HHS Public Access

Author manuscript

Ann Appl Stat. Author manuscript; available in PMC 2019 March 01.

Published in final edited form as:

Ann Appl Stat. 2018 September ; 12(3): 1422–1450. doi:10.1214/17-AOAS1116.

TPRM: TENSOR PARTITION REGRESSION MODELS WITH APPLICATIONS IN IMAGING BIOMARKER DETECTION

Michelle F. Miranda[*,¶,‖], **Hongtu Zhu**[‡,¶,**], **Joseph G. Ibrahim**[†,**], and **for the Alzheimer's Disease Neuroimaging Initiative**[§]

[¶]University of Texas MD Anderson Cancer Center

[‖]Universidade de São Paulo

[**]University of North Carolina at Chapel Hill

Abstract

Medical imaging studies have collected high dimensional imaging data to identify imaging biomarkers for diagnosis, screening, and prognosis, among many others. These imaging data are often represented in the form of a multi-dimensional array, called a tensor. The aim of this paper is to develop a tensor partition regression modeling (TPRM) framework to establish a relationship between low-dimensional clinical outcomes (e.g., diagnosis) and high dimensional tensor covariates. Our TPRM is a hierarchical model and efficiently integrates four components: (i) a partition model, (ii) a canonical polyadic decomposition model, (iii) a principal components model, and (iv) a generalized linear model with a sparse inducing normal mixture prior. This framework not only reduces ultra-high dimensionality to a manageable level, resulting in efficient estimation, but also optimizes prediction accuracy in the search for informative subtensors. Posterior computation proceeds via an efficient Markov chain Monte Carlo algorithm. Simulation shows that TPRM outperforms several other competing methods. We apply TPRM to predict disease status (Alzheimer versus control) by using structural magnetic resonance imaging data obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

Keywords and phrases

Bayesian hierarchical model; Big data; MCMC; Tensor decomposition; Tensor regression

[*]Dr. Miranda's research was partially supported by grant 2013/ 07699-0 and 2014/07254-0, Sao Paulo Research Foundation, and grant CA-178744.

[†]Dr. Ibrahim's research was partially supported by NIH grants #GM 70335 and P01CA142538.

[‡]Dr. Zhu was partially supported by NIH grant MH086633, NSF Grants SES-1357666 and DMS-1407655, and a grant from Cancer Prevention Research Institute of Texas.

[§]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Michelle F. Miranda. University of Texas MD Anderson Cancer Center and Universidade de São Paulo, SP Brazil, michellemirandaest@gmail.com

Hongtu Zhu. Department of Biostatistics, University of Texas MD Anderson Cancer Center and University of North Carolina at Chapel Hill, hzhu5@mdanderson.org

Joseph. G. Ibrahim. Department of Biostatistics, University of North Carolina at Chapel Hill, ibrahim@bios.unc.edu

1. Introduction

Medical imaging studies have collected high dimensional imaging data (e.g., Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)) to extract information associated with the pathophysiology of various diseases. These information, or imaging biomarkers, could potentially aid detection and improve diagnosis, assessment of prognosis, prediction of response to treatment, and monitoring of disease status. Thus, efficient imaging biomarker extraction is crucial to the understanding of many disorders, including different types of cancer (e.g. lung cancer), and brain disorders such as Alzheimer's disease and autism, among many others.

A critical challenge is to convert medical images into clinically useful information that can facilitate better clinical decision making (Gillies et al., 2016). Existing statistical methods are not always efficient for such conversion due to the high-dimensionality of array images as well as their complex structure, such as spatial smoothness, correlation, and heterogeneity. Although a large family of regression methods has been developed for supervised learning of a scalar response (e.g. clinical outcome) (Hastie et al., 2009; Breiman et al., 1984; Friedman, 1991; Zhang and Singer, 2010), their computability and theoretical guarantee are compromised by the ultra-high dimensionality of the imaging data covariates. To address this challenge, many modeling strategies have been proposed to establish association between high-dimensional array covariates and scalar response variables.

The first set of promising solutions is the high-dimensional sparse regression (HSR) models, which often take high-dimensional imaging data as unstructured predictors. A key assumption of HSR is its sparse solutions. HSRs not only suffer from diverging spectra and noise accumulation in ultra-high dimensional feature space (Fan and Fan, 2008; Bickel and Levina, 2004), but also their sparse solutions may lack clinically meaningful information. Moreover, standard HSRs ignore the inherent spatial structure of medical image, such as spatial correlation and spatial smoothness. To address some limitations of HSRs, a family of tensor regression models has been developed to preserve the tensor structure of imaging data, while achieving substantial dimension reduction (Zhou et al., 2013).

The second set of solutions adopts functional linear regression (FLR) approaches, which treat imaging data as functional predictors. However, since most existing FLR models focus on one-dimensional curves (Müller and Yao, 2008; Ramsay and Silverman, 2005), generalizations to two and higher dimensional images is far from trivial and requires substantial research (Reiss and Ogden, 2010). Most estimation approaches of FLR approximate the coefficient function as a linear combination of a set of fixed (or data-driven) basis functions. For instance, most estimation methods of FLR based on the fixed basis functions (e.g., tensor product wavelet) are required to solve an ultra-high dimensional optimization problem and can suffer from the same limitations as those of HSR.

The third set of solutions usually integrates supervised (or unsupervised) dimension reduction techniques with various standard regression models. Given the high dimension of imaging data, it is imperative to use some dimension reduction methods to extract and select important 'low-dimensional' features, while eliminating most noises (Johnstone and Lu,

2009; Bair et al., 2006; Fan and Fan, 2008; Tibshirani et al., 2002; Krishnan et al., 2011). Most of these methods first carry out an unsupervised dimension reduction step, often by principal component analysis (PCA), and then fit a regression model based on the top principal components (Caffo et al., 2010). Recently, for ultra-high tensor data, unsupervised higher order tensor decompositions (e.g. parallel factor analysis and Tucker) have been extensively proposed to extract important information of neuroimaging data (Martinez et al., 2004; Beckmann and Smith, 2005; Zhou et al., 2013). These methods are intuitive and easy to implement, but features extracted from PCA and tensor decomposition can miss small and localized information that is relevant to the response. We propose a novel model that efficiently extracts these information, while performing dimension reduction and feature selection for better prediction accuracy.

The aim of this paper is to develop a novel modeling framework to extract imaging biomarkers from high-dimensional imaging data, denoted by **x**, to predict a scalar response, denoted by y. The scalar response y may include cognitive outcome, disease status, and the early onset of disease, among others. The imaging data provided by neuroimaging studies is often represented in the form of a multi-dimensional array, called a tensor. We develop a novel Tensor Partition Regression Model (TPRM) to establish an association between imaging tensor predictors and clinical outcomes. Our TPRM is a hierarchical model with four components, including (i) a partition model that divides the high-dimensional tensor covariates into sub-tensor covariates; (ii) a canonical polyadic decomposition model that reduces the sub-tensor covariates to low-dimensional feature vectors; (iii) a projection of these feature vectors into the space of the principal components, and (iv) a generalized linear model with a sparse inducing normal mixture prior that is used to select informative feature vectors for predicting clinical outcomes. Although the four components of TPRM have been independently developed, the key novelty of TPRM lies in the integration of (i)–(iv) into a single framework for imaging prediction. In particular, the first two components (i) and (ii) are designed to specifically address the three key features of neuroimaging data, including relatively low signal to noise ratio, spatially clustered effect regions, and the tensor structure of imaging data.

In Section 2, we introduce TPRM, the priors, and a Bayesian estimation procedure. In Section 3, we use simulated data to compare the Bayesian decomposition with several competing methods. In Section 4, we apply our model to the ADNI data set. This data set consists of 181 subjects with Alzheimer's disease and 221 controls and the correspondent covariates are MRI images of size 96 × 96 × 96. In Section 5, we present some concluding remarks.

2. Methodology

2.1. Preliminaries

We review a few basic facts about tensors (Kolda and Bader, 2009). A **tensor** $\mathbf{x} = (x_{j_1 \ldots j_D}) \in \mathbb{R}^{J_1 \times \ldots \times J_D}$ is a multidimensional array, whose order D is determined by its dimension. For instance, a vector is a tensor of order 1 and a matrix is a tensor of order 2. The **inner**

product between two tensors $\mathcal{X} = (x_{j1\ldots jD})$ and $\mathcal{X}' = (x'_{j_1\ldots j_D})$ in $\mathbb{R}^{J1\times\ldots\times JD}$ is the sum of the product of their entries given by

$$\langle \mathcal{X}, \mathcal{X}' \rangle = \sum_{j_1=1}^{J_1} \cdots \sum_{j_D=1}^{J_D} x_{j_1\ldots j_D} x'_{j_1\ldots j_D}.$$

The **outer product** between two vectors $\boldsymbol{a}^{(1)} = (a^{(1)}_{j_1}) \in \mathbb{R}^{J_1}$ and $\boldsymbol{a}^{(2)} = (a^{(2)}_{j_1}) \in \mathbb{R}^{J_2}$ is a matrix $M = (m_{j_1 j_2})$ of size $J_1 \times J_2$ with entries $m_{j_1 j_2} = a^{(1)}_{j_1} a^{(2)}_{j_2}$. A tensor $\mathcal{X} \in \mathbb{R}^{J_1\times\ldots\times J_D}$ is a *rank one tensor* if it can be written as an outer product of D vectors such that $\mathcal{X} = \boldsymbol{a}^{(1)} \circ \boldsymbol{a}^{(2)} \ldots \circ \boldsymbol{a}^{(D)}$, where $\boldsymbol{a}^{(k)} \in \mathbb{R}^{J_k}$ for $k = 1, \ldots, D$. Moreover, the canonical polyadic decompositic (**CP decomposition**), also known as parallel factor analysis (PARAFAC), factorizes a tensor into a sum of rank-one tensors such that

$$\mathcal{X} = \sum_{r=1}^{R} \boldsymbol{a}^{(1)}_r \circ \boldsymbol{a}^{(2)}_r \circ \ldots \circ \boldsymbol{a}^{(D)}_r,$$

where $\boldsymbol{a}^{(k)}_r = (a^{(k)}_{j_k r}) \in \mathbb{R}^{J_k}$ for $k = 1, \ldots, D$ and $r = 1, \ldots, R$. See Figure 1 for an illustration of a 3D array.

It is convenient and assumed in this paper that the columns of the factor matrices are normalized to length one with weights absorbed into a diagonal matrix $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_R)$ such that

$$\mathcal{X} = \sum_{r=1}^{R} \lambda_r \boldsymbol{a}^{(1)}_r \circ \boldsymbol{a}^{(2)}_r \circ \ldots \circ \boldsymbol{a}^{(D)}_r \equiv \|\boldsymbol{\Lambda}; \boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(D)}\|, \quad (2.1)$$

where $\boldsymbol{A}^{(d)} = [\boldsymbol{a}^{(d)}_1 \boldsymbol{a}^{(d)}_2 \ldots \boldsymbol{a}^{(d)}_R]$ for $d = 1, \ldots, D$.

