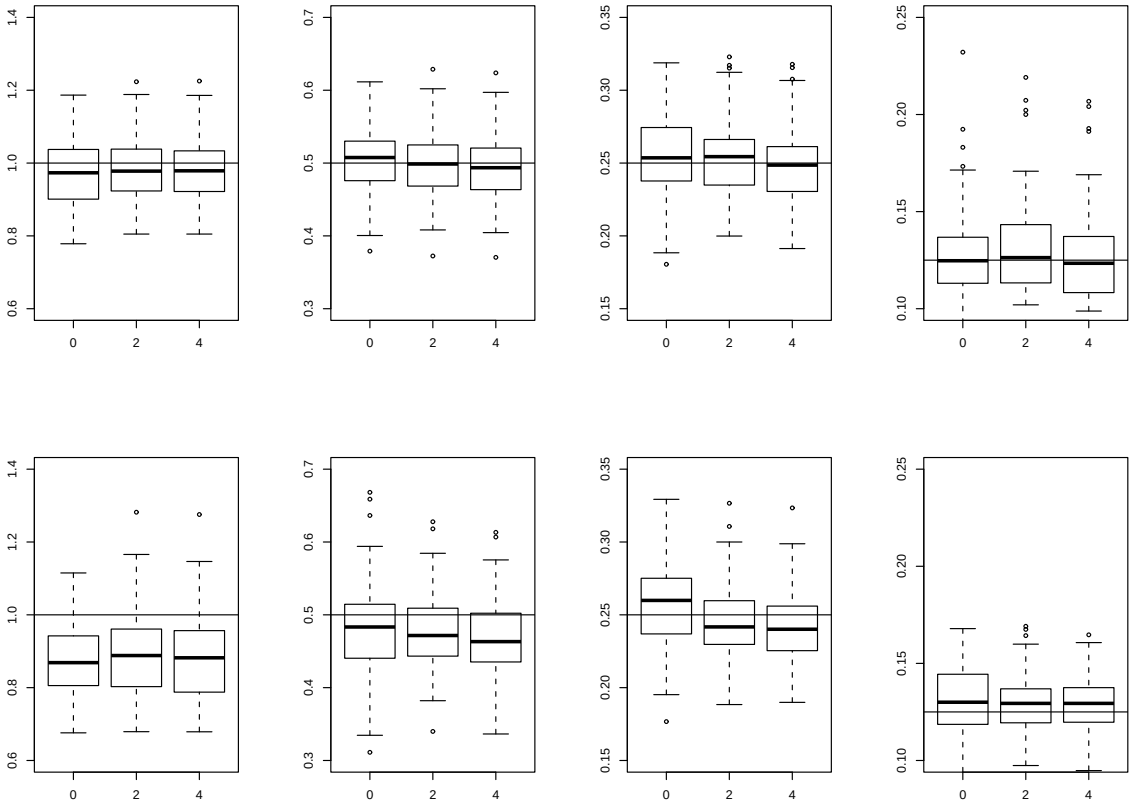

Figure 2: Boxplots of estimated eigenvalues using unsmooth MFPCA-3; the true functions are without and with noise. The solid gray lines are the true eigenvalues. The x-axis labels indicate the standard deviation of the noise.

4.2 Principal component scores

To estimate the scores we used Bayesian inference via posterior simulations using Markov Chain Monte Carlo (MCMC) methods. We used the software developed by [11] applied to the mixed effect model (8). Because this uses the full model the method will be referred to as PC-F. Because Bayesian calculations can be slow when the dimension of $\boldsymbol{V}_{ij}$ is very large, [11] introduced a projection model, which reduces computation time by orders of magnitude. Because this uses a projection in the original mixed effect model the method will be referred to as PC-P. In simulations

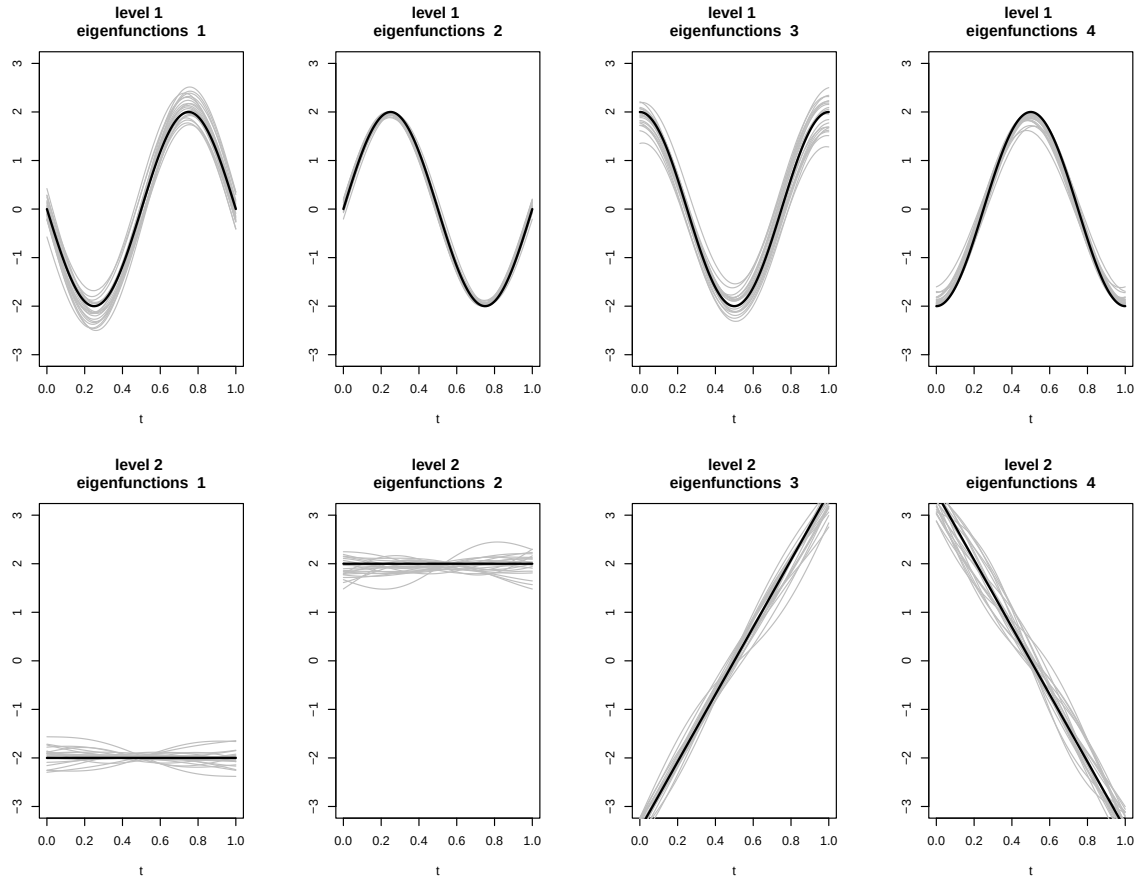

Figure 3: Estimated eigenfunctions at four randomly-selected frequencies from 20 simulated datasets when the frequency-time images are observed without noise, i.e., $\sigma = 0$. Thick black lines: true eigenfunctions at those randomly-selected frequencies. Gray lines: estimated eigenfunctions.