It is sometimes convenient to arrange the tensor $\mathcal{X}$ as a matrix. This arrangement can be done in various ways but we will rely on the following definition detailed in Kolda and Bader (2009). We define the **mode-d matricized** version of $\mathcal{X}$ as

$$\boldsymbol{X}_{(d)} = \boldsymbol{A}^{(d)} \boldsymbol{\Lambda} (\boldsymbol{A}^{(D)} \odot \ldots \odot \boldsymbol{A}^{(d+1)} \odot \boldsymbol{A}^{(d-1)} \odot \ldots \odot \boldsymbol{A}^{(1)})^T,$$

where $\odot$ denotes the Khatri–Rao product. Then, we can write the factor matrix corresponding to the dimension d as a projection of $\boldsymbol{X}_{(d)}$ in the following way

$$\boldsymbol{A}^{(d)} = \boldsymbol{X}_{(d)}(\boldsymbol{A}^{(D)} \odot \ldots \odot \boldsymbol{A}^{(d+1)} \odot \boldsymbol{A}^{(d-1)} \odot \ldots \odot \boldsymbol{A}^{(1)})\boldsymbol{V}^{\dagger}\boldsymbol{\Lambda}^{-1}, \quad (2.2)$$

where $\boldsymbol{V}^{\dagger}$ is the Moore-Penrose inverse of

$$\boldsymbol{V} = \boldsymbol{A}^{(1)T}\boldsymbol{A}^{(1)} * \ldots * \boldsymbol{A}^{(d-1)T}\boldsymbol{A}^{(d-1)} * \boldsymbol{A}^{(d+1)T}\boldsymbol{A}^{(d+1)} * \ldots * \boldsymbol{A}^{(D)T}\boldsymbol{A}(D),$$

in which * indicates the Hadamard product of matrices (Kolda and Bader, 2009; Kolda, 2006).

We need the following notation throughout the paper. Suppose that we observe data $\{(y_i, \mathcal{X}_i, \boldsymbol{z}_i) : i = 1, \ldots, N\}$ from N subjects, where the $\mathcal{X}_i$'s are tensor imaging data, $\boldsymbol{z}_i$ is a $p_Z \times 1$ vector of scalar covariates, and y_i is a scalar response, such as diagnostic status or clinical outcome. In the ADNI example, $N = 402$ and $y_i = 1$ if subject i is a patient with Alzheimer's disease and $y_i = 0$ otherwise. If we concatenate all D-dimensional tensor $\mathcal{X}_i$'s into a $(D+1)$-dimensional tensor $\tilde{\mathcal{X}} = \{\mathcal{X}_i, i = 1, \ldots, N\} = (x_{j1,\ldots,jD,i})$, then we consider the CP decomposition of $\tilde{\mathcal{X}}$ as follows:

$$\tilde{\mathcal{X}} = \|\boldsymbol{\Lambda}; \boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(D)}, \boldsymbol{L}\| \text{ or } x_{j_1,\ldots,j_D,i} = \sum_{r=1}^{R} \lambda_r a^{(1)}_{j_1 r} a^{(2)}_{j_2 r} \ldots a^{(D)}_{j_D r} l_{ir}, \quad (2.3)$$

where $\boldsymbol{L} = (l_{ir})$ is an $N \times R$ matrix. The matrices $\boldsymbol{A}^{(d)}$'s and $\boldsymbol{L}$ are the factor matrices. In this paper, we introduce the notation $\boldsymbol{L}$ in order to differentiate between matrices that carry common features among subjects ($\boldsymbol{A}^{(d)}$'s) and the matrix $\boldsymbol{L}$, that is subject specific.

2.2. Tensor Partition Regression Models

Our interest is to develop TPRM for establishing the association between responses y_i and their corresponding imaging covariate $\mathcal{X}_i$ and clinical covariates $\boldsymbol{z}_i$. The first component of TPRM is a partition model that divides the high-dimensional tensor $\mathcal{X}_i \in \mathbb{R}^{J_1 \times \ldots \times J_D}$ into S disjoint sub-tensor covariates $\mathcal{X}_i^{(s)} \in \mathcal{R}^{p_1 \times \ldots \times p_D}$ for $s = 1, \ldots, S$. Although the size of $\mathcal{X}_i^{(s)}$ can vary across s, it is assumed that, without loss of generality, $\mathcal{X}_i^{(s)}$ and the size of $\mathcal{X}_i^{(s)}$ is homogeneous such that $S = \prod_{k=1}^{D}(J_k/p_k)$. We defined the partitions as follows:

$$\mathcal{X}_i^{(s)} = \{x_{j_1, j_2, \ldots, j_D i} : j_d \in I_d^{(s_d)}, d = 1, \ldots, D\}, \quad (2.4)$$

$$s = s_1 + \sum_{d=2}^{D}(s_d - 1)\{\prod_{k=1}^{d-1}(J_k/p_k)\},$$

$$1 \le s_d \le J_d/p_d, I_d^{(s_d)} = \{(s_d - 1)p_d + 1, (s_d - 1)p_d + 2, \ldots, s_d p_d\},$$

$$\bigcup_{s_d=1}^{J_d/p_d} I_d^{(s_d)} = I_d = \{1, 2, \ldots, J_d\} \text{ and } I_d^{(s_d)} \cap I_d^{(s'_d)} = \varnothing \text{ for } s_d \neq s'_d.$$

These sub-tensors $\mathcal{X}_i^{(s)}$'s are cubes of neighboring voxels that do not overlap and collectively form the entire 3D image. Figure 2 presents a three-dimensional tensor with sub-tensors.

The second component of TPRM is a canonical polyadic decomposition model that reduces the sub-tensor covariates $\tilde{\mathcal{X}}^{(s)} = (\mathcal{X}_i^{(s)})$ to low-dimensional feature vectors. Specifically, it is assumed that for each s, we have

$$\tilde{\mathcal{X}}^{(s)} = \|\mathbf{\Lambda}_s; \boldsymbol{A}_s^{(1)}, \boldsymbol{A}_s^{(2)}, \ldots, \boldsymbol{A}_s^{(D)}, \boldsymbol{L}_s\| + \mathcal{E}^{(s)}, \quad (2.5)$$

where $\mathbf{\Lambda}_s = \text{diag}(\lambda_1^{(s)}, \ldots, \lambda_R^{(s)})$ consists of the weights for each rank of the decomposition in (2.5), $\boldsymbol{A}_s^{(d)} = (A_{s1}^{(d)} \cdots A_{sr}^{(d)}) \in \mathbb{R}^{p_d \times R}$ is the factor matrix along the d-th dimension of $\tilde{\mathcal{X}}^{(s)}$, and $\boldsymbol{L}_s \in \mathbb{R}^{N \times R}$ is the factor matrix along the subject dimension. The error term $\mathcal{E}^{(s)}$ is usually specified in order to find a set of $\boldsymbol{A}_s^{(d)}$'s and $\boldsymbol{L}_s$ that best approximates $\tilde{\mathcal{X}}^{(s)}$ (Kolda and Bader, 2009). We assume that the elements of $\mathcal{E}^{(s)} = (e_{j_1 \cdots j_D i}^{(s)})$ are measurement errors and $e_{j_1 \cdots j_D i}^{(s)} \sim N(0, (\tau^{(s)})^{-1})$.

The elements of $\boldsymbol{L}_s$ capture local imaging features in $\mathcal{X}^{(s)}$ across subjects, while the factor matrix $\boldsymbol{A}_s^{(d)}$ represents the common structure of all subjects in the d–th dimension for $d = 1, \ldots, D$ (Kolda and Bader, 2009). In our ADNI analysis, we have $D = 3$ and $\boldsymbol{A}^{(1)}$, $\boldsymbol{A}^{(2)}$, and $\boldsymbol{A}^{(3)}$ contain the vectors associated with the common features of the images along the coronal, saggital, and axial planes, respectively.

The use of (2.4) and (2.5) has two key advantages. First, the partition model (2.4) allows us to concentrate on the most important local features of each sub-tensor, instead of the major variation of the whole image, which may be unassociated with the response of interest. In many applications, although the effect regions (e.g. tumor) associated with responses (e.g. breast cancer) may be relatively small compared with the whole image, their size can be comparable with that of each sub-tensor. Therefore, one can extract more informative features associated with the response with a higher probability. Second, the canonical polyadic decomposition model (2.5) can substantially reduce the dimension of the original imaging data. For instance, consider a standard 256 × 256 × 256 3D array with 16,777,216

voxels, and its partition model with $32^3 = 32,768$ sub-arrays of size $8 \times 8 \times 8$. If we reduce each $8 \times 8 \times 8$ into a small number of components by using component (ii), then the total number of reduced features is around $O(10^4)$. We can further increase the size of each subarray in order to reduce the size of neuroimaging data to a manageable level, resulting in efficient estimation.

The third component of TPRM is a projection of $\boldsymbol{L} = [\boldsymbol{L}_1, \ldots, \boldsymbol{L}_S] \in \mathbb{R}^{N \times P_L}$ ($P_L = S \times R$) into the space spanned by the eigenvectors of $\boldsymbol{L}$. The i–th row of $\boldsymbol{L}$, $\boldsymbol{l}_i$ represents the vector of local image features across all partitions. It is assumed that

$$\boldsymbol{G} = \boldsymbol{L}\boldsymbol{D}^T, \quad (2.6)$$

where each row of $\boldsymbol{G}$ is a $1 \times K$ vector of common unobserved (latent) factors $\boldsymbol{g}_i$ and $\boldsymbol{D} \in \mathbb{R}^{K \times P_L}$ corresponds to the matrix of K basis functions used to represent $\boldsymbol{L}$. Notice that $\boldsymbol{D}$ is the intrinsic low-dimensional space spanned by all vectors of local image features and, therefore, $\boldsymbol{G}$ is the projection of $\boldsymbol{L}$ onto $\boldsymbol{D}$.

The number of latent basis functions K can be chosen by determining the percentage of data variability in oder to represent $\boldsymbol{L}$ in the basis space. The proposed basis representation has two purposes, including (i) reducing the feature matrix by selecting a small number of basis K and (ii) treating the multicolinearity induced by adjacent partitions in $\boldsymbol{L}$.

The fourth component of TPRM is a generalized linear model that links scalar responses y_i and their corresponding reduced imaging features $\boldsymbol{g}_i$ and clinical covariates $\boldsymbol{z}_i$. Specifically, y_i given $\boldsymbol{g}_i$ and $\boldsymbol{z}_i$ follows an exponential family distribution with density given by

$$f(y_i|\boldsymbol{\theta}_i) = m(y_i) \exp\{\eta(\boldsymbol{\theta}_i)T(y_i) - a(\boldsymbol{\theta}_i)\}, \quad (2.7)$$

where $m(\cdot)$, $\eta(\cdot)$, $T(\cdot)$, and $a(\cdot)$ are pre-specified functions. Moreover, it is assumed that $\mu_i = E(y_i|\boldsymbol{g}_i, \boldsymbol{z}_i)$ satisfies

$$h(\mu_i) = z_i^T\boldsymbol{\gamma} + \boldsymbol{g}_i^T\boldsymbol{b}, \quad (2.8)$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{b} = (b_k)$ are coefficient vectors associated with $\boldsymbol{z}_i$ and $\boldsymbol{g}_i$, respectively, and $h(\cdot)$ is a link function.

2.3. Prior Distributions

We consider the priors on the elements of $\boldsymbol{b}$ by assuming a bimodal sparsity promoting prior (Mayrink and Lucas, 2013; George and McCulloch, 1993, 1997) and the following hierarchy:

$$b_k|\delta_k, \sigma^2 \sim (1 - \delta_k)F(b_k) + \delta_k \mathrm{N}(0, \sigma^2), \quad (2.9)$$

$$\delta_k|\pi \sim \text{Bernoulli}(\pi) \text{ and } \pi \sim \text{Beta}(\alpha_{0\pi}, \alpha_{1\pi}),$$

where $F(\cdot)$ is a pre-specified probability distribution and $a_{0\pi}$ and $a_{1\pi}$ are pre-specified contants. If $F(\cdot)$ is a degenerate distribution at 0, then we have the *spike and slab* prior (Mitchell and Beauchamp, 1988). A different approach is to consider $F = \text{N}(0, \varepsilon)$ with a very small $\varepsilon > 0$ (Ročková and George, 2014). In this case, the hyperparameter σ^2 should be large enough to give support to values of the coefficients that are substantively different from 0, but not so large that unrealistic values of b_k are supported. In this article, we opt for the latter approach.

The probability $\boldsymbol{\pi}$ determines whether a particular component of $\boldsymbol{g}_j$ is informative for predicting y_i. A common choice for its hyperparameters is $a_{0\pi} = a_{1\pi} = 1$. However, with this choice, the posterior mean of $\boldsymbol{\pi}$ is restricted to the interval [1/3, 2/3], an undesirable feature in variable selection. The 'bathtub' shaped beta distribution with $a_{0\pi} = a_{1\pi} = 0.5$ concentrates most of its mass in the extremes of the interval (0, 1) being more suitable for variable selection (Gonçalves et al., 2013).

It is assumed that $\boldsymbol{\gamma} \sim \text{N}(\boldsymbol{\gamma}^*, \upsilon^{-1}\boldsymbol{I}_q)$ and $\upsilon \sim \text{Gamma}(\nu_{0\upsilon}, \nu_{1\upsilon})$, where $\boldsymbol{\gamma}^*$ is a pre-specified vector and $\nu_{0\upsilon}$, and $\nu_{1\upsilon}$ are pre-specified constants.

If a Bayesian model for the decomposition (2.5) is selected, we consider the priors on the elements of $A_{sr}^{(d)}, l_r^{(s)}$, and $\boldsymbol{\tau}^{(s)}$. For $d = 1, \ldots, D$ and $r = 1, \ldots, R$, we assume

$$\boldsymbol{A}_{sr}^{(d)} \sim \text{N}(0, p_d^{-1}\boldsymbol{I}_{p_d}), \boldsymbol{l}_r^{(s)} \sim \text{N}(0, (\tau^{(s)})^{-1}\boldsymbol{I}_N), \text{ and } \tau^{(s)} \sim \text{Gamma}(\nu_{0\tau}, \nu_{1\tau}),$$

where $\boldsymbol{I}_N$ is an $N \times N$ identity matrix and $\nu_{0\tau}$ and $\nu_{1\tau}$ are pre-specified constants. When p_d is large, the columns of the factor matrix $A_{sr}^{(d)}$ are approximately orthogonal, which is consistent with their role in the decomposition (2.1) (Ding et al., 2011). However, we do not explicitly require orthonormality, which leads to substantial computational efficiency.

2.4. Posterior Inference

Let $\boldsymbol{\theta} = \{\boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \upsilon\}$. A Gibbs sampler algorithm is used to generate a sequence of random observations from the joint posterior distribution given by

$$p(\boldsymbol{\theta}|\mathcal{X}, \boldsymbol{y}) \propto p(\boldsymbol{y}|z, \boldsymbol{g}, \boldsymbol{\theta})p(\boldsymbol{b}|\boldsymbol{\delta})p(\boldsymbol{\delta}|\pi)p(\pi)p(\boldsymbol{\gamma}|\upsilon)p(\upsilon). \quad (2.10)$$

The Gibbs sampler essentially involves sampling from a series of conditional distributions, while each of the modeling components is updated in turn. If the Bayesian model is considered for the tensor decomposition in Equation (2.5), then $\boldsymbol{\theta} = \{\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(D)}, \boldsymbol{L}, \boldsymbol{b}, \boldsymbol{\delta}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \upsilon\}$, where $\boldsymbol{\tau} = [\boldsymbol{\tau}^{(1)}, \ldots, \boldsymbol{\tau}^{(S)}]$. Also, we include $p(\mathcal{X}|\boldsymbol{A}^{(1)} \ldots, \boldsymbol{A}^{(D)}, \boldsymbol{L}, \boldsymbol{\tau})$ to the right hand side of (2.10). The detailed sampling algorithm is described in Appendix B.

3. Simulation Studies

We carried out simulation studies to examine the finite-sample performance of TPRM and its associated Gibbs sampler. The first study aims at comparing the Bayesian tensor decomposition method with the alternating least squares and to assess the importance of the partition model in the reconstruction of real image. The results are shown in Table 3 of Appendix A, indicating that the Bayesian estimation for the tensor components improves the reconstruction error. However, an important issue associated with using the Bayesian estimation for (2.5) is its computational burden. For a single 3-dimensional image, running one iteration of the MCMC steps (a.1)–(a.4) for a partition of size $33 \times 33 \times 35$ takes 0.72 seconds on a Macintosh OS X, processor 1.4GHz Intel Core i5, memory 8Gb 1600MHz DDR3. However, when we introduce multiple subjects, as in the examples of the next simulation section and as in the real data application, the computational time increases to 16 seconds per iteration even for a single partition. Thus, fitting a full Bayesian TPRM to multiple data sets may become computationally infeasible. Instead, we calculate the ALS estimates of $\boldsymbol{L}_s$ and then apply MCMC to the fourth component (2.8) of TPRM. This approach is computationally much more efficient than the full Bayesian TPRM.

3.1. A three-dimensional (3D) simulation study

The goal of this set of simulations is to examine the classification performance of the partition model in the 3D imaging setting. We compare three feature extraction methods including (i) functional principal component model (fPCA); (ii) tensor alternating least squares (TALS); and (iii) our TPRM. Let $\mathcal{X}_i \in \mathbb{R}^{64\times64\times50}$ be the image covariate for subject i as defined in Section 2.1. We simulated $\mathcal{X}_i$'s as follows:

$$\mathcal{X}_i(y_i) = \mathcal{G}_0 + y_i\mathcal{X}_0 + \mathcal{E}_i \text{ for } i = 1, \ldots, 200,$$

where $\mathcal{G}_0 \in \mathbb{R}^{64\times64\times50}$ is a fixed brain template with values ranging from 0 to 250, the elements of the tensor $\mathcal{E}_i \in \mathbb{R}^{64\times64\times50}$ are a noise term, and $\mathcal{X}_0$ is the true signal image. Moreover, $\mathcal{X}_0$ is the true signal image and was generated according to the following different scenarios.

- **(S.1)** $\mathcal{X}_0$ is composed by two spheres of radius equal to 4 (in voxels) and the signal decays as it gets farther from their centers;
- **(S.2)** $\mathcal{X}_0$ is a sphere of radius equal to 4 (in voxels) and the signal decays as it gets farther from the center of the sphere;
- **(S.3)** $\mathcal{X}_0 = \|50 \times \boldsymbol{I}_4; \boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}\|$, where $\boldsymbol{A}_0^{(1)} \in \mathbb{R}^{64\times4}, \boldsymbol{A}_0^{(2)} \in \mathbb{R}^{64\times4}$, and $\boldsymbol{A}_0^{(3)} \in \mathbb{R}^{50\times4}$, and $\boldsymbol{A}_0^{(d)}$s' are matrices whose $(c_d + j)$-th element of each column is equal to $\sin(j\pi/14)$ with c_d indicating the position at the d-th coordinate;
- **(S.4)** $\mathcal{X}_0 = \|65 \times \boldsymbol{I}_4; \boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}\|$, where $\boldsymbol{A}_0^{(d)}$'s are the same as those in (S.3).

(S.5) – (S.8) $\mathcal{X}_0$ is equivalent to scenarios (S.1) – (S.4) except that the elements of $\mathcal{E}_i$ are generated from the short range spacial dependency as described in the first paragraph of this section.

For scenarios (S.1) – (S.4), the elements of $\mathcal{E}_i$ were independently generated from a N(0, 70^2) generator. For scenarios (S.5) – (S.8), the elements of $\mathcal{E}_i = (\mathcal{E}_i(g))$ were generated to reflect a short range spatial dependency. Specifically, let $\mathcal{E}_i(g) = \sum_{\|g' - g\|_1 \leq 1} E_i^*(g')/m_g$, where g is a voxel in the three-dimensional space, $E_i^*(g) \sim N(0, 70^2)$, $\|.\|_1$ is the L_1 norm of a vector, and m_g is the number of locations in the set $\{\|g' - g\|_1 \leq 1\}$. Figure 3 shows the 3D rendering of $\mathcal{X}_0$ overlaid on the template $\mathcal{G}_0$.

We consider a specific choice of parameters by setting $R = K = 20$ and $S = 32$ partitions. Since the signals in $\mathcal{X}_0$ are simple geometric forms, 20 basis may be a reasonable choice. In addition, we use the same number of features for all models to ensure their comparability. The code for this simulation study is included in the Supplementary Material E.

With these choices being made, we consider the following criteria. First, we generate the data as described in scenarios (S.1)–(S.8) and split the 200 pairs $(y_i, \mathcal{X}_i)$ into 180 as training samples and 20 as test samples. We perform this splitting 10 times in a 10-fold cross validation procedure. For each combination of training and test set, we use the training set to fit FPCA, TALS, and TPRM. The TPRM model is fitted by running an MCMC algorithm with 10,000 iterations with a burn-in of 5,000. The prediction accuracy, the false positive rate and the false negative rate are then computed for each test set. These measurements are the average values across the ten folds for each model under each scenario. The prediction accuracy (10-fold Accuracy) is the average of the prediction accuracy evaluated at the testing set. Results for each scenario and each fold are presented in Tables 5 and 6 of Appendix E.

Next, we generate 200 pairs $(y_i, \mathcal{X}_i)$, randomly separate them into 180 training samples and 20 test samples, and repeat it 100 times. For each run from 1 to 100, we use the training set to fit the models and calculate the prediction accuracy based on the test set. Monte Carlo Accuracy is the average across all these runs.

Table 1 shows the average measurements across multiple runs and also across the ten folds for each model under each scenario. For all scenarios, TPRM outperforms FPCA and TALS with higher prediction accuracy and smaller FPR and FNR (an exception is the FPR rate for FPCA, since the model is wrongly classifying everyone as positive). For (S.3), the three models are almost equivalent; the prediction accuracy and FNR slightly favor TPRM, but FPR alone favors TALS.

4. Real data analysis

“Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging

(MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org."[1]

We applied the proposed model to the anatomical MRI data collected at the baseline of ADNI. We considered 402 MRI scans from ADNI1, 181 of them were diagnosed with AD ($y_i = 1$), and 221 healthy controls ($y_i = 0$). These scans were performed on a 1.5T MRI scanners using a sagittal MPRAGE sequence and the typical protocol includes the following parameters: repetition time (TR) = 2400 ms, inversion time (TI) = 1000 ms, ip angle = 8°, and field of view (FOV) = 24 cm with a 256 × 256 × 170 mm^3 acquisition matrix in the x, y, and z dimensions, which yields a voxel size of 1.25 × 1.26 × 1.2 mm^3 (Huang et al., 2015).

The T1-weighted images were processed using the Hierarchical Attribute Matching Mechanism for Elastic Registration (HAMMER) pipeline. The processing steps include anterior commissure and posterior commissure correction, skull-stripping, cerebellum removal, intensity inhomogeneity correction, and segmentation. Then, registration was performed to warp the subject to the space of the Jacob template (size 256 × 256 × 256 mm^3). Finally, we used the deformation field to compute the RAVENS maps. The RAVENS methodology precisely quantifies the volume of tissue in each region of the brain. The process is based on a volume-preserving spatial transformation that ensures that no volumetric information is lost during the process of spatial normalization (Davatzikos et al., 2001).