PC-P proved to be slightly less efficient, but much faster than PC-F. For a thorough introduction to Bayesian functional data analysis using WinBUGS [24] see [8].

We use the full model (PC-F) and the projection model (PC-P) proposed in [11] to estimate PC scores after obtaining the estimated eigenvalues and eigenfunctions using PVD. To compare the performance of these two models, we compute the root mean square errors (RMSE). In each scenario, we randomly select 10 simulated datasets and estimate the PC scores using posterior means from the Markov Chain Monte Carlo (MCMC) runs. The MCMC convergence and mixing properties are assessed by visual

inspection of the chain histories of many parameters of interest. The history plots (not shown) indicate very good convergence and mixing properties. Table 2 reports the means of the RMSE, indicating that when the amount of noise increases, the RMSE also increases. A direct comparison of the RMSE with the standard deviation of the scores at the four levels (1, 0.71, 0.50, 0.35) indicates that scores are well estimated, especially at level 1. Moreover, PC-F performs slightly better than PC-P in terms of RMSE; however, PC-P might still be preferred in applications where PC-F in computationally expensive.

Table 2: Root mean square errors for estimating scores using PC-F and PC-P

		Level 1 Component				Level 2 Component			
Method	σ	1	2	3	4	1	2	3	4
Case 2: PC-F	0	0.056	0.036	0.053	0.044	0.122	0.111	0.153	0.122
	2	0.065	0.051	0.065	0.060	0.132	0.121	0.178	0.131
	4	0.120	0.089	0.095	0.100	0.145	0.125	0.167	0.145
Case 2: PC-P	0	0.068	0.063	0.074	0.052	0.135	0.196	0.212	0.130
	2	0.079	0.087	0.087	0.060	0.138	0.227	0.258	0.150
	4	0.139	0.160	0.103	0.104	0.161	0.138	0.223	0.175

5 Application to the SHHS

In section 2 we introduced the SHHS, which collected two PSG for thousands of subjects roughly 5 years apart. Here we focus on analyzing the frequency by time spectrograms for $N = 3,201$ subjects at $J = 2$ visits. We analyze all frequencies from 0.2Hz to 32Hz in increments of 0.2Hz for a total number of $F = 160$ grid points in frequency and the first 4 hours of sleep in increments of 30 seconds for a total number of $T = 480$ grid points in time. The total number of observations per subject per visit is $FT = 76,800$ and the total number of observations across all subjects and visits is

$FTNJ = 491,673,600$. The same methods could easily be applied to fMRI studies, where one image would contain more than $V = 2,000,000$ voxels and $T = 500$ time points for a total of $VT = 1,000,000,000$ observations per image. Methods described in this paper are designed to scale up well to these larger imaging studies.

For each subject i, $i = 1, \ldots, I = 3,201$ and visit j, $j = 1, J = 2$, we obtained $\boldsymbol{Y}_{ij}$, the $F \times T = 160 \times 480$ dimensional frequency by time spectrogram. We demean the row and column vectors of each matrix using the transformation $\boldsymbol{Y}_{ij} \mapsto \{\boldsymbol{I}_F - \boldsymbol{E}_F/F\}\boldsymbol{Y}_{ij}\{\boldsymbol{I}_T - \boldsymbol{E}_T/T\}$, where $\boldsymbol{I}_F, \boldsymbol{I}_T$ denote the identity matrices of size F and T, and $\boldsymbol{E}_F$ and $\boldsymbol{E}_T$ are square matrices with each entry equal to 1 of size F and T, respectively. Note that any image $\boldsymbol{Y}_{ij}$ can be written as

$$\boldsymbol{Y}_{ij} = \{\boldsymbol{I}_F - \boldsymbol{E}_F/F\}\boldsymbol{Y}_{ij}\{\boldsymbol{I}_T - \boldsymbol{E}_T/T\} + \boldsymbol{E}_F\boldsymbol{Y}_{ij}/F + \boldsymbol{Y}_{ij}\boldsymbol{E}_T/T - \boldsymbol{E}_F\boldsymbol{Y}_{ij}\boldsymbol{E}_T/(FT).$$

The last term of the equality, $\boldsymbol{E}_F\boldsymbol{Y}_{ij}\boldsymbol{E}_T/(FT)$, is an $F \times T$ dimensional matrix with all entries equal to the average of all entries in $\boldsymbol{Y}_{ij}$. The third term of the equality, $\boldsymbol{Y}_{ij}\boldsymbol{E}_T/T$ is a matrix with T identical columns equal to the row means of the matrix $\boldsymbol{Y}_{ij}$. Similarly, $\boldsymbol{E}_F\boldsymbol{Y}_{ij}/F$ is a matrix with F identical rows equal to the column means of the matrix $\boldsymbol{Y}_{ij}$. We conclude that the inherently bi-variate information in the image $\boldsymbol{Y}_{ij}$ is encapsulated in $\{\boldsymbol{I}_F - \boldsymbol{E}_F/F\}\boldsymbol{Y}_{ij}\{\boldsymbol{I}_T - \boldsymbol{E}_T/T\}$. Methods for analyzing the average of the entire image are standard. Methods for analyzing the column and row means of the image are either classical or have been recently developed [10, 11, 25]. Thus, we focus on analyzing $\{\boldsymbol{I}_F - \boldsymbol{E}_F/F\}\boldsymbol{Y}_{ij}\{\boldsymbol{I}_T - \boldsymbol{E}_T/T\}$ and we continue to denote this $F \times T$ dimensional matrix by $\boldsymbol{Y}_{ij}$. With this definition of $\boldsymbol{Y}_{ij}$ we proceed with the main steps of our analysis. First, we obtain the subject/visit specific SVD decomposition $\boldsymbol{Y}_{ij} = \boldsymbol{U}_{ij}\boldsymbol{\Sigma}_{ij}\boldsymbol{V}_{ij}^T$. We then store the first $L_i = 10$ columns of the matrix $\boldsymbol{U}_{ij}$ in the matrix $\boldsymbol{U}_{L_i,j}$ and construct the two matrices $\boldsymbol{U}_j = [\boldsymbol{U}_{L_1,j}| \ldots |\boldsymbol{U}_{L_I,j}]$

for $j = 1, 2$. Both matrices $\boldsymbol{U}_j$ are $160 \times 32,010$ dimensional and we obtain the $160 \times 64,020$ dimensional matrix $\boldsymbol{U} = [\boldsymbol{U}_1|\boldsymbol{U}_2]$ by column binding $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$. To obtain the main directions of variation in the space spanned by the column space of the matrix $\boldsymbol{U}$, we diagonalize the 160×160 dimensional matrix $\boldsymbol{U}\boldsymbol{U}^T$. The eigenvectors of the matrix $\boldsymbol{U}\boldsymbol{U}^T$ will be called population eigenfrequencies. In fMRI these are called eigenimages [1]. A similar construction and decomposition is applied to the matrix $\boldsymbol{V}\boldsymbol{V}^T$ whose eigenvectors are called eigenvariates. Because $\boldsymbol{V}\boldsymbol{V}^T$ is much noisier than $\boldsymbol{U}\boldsymbol{U}^T$ we first apply row-by-row smoothing of $\boldsymbol{V}\boldsymbol{V}^T$. Bivariate smoothing is prohibitively slow, but this approach proved to be fast.