4.1. Functional principal component

Following the pre-processing steps, we downsampled the images, cropped them, and obtained images of size 96 × 96 × 96 mm^3. The simple solution is to consider a classification model, with the response Y being the diagnostics status as described in the previous section, and the design matrix of size $N \times 884,736 (96^3)$. Here, each column of the design matrix is a location in the 3–D voxel space. Due to the high-dimensionality of the design matrix, we need to consider a dimension reduction approach before fitting a classification model. We consider three classifiers: a classification tree, a support vector machine (SVM) classifier, and a regularized logistic regression with lasso penalty. To evaluate the finite sample performance of the models, we performed a 10-fold cross validation procedure. For each combination of training and test set, we use the training set to extract the first M principal components, with M selected to represent 99% of the data variability. Next, we use principal components as predictors to fit the models. Then, we evaluate the prediction accuracy on the test set for each data split. The average prediction accuracies across the 10 split sets are 0.6467, 0.5818, and 0.5696 for the tree model, SVM and regularized logistic, respectively. The FPCA approach used here is equivalent to selecting the smallest partition possible (size 1 × 1 × 1). In this case, our feature matrix $\boldsymbol{L}$ is formed by all data points in the tensor $\mathscr{X}$. Since the accuracy for these models is low, it is likely that many brain regions associated with the response are not captured by this

[1]ADNI manuscript citation guidelines. https://adni.loni.usc.edu/wp-content/uploads//how_to_apply/ADNI_DSP_Policy.pdf

approach. This limitation highly motivates us to consider the proposed partition model. We believe that finding local features before applying a projection into the principal components space will not only improve prediction accuracy, but also find new and important brain regions that are associated with AD.

4.2. Selecting the partition model

We then considered: (i) 64 partitions of size 24 × 24 × 24 mm^3; (ii) 512 partitions of size 12 × 12 × 12 mm^3; and (iii) 4096 partitions of size 6 × 6 × 6 mm^3. For different values of R, we selected the number of partitions based on the prediction accuracy of a 10-fold cross validation with the following steps. First, we extracted the features determined by tensor decomposition for different values of rank R. Second, to reduce the dimension of the extracted feature matrix, we projected the matrix $\boldsymbol{L}$ into the principal component space with K basis that keeps 90% of the data variability. Third, we run 100,000 iterations of the Bayesian probit model with the mixture prior described in Section 2.3 with a burn-in of 5,000 samples, and thinning interval of 50. Finally, we computed the mean prediction accuracy, the false positive rate and the false negative rate for each data split. Results are shown in Table 2. We observe that the prediction accuracy does not always increase as R increases. This shows that the locations associated with the response can be represented by a small combination of basis functions. In addition, the accuracy is higher for smaller partitions. This is expected in real data problems when signals are relatively small and their locations spread throughout the brain.

4.3. Final analysis based on the selected model

Based on the prediction accuracy, we selected the model with all partitions of size 6 × 6 × 6 mm^3 and $R = 5$. For the selected model, we fitted TPRM with $\sigma^2 = 10^4$, $\varepsilon = 10^{-4}$, and $a_{0\pi} = a_{1\pi} = 0.5$ to reflect the bathtub prior. In the first screening procedure, we eliminated the partitions whose features, extracted from the tensor decomposition, are zero because they are not relevant in the prediction of AD. From the 4,096 original partitions, only 1,720 passed the first screening, totaling 8,600 features. Figure 4 shows the correlation between the features extracted in the first screening step. Inspecting Figure 4 reveals high correlations between features within most partitions and across nearby partitions. Thus, adding the third component of TPRM can reduce correlation in the selected features.

Next, we projected the features into the space of principal components. We chose the number of principal components K to enter the final model as follows. Specifically, we chooe K by specifying the amount of data variation to be 90%. For this application, we checked the traceplots of the parameters estimates for convergence. The number of final components came down to $K = 50$.

Finally, we run the Gibbs sampler algorithm described in Section 2.4 for 150, 000 iterations with a burn-in period of 5, 000 iterations and thining interval of 50. Based on a 95% credible interval corrected by the number of test using Bonferroni ($a = 0.05/50$), we considered seven components to be important for predicting AD outcome. Convergence plots for the 7 coefficients and their correspondent qqplots are shown in Figure 9 of Appendix D. Panels (a) to (g) of Figure 5 present an axial slice of the 7 important features represented in the image

space in their order of importance. The importance is quantified by the absolute value of the posterior mean for each selected feature. We also present a sensitivity analysis for the hyperparameters $a_{0\pi}$ and $a_{1\pi}$ and conclude that the selected features are consistent across different combinations of these hyperparameters. Results are included in Appendix C.

Second, let $\tilde{\boldsymbol{p}} = \hat{\boldsymbol{b}}^T \boldsymbol{D}$ be a $1 \times P_L$ vector representing the estimated coefficient vector $\hat{\boldsymbol{b}}$ in the local image feature space spanned by the columns of $\boldsymbol{L}$. We computed the projection $\mathscr{P} = \|\Lambda; \boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \boldsymbol{A}^{(3)}, \tilde{\boldsymbol{p}}\|$. The projection $\mathscr{P}$ is a representation of the estimated coefficient vector $\hat{\boldsymbol{b}}$ in the three-dimensional image space. Panel (g) of Figure 5 presents the absolute value of $\mathscr{P}$, indicating regions of differences between the control group and the Alzheimer's group. Values on the right hand side of the colorbar are the regions where differences between AD and controls are high. To highlight these biomarkers, we thresholded $\mathscr{P}$ to reveal some of the important regions for AD prediction (Panel (h) of Figure 5). The threshold value was chosen to select the 5% highest absolute values of the projection $\mathscr{P}$.

To find specific brain locations that are meaningful for predicting AD, we label the signal locations and present it in Figure 5 based on the Jülich atlas (Eickhoff et al., 2005). The largest biomarker is the insula, as shown in Table 4, Appendix D. The insula is associated with perception, self-awareness, and cognitive function. Many studies have revealed its importance as an AD biomarker (Foundas et al., 1997; Karas et al., 2004; Jr. and Holtzman, 2013; Hu et al., 2015). Other important biomarkers are located along the white-matter fiber tracts (fascicles), in particular a region known as the uncinate fascicle, which contains fiber tracts linking regions of the temporal lobe (such as hippocampus and amygdala) to several frontal cortex regions. Abnormalities within the fiber bundles of the uncinate fasciculus have been previously associated with AD (Yasmin et al., 2008; Salminen et al., 2013).

Another important biomarker is the hippocampus, which is associated with learning and consolidation of explicit memories from short-term memory to cortical memory storage for the long term (Campbell and MacQueen, 2004). Previous studies have shown that this region is particularly vulnerable to Alzheimer's disease pathology and already considerably damaged at the time clinical symptoms first appear (Schuff et al., 2009; Braak and Braak, 1998). Other important biomarkers found by TPRM are shown in Table 4, Appendix D.

5. Discussion

We have proposed a Bayesian tensor partition regression model (TPRM) to correlate imaging tensor predictors with clinical outcomes. The ultra-high dimensionality of imaging data is dramatically reduced by using the proposed partition model. Our TPRM efficiently addresses the three key features of imaging data, including relatively low signal to noise ratio, spatially clustered effect regions, and the tensor structure of imaging data. Our simulations and real data analysis confirm that TPRM outperforms some state-of-art methods, while efficiently reducing and identifying relevant imaging biomarkers for accurate prediction.

Many important issues need to be addressed in future research. One limitation of TPRM is that the partition tensors are taken from consecutive voxels and therefore do not represent a

meaningful brain regions. Such partition is critical for the tensor decomposition that accounts for the spatial structure of medical imaging data. If a prior partition obtained from the existing biological brain regions is preferred, a different basis choice, such as principal components or wavelets, is necessary, since the shapes of these regions will not form a hypercube and therefore tensor decomposition is not applicable. Another limitation of TPRM is that we only offer an ad hoc approach to select the number of partitions. This approach is not efficient because we have to run many models with different partition sizes in order to identify the best one according to a criterion, such as the prediction accuracy used here. An automated way of selecting the number of partitions is ideal and a topic for future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare;; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern Rev December 5, 2013 California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

APPENDIX A

SIMULATION FOR BAYESIAN TENSOR DECOMPOSITION

The two goals of the first set of simulations are (i) to compare the Bayesian tensor decomposition method with the alternating least squares method and (ii) to assess the importance of the partition model in the reconstruction of the original image. We considered 3 different imaging data sets (or tensors) including (I·1) a diffusion tensor image (DTI) of size 90 × 96 × 96, (I·2) a white matter RAVENS image of size 99 × 99 × 70, and (I·3) a T2-weighted MRI image of size 64 × 108 × 99. We fitted models (2.4) and (2.5) to the three types of image tensor and decomposed each of them with $R = 5$, 10, and 20. We consider 27 partitions of size 30 × 30 × 32 for the DTI image, 18 partitions of size 33 × 33 × 35 for the RAVENS map, and 24 partitions of size 32 × 27 × 33 for the T2 image, respectively. The hyperparameters $\nu_{0\tau} = 1$, $\nu_{1\tau} = 10^{-2}$, and $\kappa = 10^{-6}$ were chosen to reflect non-informative priors.

We run steps (a.1) – (a.4) of the Gibbs sampler algorithm in Section 2.4 for 5, 000 iterations. Figure 6 shows the trace plots of Gibbs sampler at 9 randomly selected voxels based on the

results for the reconstructed RAVENS map decomposed with $R = 20$. The proposed algorithm converges very fast in all voxels. At each iteration, we computed the quantity $\mathcal{I} = \sum_{s=1}^{S} \left\| \Lambda_s; \boldsymbol{A}_s^{(1)}, \boldsymbol{A}_s^{(2)}, \boldsymbol{A}_s^{(3)} \right\|$ for each rank and each partition. Subsequently, we computed the reconstructed image, defined as $\hat{\mathcal{X}}$, and the posterior mean estimate of $\mathcal{S}$ after a burn-in sample of 3, 000 iterations. For each reconstructed image $\hat{\mathcal{X}}$, we computed its root mean squared error, RMSE $= \|\widehat{\mathcal{X}} - \mathcal{X}\|_2 / \sqrt{J_1 J_2 J_3}$.

We consider the non-partition model and compare the Bayesian method with the standard alternating least squares method (Kolda and Bader, 2009). Figure 7 shows an axial slice of the original white Matter RAVENS map and the reconstructed images for ranks $R = 5$, 10, and 20 as $S = 1$. Table 3 presents RMSEs obtained from the three methods in all scenarios. The Bayesian decomposition method gives a smaller RMSE for all cases. As expected, the higher the rank, the smaller the reconstruction error.

APPENDIX B

GIBBS SAMPLING ALGORITHM FOR TPRM

We provide the Gibbs sampling algorithm to sample from the posterior distribution (2.10) in Section 2.4. It involves sampling from a series of conditional distributions, while each of the modeling components is updated in turn. As an illustration, we divide the whole image into S equal sized regions and assume $y_i \sim$ Bernoulli(μ_i) with the link function $h(\cdot)$ being the probit function. By following Albert and Chib (1993), we introduce a normally distributed latent variable, w_i, such that $w_i \sim N(\mu_i, 1)$ and $y_i = \mathbf{1}(w_i > 0)$, where $\mathbf{1}(\cdot)$ is an indicator function of an event.