Table 3 displays some important eigenvalues of $\boldsymbol{U}\boldsymbol{U}^{\boldsymbol{T}}$ and $\boldsymbol{V}\boldsymbol{V}^T$, respectively. Results are reassuring and indicate that our intuition that samples of images have many common features is warranted. Indeed, the first 13 population level eigenfrequencies explain more than 90% of the variability of collection of first 10 subject-specific eigenfrequencies over more than $3,000$ subjects. Another interesting property of the population eigenfrequencies is that the most important 5-7 of them explain a similar amount of variability; please note the very slow decay in the associated variance components. The variance explained decays exponentially starting with component 8 and becomes practically negligible for components 15 and beyond. Returning to our thought experiment this means that if we look at the frequency (X) dimension across subjects there will be a lot of consistency in terms of the shape and location of the observed signal. This is consistent with the population data, which, across subjects, shows higher proportion power and variability in the δ and α power bands. Our results provide a quantification for this general observation while remaining agnostic to the classical partition of the frequency domain.

A similar story can be told about the eigenvariates, but some of the specifics are different. More precisely, the variance explained by individual eigenvariates de-

creases more linearly and does not exhibit any sudden drop. Moreover, the first 13 eigenvariates explain roughly 80% of the observed variability of the subject-specific eigenvariates and 20 eigenvariates are necessary to explain 90% of the variability.

The shape of the first 10 population level eigenfrequencies and eigenvariates are displayed in Figures 4 and 5. The most remarkable aspect of these plots is that they make sense. Indeed, Figure 4 indicates that most of the variability is in a range of frequencies that roughly overlaps with the δ power band range $[0.8, 4\text{Hz}]$. This should not be surprising as most of the observed variability is *obviously* in this frequency range. However, the level and type of variability we identified in the δ power band is novel. For example, subjects who are positively loaded on the first eigenfrequency (top-left plot in Figure 4) will tend to have much higher percent power around frequency 0.6Hz than around 1.2Hz. Similarly, a subject who is positively loaded on the second eigenfrequency (top-right panel in Figure 4) will have higher percent power around frequencies 0.4 and 1.2Hz than around 0.8Hz. Moreover, differences between percent power in these frequencies are quite sharp. Another interesting finding is that the first 5 eigenfrequencies seem "dedicated" to discrepancies in the low part of the frequency range $[0.2, 2\text{Hz}]$. Each of these eigenfrequencies explain roughly 10% of the eigenfrequency variability for a combined 49% explained variability. Starting with eigenfrequency 6 there is a slow but steady shift towards discrepancies at higher frequency. Moreover, higher eigenfrequencies display more detail in the 8 to 10Hz range, which is well within the α power range $[8.1, 13.0\text{Hz}]$.

The eigenvariates shown in Figure 5 tell an equally interesting, but different, story. First all eigenvariates indicate that in the time domain differences tend to be smooth with very few sudden changes. An alternative interpretation would be that some transitions may happen very fast in time, but are undetectable in the signal. A closer look at the first eigenvariate indicates that, relative to the population aver-

age, subjects who are positively loaded on this component (top-left plot) will tend to have: 1) higher percent power between the 30th and 50th minute; 2) slightly lower percent power between minute 70 and 80; 3) higher percent power between minutes 120 and 140, but the discrepancy is smaller than the one around minute 40; and 4) smaller percent power between minutes 180 and 210. The other eigenvariates have similarly interesting interpretations. It is worth noting that eigenvariates become roughly sinusoidal starting with the 7th eigenvariate. There are at least two alternative explanations for this occurrences. First, it could be that there are, indeed, high frequency cycles in the population. Another possible explanation is that the distances between peaks and valleys vary randomly across subjects; see Woodard, Crainiceanu and Ruppert [29] for an explanation of this behavior.

The eigenfrequencies and eigenvariates are interesting in themselves, but it is the Kronecker product of these bases that provides the projection basis for the actual images. Figure 6 displays some population level basis components obtained as Kronecker products of eigenfrequencies and eigenvariates. We call these eigenimages; in the fMRI context [1] eigenimages are referred to as eigenvectors or eigencomponents and eigenfrequencies are referred to as eigenimages. The x-axis represents the frequencies from 0.2 to 8Hz and the y-axis represent the time from sleep onset until the end of the 4th hour are on the y-axis. Images are cut at 8Hz to focus on the more interesting part of the graph, but analysis was conducted on frequencies up to 32Hz. The title of each image indicates the eigenfrequency number (F) and eigenvariate number (T), as ordered by their corresponding eigenvalues. For example, F = 1, T = 7 indicates the basis component obtained as a Kronecker product of the 1st eigenfrequency and the 7th eigenvariate. The checkerboard patterns observed in the right panels are due to the 7th and 10th eigenvariate, which are the sinus-like functions displayed in Figure 5.

We now investigate the smoothing effects of the population level eigenimages. The top left panel in Figure 7 displays the frequency by time plot of the fraction power for the same subject shown in Figure 1. The only difference is that the time interval was reduced to the first 4 hours after sleep onset. The top-right panel displays the projection of the frequency by time image on a basis with 225 components obtained as Kronecker products of the first 15 population level eigenfrequencies and the first 15 population level eigenvariates. The smooth surface provides a pleasing summary of the main features of the original data by reducing some of the observed noise. The bottom-left plot displays the projection of the frequency by time image on a basis with 45 components obtained as Kronecker products of the first 15 subject level eigenfrequencies and the first 3 subject level eigenvariates. We did not include more subject level eigenvariates because they were indistinguishable from noise. The bottom-right plot displays the difference between the projection on the subject level basis (bottom-left panel) and the projection on the population level basis (top-right panel). We conclude that both projections on the subject level and the population level bases reduce the noise in the original image and provide pleasing summaries of the main features of the data. The two summaries are not identical, with the subject-level smooth being slightly closer to the original data in the δ frequency range (note the sharper peaks) and the population-level smooth being closer to the original data in the α frequency range (compare the number and size of peaks). While one could argue about what basis one should use at the subject level, there is no doubt that having a population level basis with reasonable smoothing properties is an excellent tool if the final goal is statistical inference on populations of images. The current practice of taking averages over frequencies in the δ power band can be viewed as a much cruder alternative. These plots also indicate a potential challenge that was not addressed. The variability around the signal seem to be roughly proportional with

the signal, a rather unexpected feature of the data that deserves farther investigation. This problem exceeds the scope of the current paper.