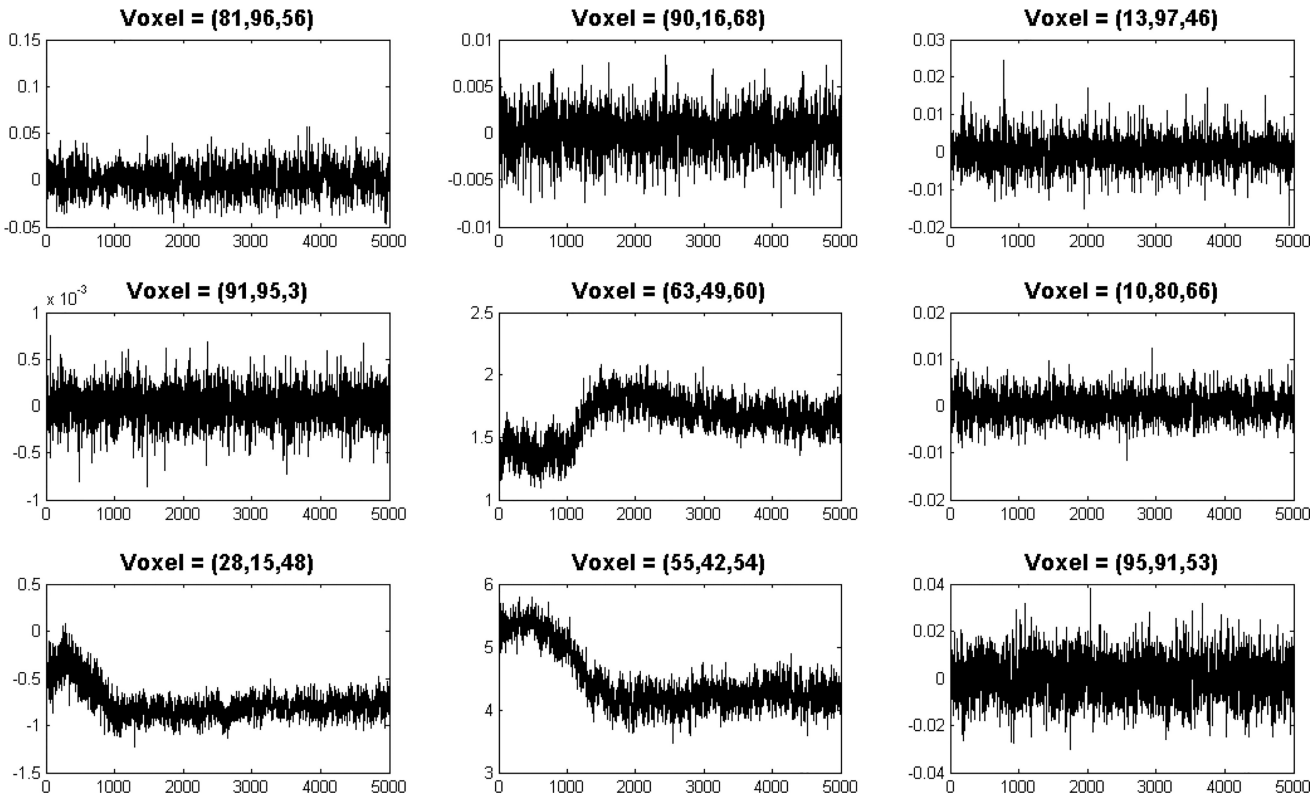

Fig 6.

Trace plots of Gibbs samplers in 9 randomly selected voxels for the RAVENS map obtained by Bayesian tensor decomposition with $R = 20$. The trace plots indicate that the Markov chains converge after around 2,000 iterations.

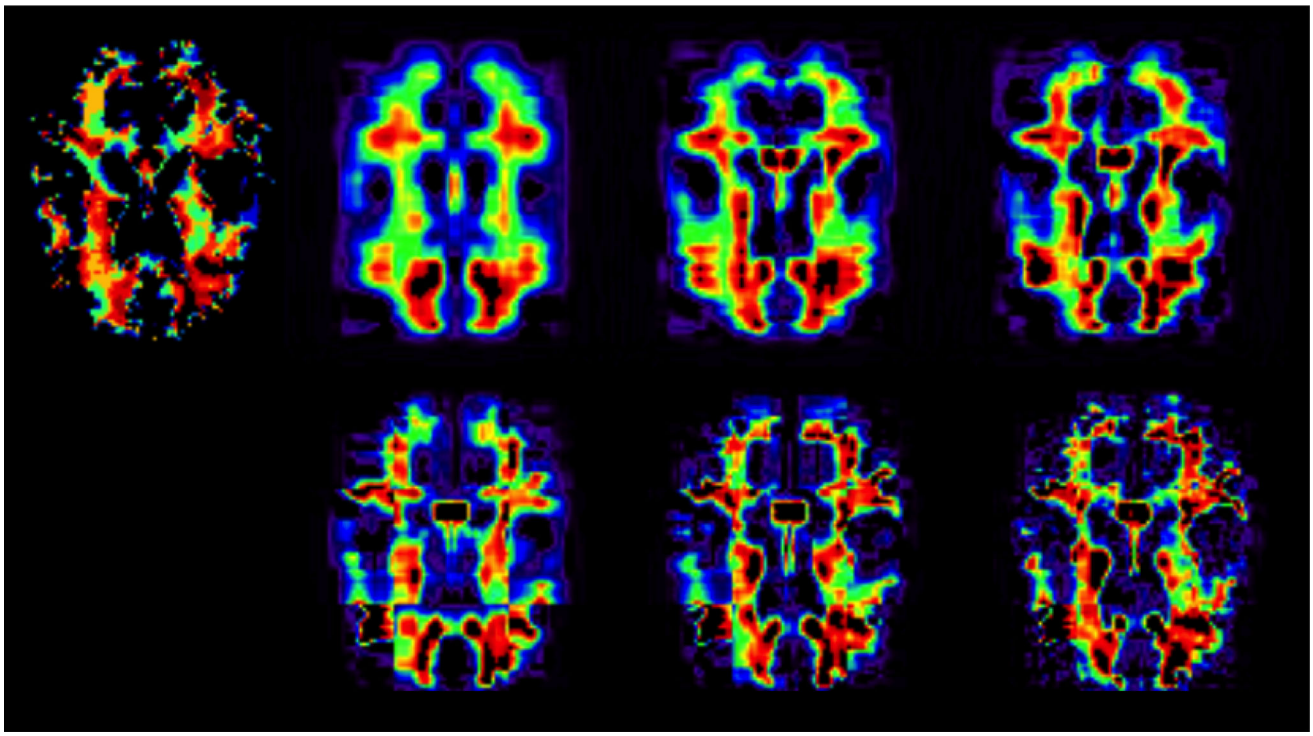

Fig 7.

Bayesian tensor decomposition results. Top panels: the image on the left represents an axial slice of the RAVENS map image, followed by reconstruction results for the nonpartition model. Bottom panels: reconstruction results for the partition model. From left to right, we have the decomposed images for ranks $R = 5$, 10, and 20, respectively.

Table 3

Root mean squared error for 3 different types of imaging data. The Bayesian decomposition outperforms the alternating least squares in all scenarios. As the rank R increases, the error decreases.

		T2-weighted	WM RAVENS	DTI
R=5	BayesianCP	45.3191	1.5853	3.1656e-004
	ALS	45.3636	1.6013	3.2506e-004
	Partition	37.3712	1.2178	2.0929e-004
R=10	BayesianCP	41.7018	1.4382	2.7367e-004
	ALS	42.4350	1.4533	2.8247e-004
	Partition	31.3836	1.0186	1.5748e-004
R=20	BayesianCP	37.1796	1.2885	2.2911e-004
	ALS	38.3166	1.3166	2.3676e-004
	Partition	25.1574	0.8085	1.1349e-004

The complete Gibbs sampler algorithm proceeds as follows.

(*a*.0) Generate $\boldsymbol{w} = (w_1, \ldots, w_n)^T$ from

$$w_i | y_i = 0 \sim \mathbf{1}(w_i \leq 0) \mathrm{N}(z_i^T \boldsymbol{\gamma} + \boldsymbol{g}_i^T \boldsymbol{b}, 1),$$

$$w_i | y_i = 1 \sim \mathbf{1}(w_i \geq 0) \mathrm{N}(z_i^T \boldsymbol{\gamma} + \boldsymbol{g}_i^T \boldsymbol{b}, 1).$$

(*a*.1) Update $\boldsymbol{\tau}(s)$ from its full conditional distribution

$$\tau(s) | - \sim \mathrm{Gamma}(\nu_{0\tau} + (N \prod_{d=1}^{D} p_d)/2, \nu_{1\tau} + (1/2) \sum_{i, j_1, \ldots, j_D} (x^*_{j_1, \ldots, j_D i}(s))^2),$$

where $x^*_{j_1, \ldots, j_D i}(s) = \{\mathcal{X}^{(s)} - \| \Lambda^{(s)}; \boldsymbol{A}_s^{(1)}, \boldsymbol{A}_s^{(2)}, \ldots, \boldsymbol{A}_s^{(D)}, \boldsymbol{L}^{(s)} \| \}_{j_1, \ldots, j_D i}$.

(*a*.2) Update $\{\boldsymbol{A}_s^{(d)}\}_{j_d r}$ from its full conditional distribution given by

$$\{\boldsymbol{A}_s^{(d)}\}_{j_d r} | - \sim \mathrm{N}\left(\frac{\tau^{(s)} \langle \widehat{\mathcal{X}}^{s(j_d)}_{(-r)}, \mathcal{F}^s_{(-d)} \rangle}{\tau^{(s)} \langle \mathcal{F}^s_{(-d)}, \mathcal{F}^s_{(-d)} \rangle + p_d}, \left(\tau^{(s)} \langle \mathcal{F}^s_{(-d)}, \mathcal{F}^s_{(-d)} \rangle + p_d \right)^{-1} \right),$$

where $\mathcal{F}^s_{(-d)} = \| \Lambda^{(s)}; \boldsymbol{A}_s^{(1)}, \ldots, \boldsymbol{A}_s^{(d-1)}, \boldsymbol{A}_s^{(d+1)}, \ldots, \boldsymbol{A}_s^{(D)}, \boldsymbol{L}^{(s)} \|, \widehat{\mathcal{X}}^s_{(-r)}$ is given by $\mathcal{X}^{(s)} - \| \Lambda^{(s)}; \boldsymbol{A}_s^{(1)}, \boldsymbol{A}_s^{(2)}, \ldots, \boldsymbol{A}_s^{(D)}, L_i^{(s)} \| + \| \Lambda^{(s)}; \{\boldsymbol{A}_s^{(1)}\}_{:,r}, \{\boldsymbol{A}_s^{(2)}\}_{:,r}, \ldots, \{\boldsymbol{A}_s^{(D)}\}_{:,r}, \{\boldsymbol{L}_i^{(s)}\}_{:,r} \|$

and $\widehat{\mathcal{X}}^{s(j_d)}_{(-r)}$ is a sub-tensor fixed at the entry j_d along the d-th dimension of $\widehat{\mathcal{X}}^s_{(-r)}$.

(*a*.3) Update $\{\boldsymbol{L}_s\}_{ir}$ from its full conditional distribution given by

$$\{\boldsymbol{L}_s\}_{ir} | - \sim \mathrm{N}\left(\frac{\tau^{(s)} \langle \widehat{\mathcal{X}}^{s(i)}_{(-r)}, \mathcal{F}^s \rangle}{\tau^{(s)} \langle \mathcal{F}^s, \mathcal{F}^s \rangle + N}, \left(\tau^{(s)} \langle \mathcal{F}^s, \mathcal{F}^s \rangle + N \right)^{-1} \right),$$

where $\mathcal{F}^s = \| \Lambda^{(s)}; \boldsymbol{A}_s^{(1)}, \ldots, \boldsymbol{A}_s^{(D)} \|, \widehat{\mathcal{X}}^s_{(-r)}$ is the same as above, and $\widehat{\mathcal{X}}^{s(i)}_{(-r)}$ is a subtensor fixed at the i-th entry along the subject dimension of $\widehat{\mathcal{X}}^s_{(-r)}$.