To analyze the clustering of images we used a basis with 100 components obtained by taking the Kronecker product of the first 10 eigenfrequencies and first 10 eigenvariates. Examples of these components are shown in Figure 6. The subject/visit-specific coefficients were obtained by projecting the original images on this basis, which resulted in a 100-dimensional vector of coefficients. Thus, we applied MFPCA [11] to $I = 3,201$ subjects observed at $J = 2$ visits, each subject/visit being characterized by a vector $\boldsymbol{v}_{ij}$ of 100 coefficients. This took less than 10 seconds using a personal computer with dual core processors 3GHz CPU and 8Gb RAM. We fit the model (6) from Section 3.2.2 in matrix form $\boldsymbol{V}_{ij} = \sum_{k=1}^{K} \xi_{ik}\boldsymbol{\phi}_k^{(1)} + \sum_{l=1}^{L} \zeta_{ijl}\boldsymbol{\phi}_l^{(2)} + \boldsymbol{\eta}_i$, where $\xi_{ik} \sim \mathrm{N}\{0, \lambda_k^{(1)}\}$, $\zeta_{ijl} \sim \mathrm{N}\{0, \lambda_l^{(2)}\}$ are mutually uncorrelated. We first focused on estimating $\lambda_k^{(1)}$, $\lambda_l^{(2)}$, $\boldsymbol{\phi}_k^{(1)}$, $\boldsymbol{\phi}_l^{(2)}$, K and L. The table in the web appendix provides the estimates for the first 10 eigenvalues indicating that the level 2 eigenvalues quantifying the visit-specific variability are roughly 100 times larger than the level 1 eigenvalues quantifying the subject-specific variability. Using the same notation as in [11] the proportion of variance explained by within-subject variability is $\rho_W = (\sum_{k=1}^{100} \lambda_k^{(1)})/(\sum_{k=1}^{100} \lambda_k^{(1)} + \sum_{l=1}^{100} \lambda_l^{(2)})$. A plug-in estimator of ρ_W is $\widehat{\rho}_W = 0.033$, which indicates that the between-subject variability is very small compared to the within-subject between-visit variability. In studies of δ-power [11] estimated a much higher ρ_W, in the range $[0.15, 0.20]$, depending on the particular application. Our results do not contradict these previous results, as the subject-specific mean over all time points was removed from the bivariate spectrogram. However, they indicate that in the SHHS most of the within-subject correlation is contained in the margins of the frequency by time image. The margins are the column and row means of the original bivariate plots.

The left panels in Figure 8 displays the first 4 subject-level eigenfunctions, $\boldsymbol{\phi}_k^{(1)}$, $k = 1, \ldots, K$, in the coefficient space. In matrix format these bases are 10×10 dimensional and are hard to interpret. However, by pre- and post-multiplying them with the population level matrices P and D, we obtain the eigenimages in the original space, $\boldsymbol{\Phi}_k^{(1)} = \boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D}$. These eigenimages are displayed in the corresponding right panels of Figure 8. The second figure in the web supplement provides the same results for the level 2 eigenimages.

6 Discussion

Statistical analysis of populations of images when even one image cannot be loaded in the computer memory is a daunting task. Historically, data compression or signal extraction methods aim at reducing the very large images to a few indices that can be then analyzed statistically. Examples are total brain volume obtained from MRI studies or average percent δ-power in sleep EEG studies. In this paper we have proposed an integrated approach to signal extraction and statistical analysis that: 1) uses the entire information available in images; 2) is computationally fast and scalable to much larger studies; 3) provides equivalence results between the analysis of populations of image coefficients and populations of images. The approach was applied to the SHHS, arguably one of the largest studies analyzed statistically. Indeed, only the EEG data in the study contains more than 85 billion observations. We are in the process of deploying our methodology to longitudinal studies of fMRI, which are roughly 3 orders of magnitude larger than the SHHS. What is called "massive" is very quickly changing in Statistics.

The most important contribution of our paper is furthering the foundation for next generation statistical studies. We call this area the large N, large P, large J

problem, where N denotes the number of subjects, P denotes the dimensionality of the problem, and J denotes the number of visits or observations within cluster. Note that the famous small N large P problem can be obtained from our problem by setting $J = 1$ and cutting N. Our methods are designed for K-dimensional matrices where dimensions naturally split into 2 different modalities, e.g. time and frequency in spectral analysis and time and space in fMRI. Because we use a two-stage singular value decomposition (SVD) our method inherits the weaknesses of the SVD: 1) sensitivity to noise, correlation and outliers; 2) dependence on methods for choosing the dimension of the underlying linear space; and 3) lack of invariance under nonlinear transformations of the data.

It is important to better position our work with respect to other methods used for image analysis, including Principal Component Analysis (PCA) [5, 15, 23], Independent Component Analysis (ICA) [6, 13, 14] and Partial Least Squares (PLS) [7, 27, 28]. In short, our method is a multi-stage PCA method. Indeed, the subject level SVD of the data matrix $\boldsymbol{Y}_i$ is a decomposition $\boldsymbol{Y}_i = \boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{V}_i$, where: 1) $\boldsymbol{V}_i$ are the right eigenvectors of the matrix $\boldsymbol{Y}_i$ and satisfy $\boldsymbol{Y}_i^T\boldsymbol{Y}_i = \boldsymbol{V}_i^T\boldsymbol{\Sigma}_i^2\boldsymbol{V}_i$; 2) $\boldsymbol{U}_i$ are the left eigenvectors of the matrix $\boldsymbol{Y}_i$ and satisfy $\boldsymbol{Y}_i\boldsymbol{Y}_i^T = \boldsymbol{U}_i^T\boldsymbol{\Sigma}_i^2\boldsymbol{U}_i^T$; and 3) $\boldsymbol{\Sigma}_i$ is a diagonal matrix containing the square roots of the eigenvalues of $\boldsymbol{Y}_i^T\boldsymbol{Y}_i$ and $\boldsymbol{Y}_i\boldsymbol{Y}_i^T$ on the main diagonal. The method we proposed is a multi-stage PCA method because it extracts the firs K left and right subject-specific eigenvectors, it stacks them and conducts a second-stage PCA analysis on the stacked eigenvectors. ICA is an excellent tool for decomposing variability in independent rather than uncorrelated components and works very well when signals are non-normal. However, statistically principled ICA analysis of populations of images is still in its infancy. Group ICA [2, 3] cannot currently be applied to, say, hundreds of functional Magnetic Resonance Images (fMRIs). Moreover, ICA uses PCA as a pre-processing

step before conducting ICA. We are aware that the team behind the 1000 Connectome [http://www.nitrc.org/projects/fcon_1000/] has reportedly used group ICA methods for analyzing thousands of fMRI. However, the software posted does not show how to conduct group ICA on these images. We speculate that the team pooled results from many small-group ICA analyses, which is likely to be computationally expensive. PVD is a simple and very fast alternative that could inform future group ICA methods. Partial least squares regression (PLS-regression) is related to principal components regression (PCR) and, thus, regression using SVD decompositions. We are not yet focusing on the regression part of the problem and are interested in smoothing and decomposing the variability of populations of images.

A simple alternative to our two-stage SVD was suggested by the Associate Editor. Using notations in equation (2), the method would sum the $\boldsymbol{Y}_i\boldsymbol{Y}_i'$ and use the SVD of this sum to estimate $\boldsymbol{P}$ and then sum the $\boldsymbol{Y}_i'\boldsymbol{Y}_i$ matrices and use the SVD of this sum to estimate $\boldsymbol{D}$. This is a very simple and compelling idea that we have also considered. This is an excellent, and potentially faster, alternative to our default PVD procedure in the particular example we consider here. However, there are many reasons for using PVD. First, in many applications one of the dimensions is very large. For example in fMRI the number of voxels is in the millions and calculating and diagonalizing the space-by-space covariance matrix would be out of the question. Second, our method provides the subject-specific left and right eigenfunctions and opens up new analysis possibilities. For example, one could be interested in studying the variability of $\boldsymbol{U}_{L_i}$, the matrix containing the first L_i left eigenvectors of the data matrix $\boldsymbol{Y}_i$, around $\boldsymbol{P}$, the population level matrix of left eigenvectors. Third, our method is probably equally fast and requires only minimal additional coding. Fourth, both methods are reasonable ways of constructing the $\boldsymbol{P}$ and $\boldsymbol{D}$ matrices. Simply putting forward the PVD formula will lead to many ways of building $\boldsymbol{P}$ and $\boldsymbol{D}$.

A few open problems remain and will need to be addressed. First, theoretic and methodological approaches are needed to determine the cutoff dimension for the number of subject-specific eigenfrequencies and eigenvariates retained for the second stage of the analysis. While we use the same number of eigenfrequencies and eigenvectors it may make sense to keep a different number of bases in each dimension. Second, methods are needed to address the noise in images. The noise in the frequency by time plots is large and its size probably depends on the size of the signal. Conducting SVD of images with complex noise structure remains an open area of research. Third, investigating the optimality properties, or lack thereof, of our procedure is needed and may lead to better or faster procedures. Fourth, better visualization tools will need to be developed to address the data onslaught. In spite of our best efforts, we believe that better ways of presenting terrabytes, and soon petabytes, of data are needed. Fifth, better understanding of the geometry of images in very-high dimensional spaces is necessary.

Acknowledgement. Research supported by Award Number R01NS060910 from the National Institute Of Neurological Disorders And Stroke. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of Neurological Disorders And Stroke or the National Institutes of Health. The authors gratefully acknowledge the suggestions and comments of the Associate Editor and two anonymous reviewers. Any remaining errors are the sole responsibility of the authors.

References

[1] B. Caffo, C.M. Crainiceanu, G. Verduzco, S.H. Mostofsky, S. Spear-Bassett, and J.J. Pekar. Two-stage Decompositions for the Analysis of Functional Connectivity for fMRI With Application to Alzheimer's Disease Risk. *NeuroImage*, 10(3):1140–1149, 2009.

[2] V. D. Calhoun, T. Adali, G. D. Pearlson, and J.J. Pekar. A method for making group inferences from functional mri data using independent component analysis. *Human Brain Mapping*, 14(3):140–151, 2001.

[3] V. D. Calhoun, J. Liu, and T. Adali. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage*, 45(1):163–172, 2009.

[4] K. Cheshire, H. Engleman, I. Deary, C. Shapiro, and N. J. Douglas. Factors impairing daytime performance in patients with sleep apnea/hypopnea syndrome. *Archives of Internal Medicine*, 152:538–541, 1992.

[5] R. Christensen. *Advanced Linear Modeling, 2nd edition.* Springer, New York, 2001.

[6] P. Comon. Independent component analysis: a new concept? *Signal Processing*, 36(3):287–314, 1994.

[7] D. Cook. Fisher lecture: Dimension reduction in regression. *Statistical Science*, 22(1):1–26, 2007.

[8] C. M. Crainiceanu and A. J. Goldsmith. Bayesian Functional Data Analysis using WinBUGS. *Journal of Statistical Software*, 32(11), 2009.

[9] C.M. Crainiceanu, B. Caffo, C. Di, and N.M. Punjabi. Nonparametric signal extraction and measurement error in the analysis of electroencephalographic activity during sleep. *Journal of the American Statistical Association*, 486:541–555, 2009.

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

[11] C. Di, C.M. Crainiceanu, B.S. Caffo, and N.M. Punjabi. Multilevel functional principal component analysis. *Annals of Applied Statistics*, 3:458–488, 2009.

[12] C. Guilleminault, M. Partinen, M.A. Quera-Salva, B. Hayes, W.C. Dement, and G. Nino-Murcia. Determinants of daytime sleepiness in obstructive sleep apnea. *Chest*, 94:32–37, 1988.

[13] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley, New York, 2001.

[14] A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and application. *Neural Networks*, 13(4–5):411–430, 2000.

[15] I. T. Jollife. *Principal Component Analysis.* Springer, New York, 2002.

[16] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiæ Scientiarum Fennicæ, Series A1: Mathematica-Physica, Suomalainen Tiedeakatemia*, 37:3–79, 1947.

[17] R. N. Kingshott, H. M. Engleman, J. J. Deary, and N. J. Douglas. Does arousal frequency predict daytime function. *European Respiratory Journal*, 12:1264–1270, 1998.

[18] M. Loève. Functions Aleatoire de Second Ordre. *Comptes Rendus de l'Acadmie des Sciences*, 220, 1945.

[19] S. E. Martin, P. K. Wraith, J. J. Deary, and N. J. Douglas. The effect of nonvisible sleep fragmentation on daytime function. *American Journal of Respiratory and Critical Care Medicine*, 155:1596–1601, 1997.

[20] S.F. Quan, B.V. Howard, C. Iber, J.P. Kiley, F.J. Nieto, and G.T. O'Connor et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*, 20:1077–85, 1997.

[21] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer-Verlag, New York, 2006.

[22] S. Redline, M. H. Sanders, B. K. Lind, S. F. Quan, C. Iber, D. J. Gottlieb, W. H. Bonekat, D. M. Rapoport, P. L. Smith, and J. P. Kiley. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. sleep heart health research group. *Sleep*, 21(7):759767, 1998.

[23] G. A. F. Seber. *Multivariate Observations*. Wiley, New York, 1984.

[24] D. J. Spiegelhalter, A. Thomas, N. G. Best, and D. Lunn. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit, Cambridge, 2003.

[25] A.-M. Staicu, C. M. Crainiceanu, and R. J. Carroll. Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 2010.

[26] C. W. Whitney, D. J. Gottlieb, S. Redline, R. G. Norman, R. R. Dodge, E. Shahar, S. Surovec, and F. J. Nieto. Reliability of scoring respiratory disturbance indices and sleep staging. *Sleep*, 21(7):749–757, 1998.

[27] H. Wold. *Partial least squares, Pp. 581-591 in Samuel Kotz and Norman L. Johnson, eds., Encyclopedia of statistical sciences*. Wiley, New York, 1985.

[28] S. Wold, A. Ruhe, H. Wold, and W.J. Dunn. The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses. *Journal on Scientific and Statistical Computing*, 5:735–743, 1984.

[29] D. Woodard, C.M. Crainiceanu, and D. Ruppert. Population Level Hierarchical Adaptive Regression Kernels. *manuscript*.

[30] F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590, 2005.