(*a*.4) Normalize the columns of $\boldsymbol{A}_s^{(d)}$ and $\boldsymbol{L}^{(s)}$ and compute $\Lambda^{(s)}$ with

$$\lambda_r^{(s)} = \| \boldsymbol{A}_s^{(1)} \| \times \ldots \times \| \boldsymbol{A}_s^{(D)} \| \times \| \boldsymbol{L}^{(s)} \|.$$

(*a*.5) Update $\boldsymbol{g}_k$ from its full conditional distribution

$$\boldsymbol{g}_k|-\sim \mathrm{N}(\mu_g, \sum_g), \sum_g = (n\boldsymbol{I}_n + \tau_\psi \sum_{j=1}^{P_L} d_{kj}^2)^{-1} \text{ and } \mu_g = \tau_\psi \sum_g \sum_{j=1}^{P_L} d_{kj} \boldsymbol{l}_j^{*-k},$$

where $\boldsymbol{l}_j^{*-k} = \boldsymbol{L} - \boldsymbol{G}\boldsymbol{d}_j + d_{kj}\boldsymbol{g}_k$ for $j = 1, \ldots, P_L$.

(*a*.6) Update d_{kj} for $j = 1, \ldots, P_L$ from its full conditional distribution

$$d_{kj}|-\sim \mathrm{N}(\tau_\psi \sum_d \sum_{j=1}^{P_L} \boldsymbol{g}_k^T \boldsymbol{l}_j^{*-k}, \sum_d),$$

where $\sum_d = \left(1 + \tau_\psi \sum_{j=1}^{P_L} \boldsymbol{g}_k^T \boldsymbol{g}_k\right)^{-1}$.

(*a*.7) Update τ_ψ from its full conditional distribution

$$\tau_\psi|-\sim \mathrm{Gamma}\,(\beta_{0\psi} + NP_L/2, \beta_{1\psi} + (\boldsymbol{L}^{*T}\boldsymbol{L}^*)/2),$$

where $\boldsymbol{L}^* = \boldsymbol{L} - \boldsymbol{G}\boldsymbol{D}$.

(*a*.8) Update δ_k from its full conditional distribution

$$\delta_k \sim \mathrm{Bernoulli}(\tilde{p}_1/(\tilde{p}_1 + \tilde{p}_0)),$$

where $\tilde{p}_1 = \pi \exp\{-(1/2\sigma^2)b_k^2\}$ and $\tilde{p}_0 = \pi \exp\{-(1/2\varepsilon)b_k^2\}$.

(*a*.9) Update $\boldsymbol{b}$ from its full conditional distribution

$$b_k|\delta_k = 1 \sim \mathrm{N}(\sum_i \tilde{w}_i g_{ik}/(\sum_i g_{ik}^2 + 1/\sigma^2), (\sum_i g_{ik}^2 + 1/\sigma^2)^{-1}),$$

$$b_k|\delta_k = 0 \sim \mathrm{N}(\sum_i \tilde{w}_i g_{ik}/(\sum_i g_{ik}^2 + 1/\varepsilon), (\sum_i g_{ik}^2 + 1/\varepsilon)^{-1}),$$

where $\tilde{w}_i = w_i - z_i^T\boldsymbol{\gamma} - \sum_{s'=1}^{S} \boldsymbol{g}_i^{(s')T}\boldsymbol{b}^{(s')} + g_{ir}^{(s)T} b_r^{(s)}$.

(*a*.10) Update $\boldsymbol{\pi}$ from its full conditional distribution

$$\pi|-\sim \mathrm{beta}(\alpha_{0\pi} + \sum_k \delta_k, \alpha_{1\pi} + K - \sum_k \delta_k).$$

(*a*.11) Update $\boldsymbol{\gamma}$ from its full conditional distribution

$$\gamma | - \sim \mathrm{N}(\sum_{\gamma}^{*-1}(\upsilon\gamma^{*} + \boldsymbol{Z}^{T}\boldsymbol{w}_{\gamma}^{*}), \sum_{\gamma}^{*-1}),$$

where $\sum_{\gamma}^{*} = \upsilon \boldsymbol{I}_{q} + \boldsymbol{Z}^{T}\boldsymbol{Z}$ and $= w_{\gamma}^{*} = \boldsymbol{w} - \boldsymbol{G}^{T}\boldsymbol{b}$.

(*a*.12) Update υ from its full conditional distribution

$$\upsilon | - \sim \mathrm{Gamma}\,(\nu_{0\upsilon} + q/2, \nu_{1\upsilon} + (\gamma^{T}\gamma)/2)\,.$$

All the tensor operations described in steps (*a*.1) – (*a*.4) can be easily computed using Bader et al. (2015), at http://www.sandia.gov/~tgkolda/TensorToolbox/index-2.5.html.

APPENDIX C

SENSITIVITY ANALYSIS

We present some results obtained from a sensitivity analysis on the hyperparameters $a_{0\pi}$ and $a_{1\pi}$ in (2.9). For different combinations of the hyperparameters, we run steps (a.8)–(a.10) in order to select a subset of variables. Figure 8 shows the MCMC results. The *x*-axis indicates the decision for each of the $K = 100$ features. A white color indicates that a specific feature was selected in TPRM, whereas a black color indicates exclusion. The selected features are similar to each other for all combinations of $a_{0\pi}$ and $a_{1\pi}$.

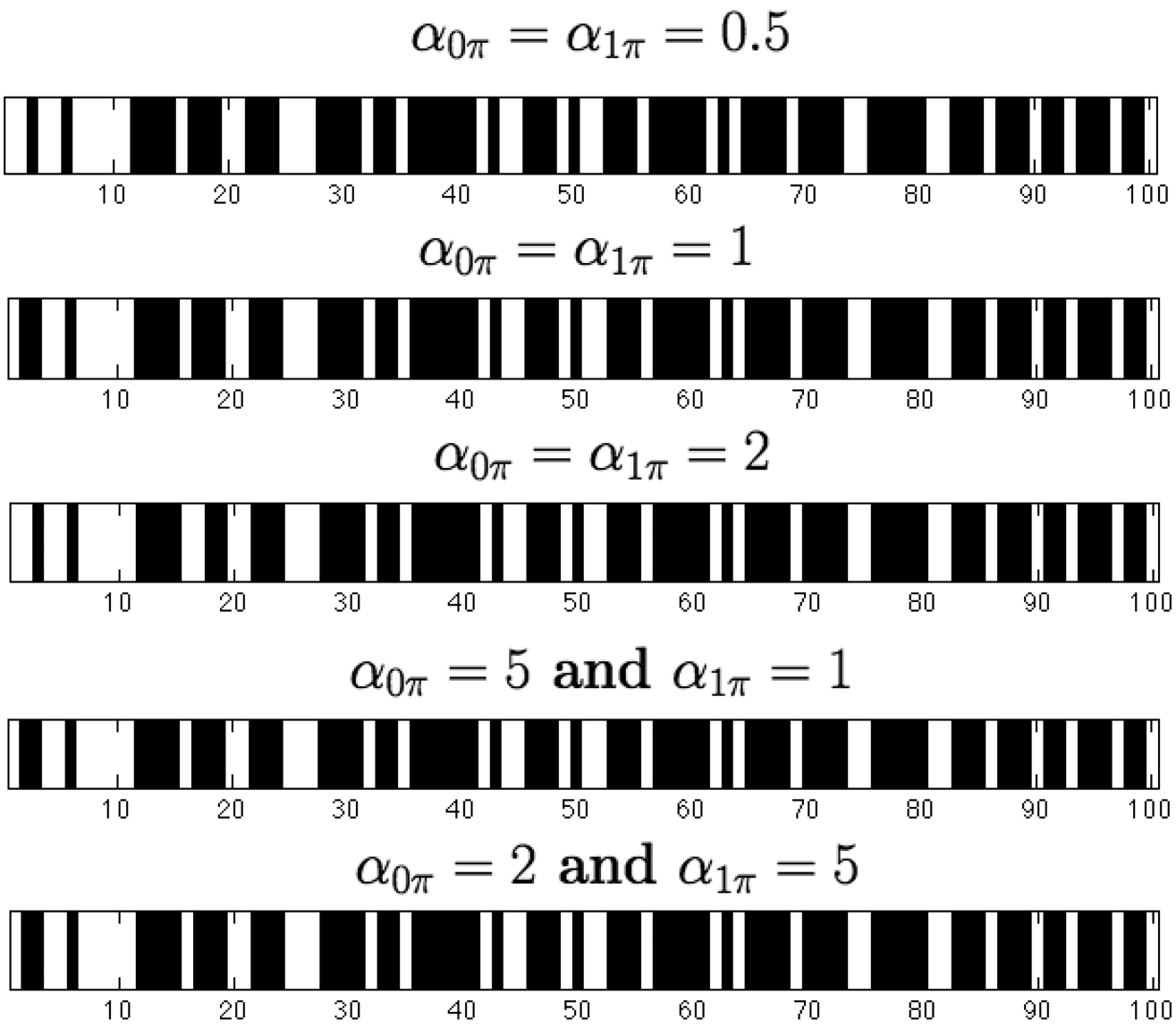

Fig 8.

Sensitivity analysis for the hyperparameters $\alpha_{0\pi}$ and $\alpha_{1\pi}$ of the bathtub prior in (2.9). A white color indicates that the feature was selected in the model, whereas a black color indicates exclusion. The selected features are similar to each other for all combinations of $\alpha_{0\pi}$ and $\alpha_{1\pi}$.

APPENDIX D

REAL DATA ANALYSIS SUPPORTING MATERIALS

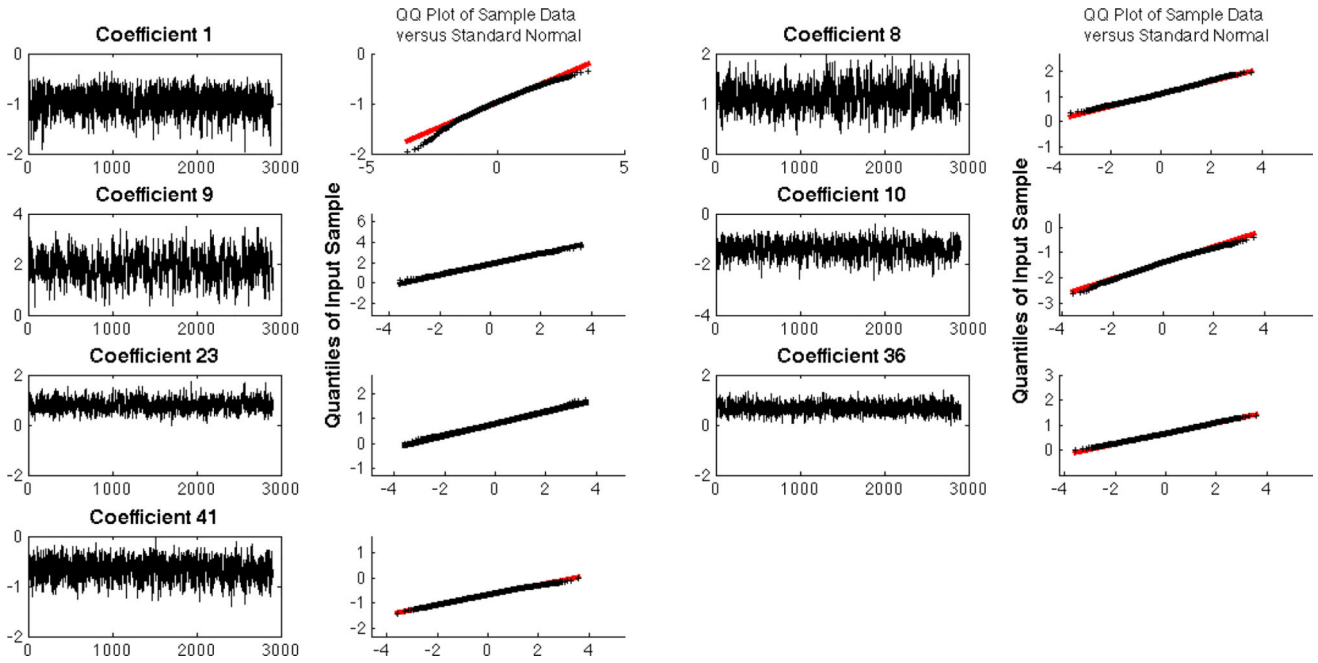

Fig 9.