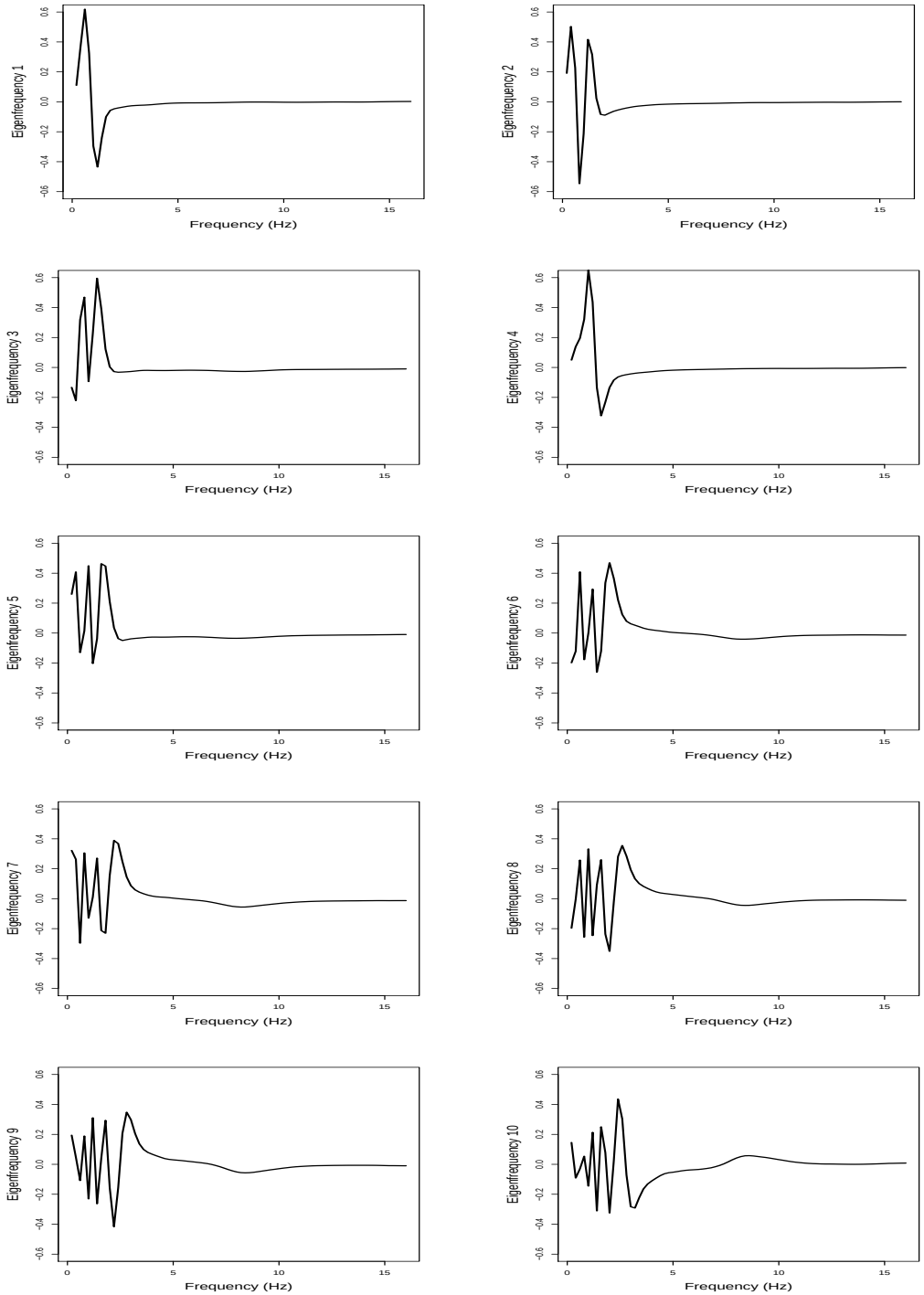

Figure 4: First 10 population level eigenfrequencies for the combined data from visit 1 and 2. The X-axis is frequency in Hz. Eigenfrequencies are truncated at 16Hz for plotting purposes, but they extend to 32Hz.

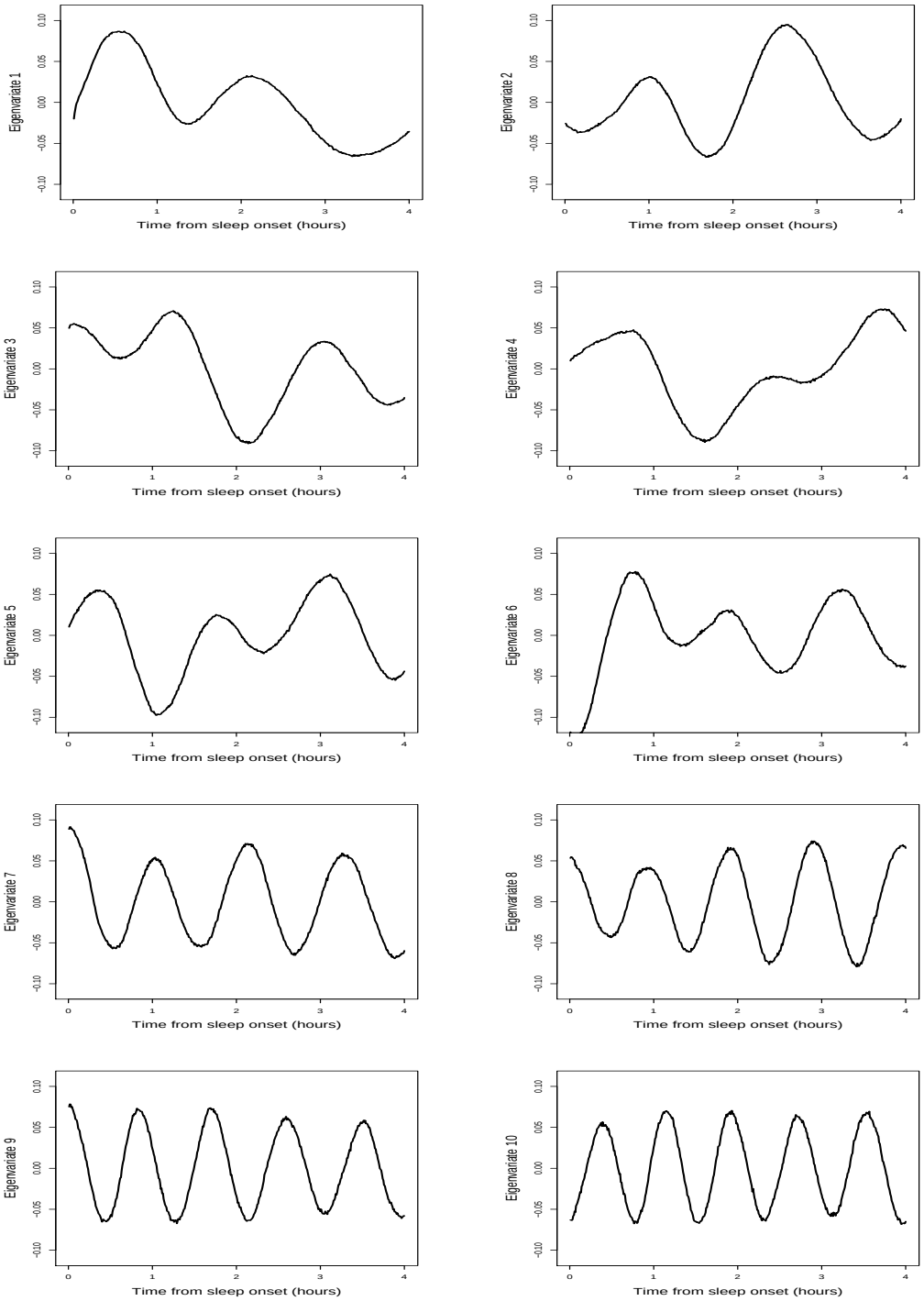

Figure 5: First 10 population level eigenvariates for the combined data from visit 1 and 2. The X-axis is time from sleep onset in hours.

Eigenfrequencies										
Comp.	1	5	6	7	8	9	10	11	12	13
λ ($\times 10^{-2}$)	9.95	9.58	9.38	8.80	8.19	6.73	4.45	2.37	1.92	1.69
sum % var	9.96	49.09	58.49	67.31	75.51	82.25	86.71	89.08	90.10	92.69
Eigenvariates										
Comp.	1	5	6	7	8	9	10	11	12	13
λ ($\times 10^{-2}$)	2.10	1.21	1.07	0.89	0.74	0.63	0.55	0.49	0.42	0.37
sum % var	12.65	47.67	54.13	59.50	63.94	67.75	71.09	74.04	76.57	78.82

Table 3: Variance and cumulated percent variance explained by population level eigenvalues from the observed variance of eigenvalues at the subject level. The labels eigenfrequencies and eigenvariates refer to the left and right eigenvectors, respectively. Population level eigenfrequencies are the eigenvectors in the $\mathbb{R}^F$ dimensional subspace spanned by the collection of the first 10 eigenfrequencies at the subject level across all subjects. Population level eigenvariaes are the eigenvectors in the $\mathbb{R}^T$ dimensional subspace spanned by the collection of the first 10 eigenvariates at the subject level across all subjects.

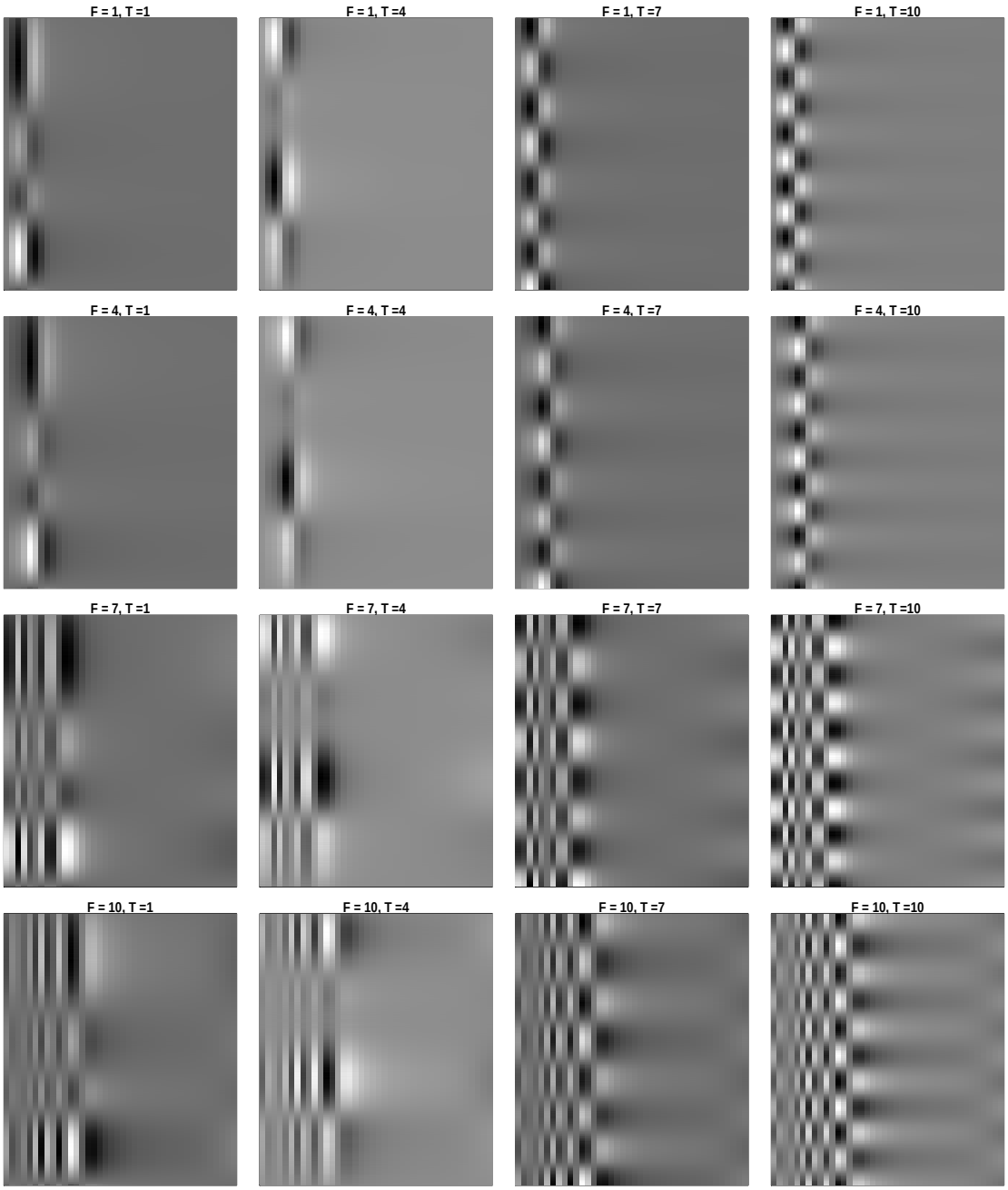

Figure 6: Some population level basis components obtained as Kronecker products of eigenfrequencies and eigenvariates. The x-axis are the Frequencies from 0.2 to 8Hz are on the x-axis time from sleep onset until the end of the 4th hour are on the y-axis. The title of each image indicates the eigenfrequency number (F) and eigenvariate number (T), as ordered by their corresponding eigenvalues. For example, $F = 1$, $T = 7$ indicates the basis component obtained as a Kronecker product of the 1st eigenfrequency and the 7th eigenvariate.

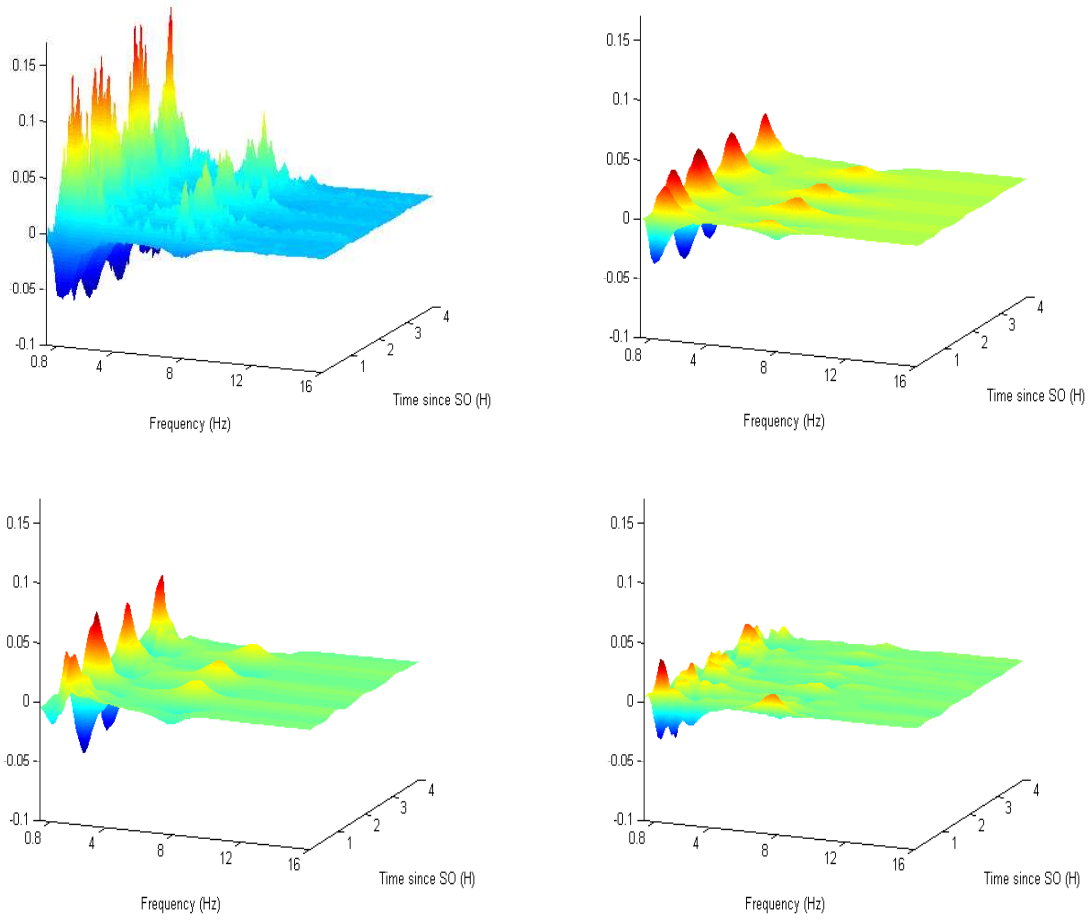

Figure 7: Image smoothing for one subject for the first 4 hours of sleep after sleep onset. Top left panel displays the normalized power up to 16Hz, even though the analysis is based on data up to 32Hz. Top right panel displays the smooth image obtained by projection on the first 15 eigenfrequencies and first 3 smoothed eigenvariates at the subject level; the other eigenvariates at the subject level are indistinguishable from white noise. Bottom left panel displays the smooth image obtained by projection on the first 15 eigenfrequencies and first 15 eigenvariates at the population level (some shown in Figure 3). Bottom right panel displays the difference between the subject-level smooth (top right panel) and population level smooth (bottom left panel).

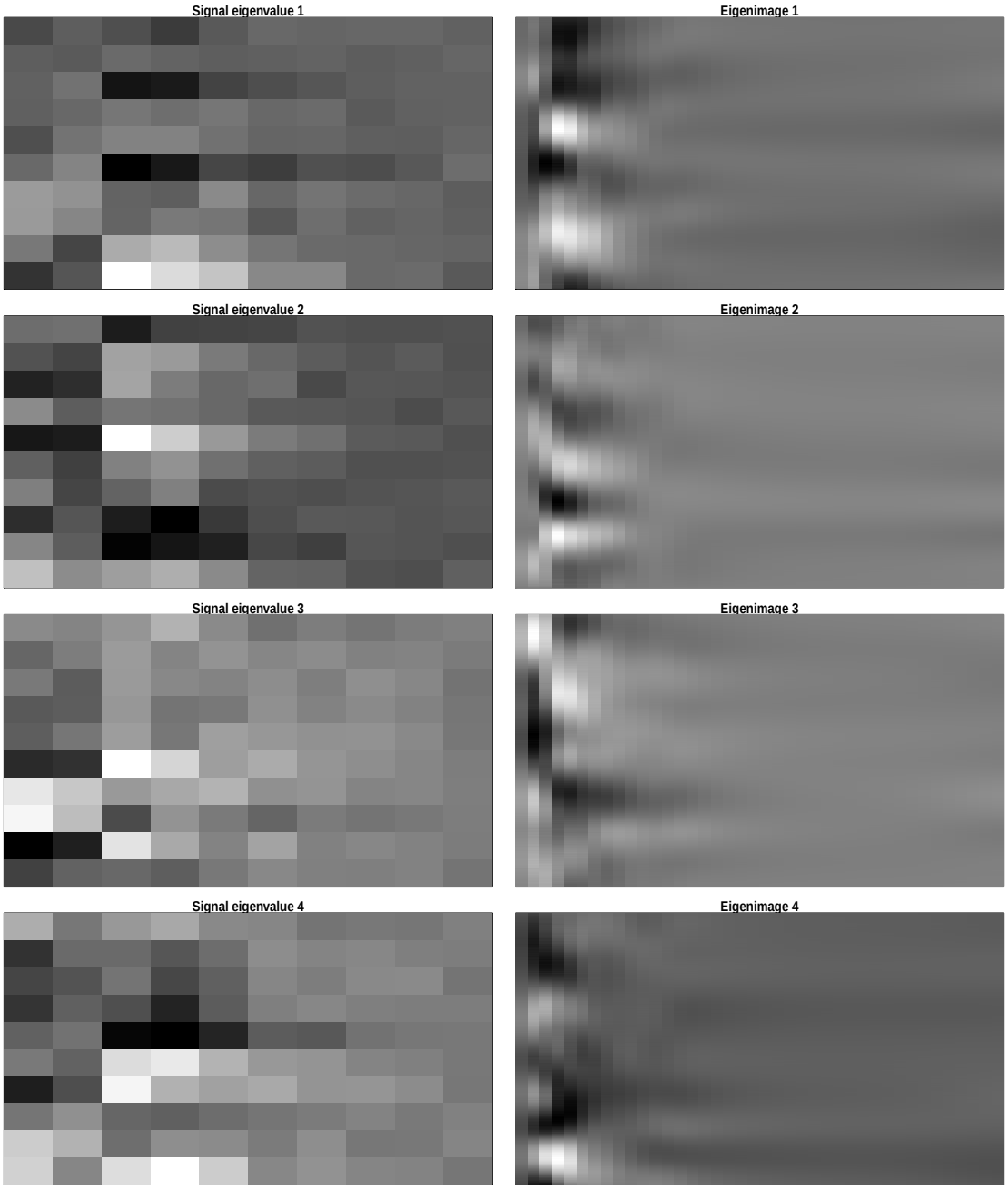

Figure 8: Left panels: first 4 subject-specific eigenimages, $\boldsymbol{\phi}_k^{(1)}$, of the multivariate process of image coefficients, $\boldsymbol{V}_{ij}$. Right panels: first 4 subject-specific eigenimages, $\boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D}$, of the image process, $\boldsymbol{Y}_{ij}$. Right panels are reconstructed from the left panels using the transformation $\boldsymbol{\phi}_k^{(1)} \rightarrow \boldsymbol{P}\boldsymbol{\phi}_k^{(1)}\boldsymbol{D}$ from the coefficient to the image space.