Traceplots for the 7 significant coefficients, with their corresponding qqplots. The results confirm convergence of the MCMC samplers. In addition, coefficients seem to follow a standard Gaussian distribution.

Table 4

Biomarkers that are relevant to predict AD outcome, based on the Jülich atlas. Columns represent the region name, the total amount of voxels in the corresponding region, the number of voxels above the threshold of the projection $\mathscr{P}$, and the percentage of significant voxels considering the total size of the region, respectively.

Region	# voxels	# sig. voxels	%
GM Insula Ig1 R	189	175	93
GM Insula Id1 L	558	441	79
GM Insula Ig2 R	743	585	79
GM Visual cortex V1 BA17 L	6367	4988	78
GM Hippocampus dentate gyrus L	6084	4721	78
WM Inferior occipito-frontal fascicle L	1708	1305	76
GM Superior parietal lobule 7A R	14507	10512	72
GM Lateral geniculate body R	1645	1180	72
WM Uncinate fascicle L	571	401	70
GM Hippocampus dentate gyrus R	647	451	70

Region	# voxels	# sig. voxels	%
GM Primary motor cortex BA4a R	7737	5208	67
GM Inferior parietal lobule PGp L	8903	5964	67
GM Inferior parietal lobule PGp R	10418	6679	64
GM Inferior parietal lobule PF R	7911	4957	63
GM Broca's area BA44 L	1555	967	62
GM Superior parietal lobule 5M R	2700	1668	62
GM Primary auditory cortex TE1.0 L	10423	6100	59
GM Inferior parietal lobule PFt L	2054	1173	57
GM Primary auditory cortex TE1.0 R	1614	895	55
GM Primary somatosensory cortex BA1 R	7170	3859	54

APPENDIX E

SIMULATION RESULTS, SECTION 3.1

Table 5

Model Comparison (FPCA, TALS, and TPRM) - prediction accuracy, false positive rate, and false negative rate for each fold and scenarios (S.1) – (S.4) described on Section 3.1.

Scenario 1								
Prediction Accuracy			**False Positive Rate**			**False Negative Rate**		
FPCA	**TALS**	**TPRM**	**FPCA**	**TALS**	**TPRM**	**FPCA**	**TALS**	**TPRM**
0.500	0.550	0.950	0.000	0.300	0.100	1.000	0.600	0.000
0.600	0.650	0.850	0.000	0.083	0.250	1.000	0.750	0.000
0.350	0.450	0.850	0.000	0.714	0.000	1.000	0.462	0.231
0.600	0.500	1.000	0.000	0.500	0.000	1.000	0.500	0.000
0.650	0.550	0.850	0.000	0.308	0.231	1.000	0.714	0.000
0.800	0.650	0.700	0.000	0.438	0.375	1.000	0.000	0.000
0.700	0.550	0.750	0.000	0.357	0.357	1.000	0.667	0.000
0.450	0.500	0.950	0.000	0.667	0.000	1.000	0.364	0.091
0.500	0.650	0.950	0.000	0.300	0.100	1.000	0.400	0.000
0.600	0.700	0.950	0.000	0.083	0.083	1.000	0.625	0.00
Scenario 2								
0.500	0.700	1.000	0.000	0.200	0.000	1.000	0.400	0.000
0.600	0.500	0.950	0.000	0.583	0.083	1.000	0.375	0.000
0.350	0.700	0.900	0.000	0.571	0.000	1.000	0.154	0.154
0.600	0.550	1.000	0.000	0.250	0.000	1.000	0.750	0.000
0.650	0.550	0.750	0.000	0.615	0.308	1.000	0.143	0.143
0.750	0.600	0.800	0.063	0.375	0.250	1.000	0.500	0.000
0.700	0.750	0.950	0.000	0.000	0.071	1.000	0.833	0.000

Scenario 1								
Prediction Accuracy			False Positive Rate			False Negative Rate		
FPCA	TALS	TPRM	FPCA	TALS	TPRM	FPCA	TALS	TPRM
0.450	0.550	1.000	0.000	0.556	0.000	1.000	0.364	0.000
0.500	0.700	0.800	0.000	0.500	0.200	1.000	0.100	0.200
0.600	0.550	0.950	0.000	0.167	0.083	1.000	0.875	0.000
Scenario 3								
0.500	0.600	0.700	0.000	0.300	0.300	1.000	0.500	0.300
0.600	0.400	0.350	0.000	0.583	0.667	1.000	0.625	0.625
0.350	0.300	0.850	0.000	0.714	0.000	1.000	0.692	0.231
0.600	0.600	0.800	0.000	0.000	0.250	1.000	1.000	0.125
0.650	0.600	0.450	0.000	0.077	0.538	1.000	1.000	0.571
0.800	0.650	0.700	0.000	0.375	0.313	1.000	0.250	0.250
0.700	0.700	0.600	0.000	0.143	0.500	1.000	0.667	0.167
0.450	0.650	0.500	0.000	0.222	0.667	1.000	0.455	0.364
0.500	0.600	0.550	0.000	0.100	0.500	1.000	0.700	0.400
0.600	0.600	0.600	0.000	0.167	0.333	1.000	0.750	0.500
Scenario 4								
0.500	0.450	0.750	0.000	0.700	0.200	1.000	0.400	0.300
0.600	0.650	0.750	0.000	0.167	0.250	1.000	0.625	0.250
0.350	0.450	0.750	0.000	0.143	0.143	1.000	0.769	0.308
0.600	0.500	0.600	0.000	0.250	0.417	1.000	0.875	0.375
0.650	0.350	0.750	0.000	1.000	0.385	1.000	0.000	0.000
0.800	0.600	0.650	0.000	0.438	0.375	1.000	0.250	0.250
0.700	0.650	0.550	0.000	0.214	0.429	1.000	0.667	0.500
0.450	0.650	0.950	0.000	0.556	0.000	1.000	0.182	0.091
0.500	0.550	0.600	0.000	0.000	0.400	1.000	0.900	0.400
0.600	0.500	0.800	0.000	0.250	0.167	1.000	0.875	0.250

Table 6

Model Comparison (FPCA, TALS, and TPRM) - prediction accuracy, false positive rate, and false negative rate for each fold and scenarios (S.5) – (S.8) described on Section 3.1.

Scenario 5								
Prediction Accuracy			False Positive Rate			False Negative Rate		
FPCA	TALS	TPRM	FPCA	TALS	TPRM	FPCA	TALS	TPRM
0.800	0.850	0.950	0.000	0.000	0.100	0.400	0.300	0.000
0.900	0.900	1.000	0.000	0.083	0.000	0.250	0.125	0.000
0.350	0.750	0.900	0.000	0.143	0.000	1.000	0.308	0.154

Scenario 5								
Prediction Accuracy			False Positive Rate			False Negative Rate		
FPCA	TALS	TPRM	FPCA	TALS	TPRM	FPCA	TALS	TPRM
0.700	0.750	0.900	0.000	0.167	0.167	0.750	0.375	0.000
0.750	0.650	0.900	0.000	0.462	0.154	0.714	0.143	0.000
0.900	0.800	0.850	0.000	0.188	0.188	0.500	0.250	0.000
0.850	1.000	0.900	0.000	0.000	0.143	0.500	0.000	0.000
0.550	0.750	0.950	0.000	0.222	0.000	0.818	0.273	0.091
0.850	0.750	1.000	0.000	0.000	0.000	0.300	0.500	0.000
0.950	0.800	0.900	0.000	0.333	0.167	0.125	0.000	0.000
Scenario 6								
0.500	0.900	0.750	0.000	0.000	0.300	1.000	0.200	0.200
0.600	0.800	0.850	0.000	0.083	0.250	1.000	0.375	0.000
0.350	0.900	0.750	0.000	0.000	0.143	1.000	0.154	0.308
0.600	0.650	1.000	0.000	0.083	0.000	1.000	0.750	0.000
0.650	0.650	0.600	0.000	0.385	0.462	1.000	0.286	0.286
0.800	0.850	0.750	0.000	0.188	0.313	1.000	0.000	0.000
0.800	0.650	0.750	0.000	0.357	0.286	0.667	0.333	0.167
0.450	0.850	0.950	0.000	0.111	0.000	1.000	0.182	0.091
0.500	0.650	0.900	0.000	0.000	0.100	1.000	0.700	0.100
0.600	0.550	0.950	0.000	0.250	0.083	1.000	0.750	0.000
Scenario 7								
0.500	0.450	0.600	0.000	0.500	0.300	1.000	0.600	0.500
0.600	0.550	0.650	0.000	0.167	0.333	1.000	0.875	0.375
0.350	0.350	0.650	0.000	0.429	0.286	1.000	0.769	0.385
0.600	0.500	0.700	0.000	0.333	0.333	1.000	0.750	0.250
0.650	0.600	0.550	0.000	0.231	0.538	1.000	0.714	0.286
0.800	0.650	0.750	0.000	0.375	0.313	1.000	0.250	0.000
0.700	0.600	0.700	0.000	0.214	0.357	1.000	0.833	0.167
0.450	0.500	0.650	0.000	0.667	0.333	1.000	0.364	0.364
0.500	0.550	0.750	0.000	0.000	0.300	1.000	0.900	0.200
0.600	0.450	0.750	0.000	0.333	0.333	1.000	0.875	0.125
Scenario 8								
0.500	0.500	0.600	0.000	0.400	0.400	1.000	0.600	0.400
0.600	0.550	0.650	0.000	0.250	0.333	1.000	0.750	0.375
0.350	0.750	0.600	0.000	0.143	0.286	1.000	0.308	0.462
0.600	0.700	0.850	0.000	0.167	0.167	1.000	0.500	0.125
0.650	0.450	0.800	0.000	0.462	0.308	1.000	0.714	0.000
0.800	0.500	0.700	0.000	0.625	0.375	1.000	0.000	0.000

Scenario 5								
Prediction Accuracy			False Positive Rate			False Negative Rate		
FPCA	TALS	TPRM	FPCA	TALS	TPRM	FPCA	TALS	TPRM
0.750	0.850	0.800	0.000	0.000	0.286	0.833	0.500	0.000
0.450	0.700	0.800	0.000	0.111	0.000	1.000	0.455	0.364
0.500	0.800	0.800	0.000	0.200	0.100	1.000	0.200	0.300
0.600	0.750	0.750	0.000	0.333	0.167	1.000	0.125	0.375

References

Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association. 1993; 88(422):669–679.

Bader BW, Kolda TG, et al. Matlab tensor toolbox version 2.6. 2015 Available online.

Bair E, Hastie T, Paul D, Tibshirani R. Prediction by supervised principal components. Journal of the American Statistical Association. 2006; 101:119–137.

Beckmann CF, Smith SM. Tensorial extensions of independent component analysis for multisubject fMRI analysis. NeuroImage. 2005; 25(1):294–311. [PubMed: 15734364]

Bickel P, Levina E. Some theory for Fisher's linear discriminant function, 'naive Bayes', and some alternatives when there are many more variables than observations. Bernoulli. 2004; 10:989–1010.

Braak H, Braak E. Evolution of neuronal changes in the course of Alzheimer's disease. In: Jellinger K, Fazekas F, Windisch M, editorsAgeing and Dementia, volume 53 of Journal of Neural Transmission. Supplementa. Springer; Vienna: 1998. 127–140.

Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Wadsworth; California: 1984.

Caffo B, Crainiceanu C, Verduzco G, Joel S, SH M, Bassett S, Pekar J. Two-stage decompositions for the analysis of functional connectivity for fMRI with application to Alzheimer's disease risk. Neuroimage. 2010; 51(3):1140–1149. [PubMed: 20227508]

Campbell S, MacQueen G. The role of the hippocampus in the pathophysiology of major depression. Journal of Psychiatry and Neuroscience. 2004; 29(6):417–426. [PubMed: 15644983]

Davatzikos C, Genc A, Xu D, Resnick SM. Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy. NeuroImage. 2001; 14(6):1361–1369. [PubMed: 11707092]

Ding X, He L, Carin L. Bayesian robust principal component analysis. Imaging Processing, IEEE Transactions on. 2011; 20(12):3419–3430.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage. 2005; 25(4):1325–1335. [PubMed: 15850749]

Fan J, Fan Y. High-dimensional classification using features annealed independence rules. Annals of Statistics. 2008; 36:2605–2637. [PubMed: 19169416]

Foundas A, Leonard C, Mahoney SM, Agee O, Heilman K. Atrophy of the hippocampus, parietal cortex, and insula in alzheimer's disease: a volumetric magnetic resonance imaging study. Neuropsychiatry, Neuropsychology, and Behavioral Neurology. 1997; 10(2):81–9.

Friedman J. Multivariate adaptive regression splines (with discussion). Annals of Statistics. 1991; 19:1–141.

George EI, McCulloch RE. Variable selection via Gibbs sampling. Journal of the American Statistical Association. 1993; 88(423):881–889.

George EI, McCulloch RE. Approaches for Bayesian variable selection. Statistica Sinica. 1997; 7:339–373.

Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology. 2016; 278:563–577. [PubMed: 26579733]

Gonçalves F, Gamerman D, Soares T. Simultaneous multifactor DIF analysis and detection in item response theory. Computational Statistics & Data Analysis. 2013; 59(0):144–160.

Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. Springer; Hoboken, New Jersey: 2009.

Hu X, Meiberth D, Newport B, Jessen F. Anatomical correlates of the neuropsychiatric symptoms in alzheimer's disease. Current Alzheimer Research. 2015; 12(3):266–277. [PubMed: 25731626]

Huang M, Nichols T, Huang C, Yang Y, Lu Z, Feng Q, Knickmeyere RC, Zhu H. for the Alzheimer's Disease Neuroimaging Initiative. FVGWAS: Fast voxelwise genome wide association analysis of large-scale imaging genetic data. NeuroImage. 2015; 118:613–627. [PubMed: 26025292]

Johnstone IM, Lu AY. On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Assocation. 2009; 104:682–693.

Jr CRJ, Holtzman DM. Biomarker modeling of alzheimer's disease. Neuron. 2013; 80(6):1347–1358. [PubMed: 24360540]

Karas G, Scheltens P, Rombouts S, Visser P, van Schijndel R, Fox N, Barkhof F. Global and local gray matter loss in mild cognitive impairment and alzheimer's disease. NeuroImage. 2004; 23(2):708–716. [PubMed: 15488420]

Kolda TG. Multilinear operators for higher-order decompositions. Technical report. 2006

Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Rev. 2009; 51(3):455–500.

Krishnan A, Williams L, McIntosh A, Abdi H. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage. 2011; 56:455–475. [PubMed: 20656037]

Martinez E, Valdes P, Miwakeichi F, Goldman RI, Cohen MS. Concurrent EEG/fMRI analysis by multiway partial least squares. NeuroImage. 2004; 22(3):1023–1034. [PubMed: 15219575]

Mayrink VD, Lucas JE. Sparse latent factor models with interactions: Analysis of gene expression data. The Annals of Applied Statistics. 2013; 7(2):799–822.

Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. Journal of the American Statistical Association. 1988; 83(404):1023–1032.

Müller H-G, Yao F. Functional additive models. Journal of the American Statistical Assocation. 2008; 103(484):1534–1544.

Ramsay JO, Silverman BW. Springer Series in Statistics. second. Springer; New York: 2005. Functional data analysis.

Reiss PT, Ogden RT. Functional generalized linear models with images as predictors. Biometrics. 2010; 66(1):61–69. [PubMed: 19432766]

Ročková V, George EI. Emvs: The em approach to bayesian variable selection. Journal of the American Statistical Association. 2014; 109(506):828–846.

Salminen LE, Schofield PR, Lane EM, Heaps JM, Pierce KD, Cabeen R, Paul RH. Neuronal fiber bundle lengths in healthy adult carriers of the apoe4 allele: A quantitative tractography dti study. brain imaging and behavior. Brain Imaging and Behavior. 2013; 7(3):81–89.

Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR Jr, Weiner MW. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain. 2009; 132(4):1067–1077. [PubMed: 19251758]

Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proceedings of the National Academy of Sciences. 2002; 99:6567–6572.

Yasmin H, Nakata Y, Aoki S, Abe O, Sato N, Nemoto K, Arima K, Furuta N, Uno M, Hirai S, Masutani Y, Ohtomo K. Diffusion abnormalities of the uncinate fasciculus in alzheimer's disease: diffusion tensor tract-specific analysis using a new method to measure the core of the tract. Neuroradiology. 2008; 50(4):293–299. [PubMed: 18246334]

Zhang HP, Singer BH. Recursive Partitioning and Applications. 2. Springer; New York: 2010.

Zhou H, Li L, Zhu H. Tensor regression with applications in neuroimaging data analysis. Journal of the American Statistical Association. 2013; 108:540–552. [PubMed: 24791032]

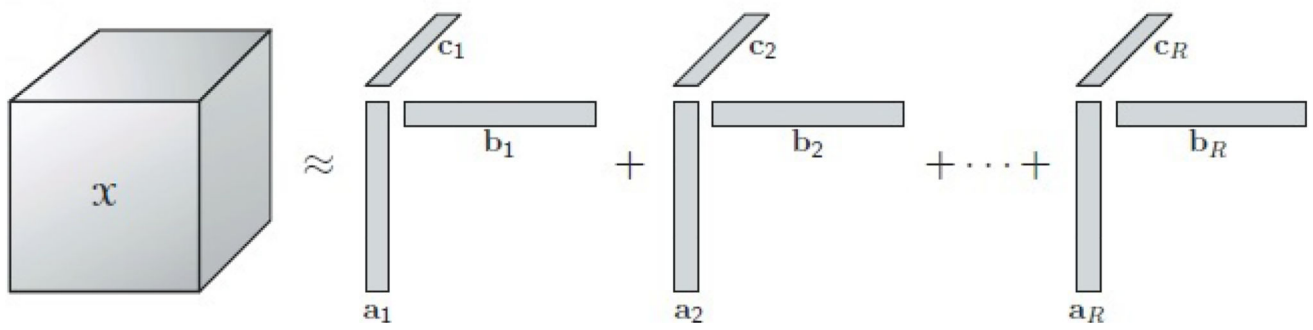

Fig 1.

Figure copied from (Kolda and Bader, 2009). Panel (a) illustrates the CP decomposition of a three way array as a sum of R components of rank-one tensors, i.e. $\mathcal{X} \approx \sum_{r=1}^{R} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r$. The approximation sign means that the right hand side is the the solution of $\min_{\tilde{\mathcal{X}}} \|\mathcal{X} - \tilde{\mathcal{X}}\|_2$, where $\|\cdot\|_2$ is the L_2 norm of tensors and $\tilde{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r$. This minimization problem can be written in a matricized version and solved using an alternating least squares (ALS) algorithm, please see Kolda and Bader (2009) for details.

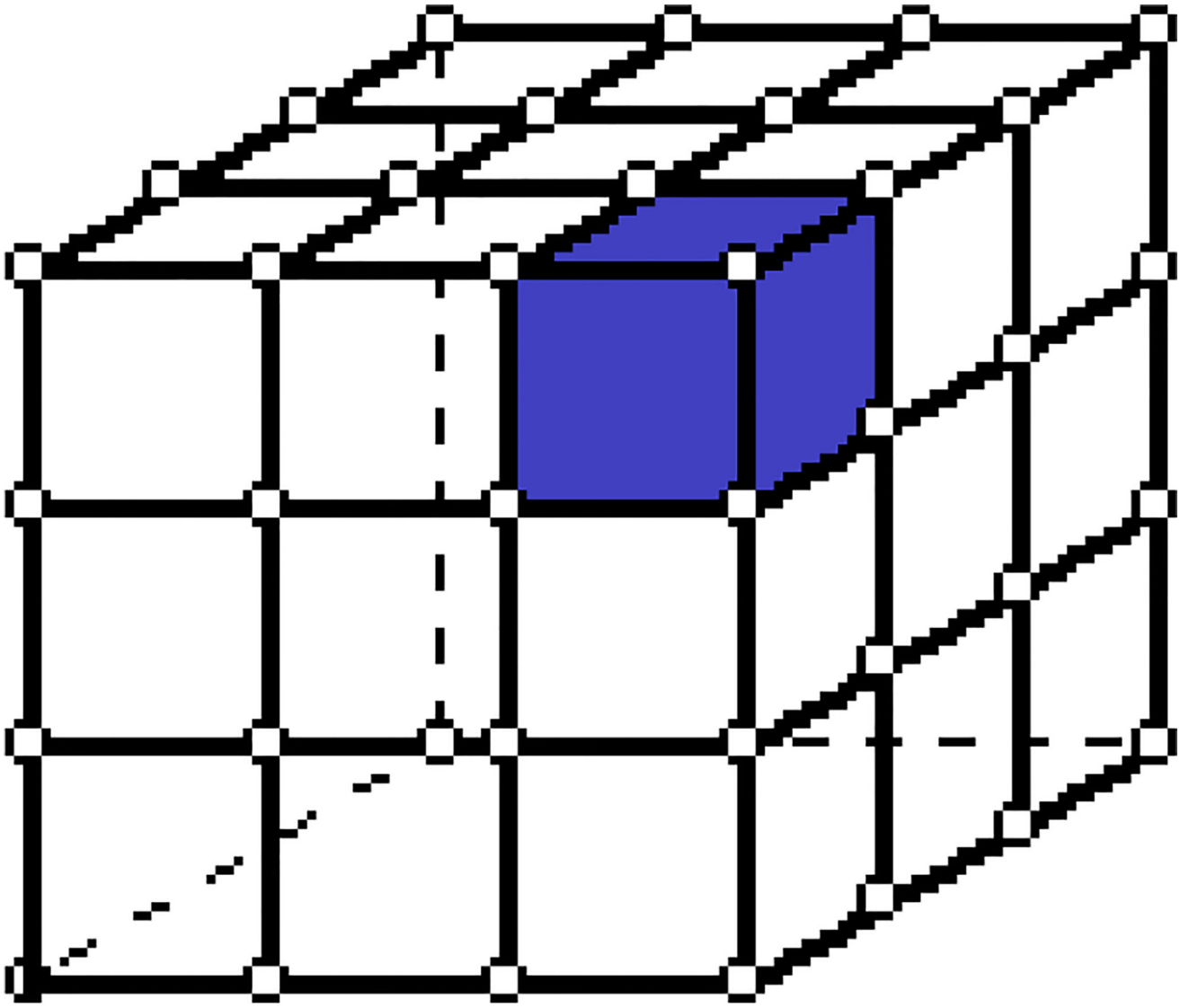

Fig 2.

Partition illustration. The purple cube illustrates a sub-tensor $\tilde{\mathcal{X}}^{(s)}$. For $s = 1, \ldots, S$, the union of $\tilde{\mathcal{X}}^{(s)}$'s form $\tilde{\mathcal{X}}$, the 3D cube.

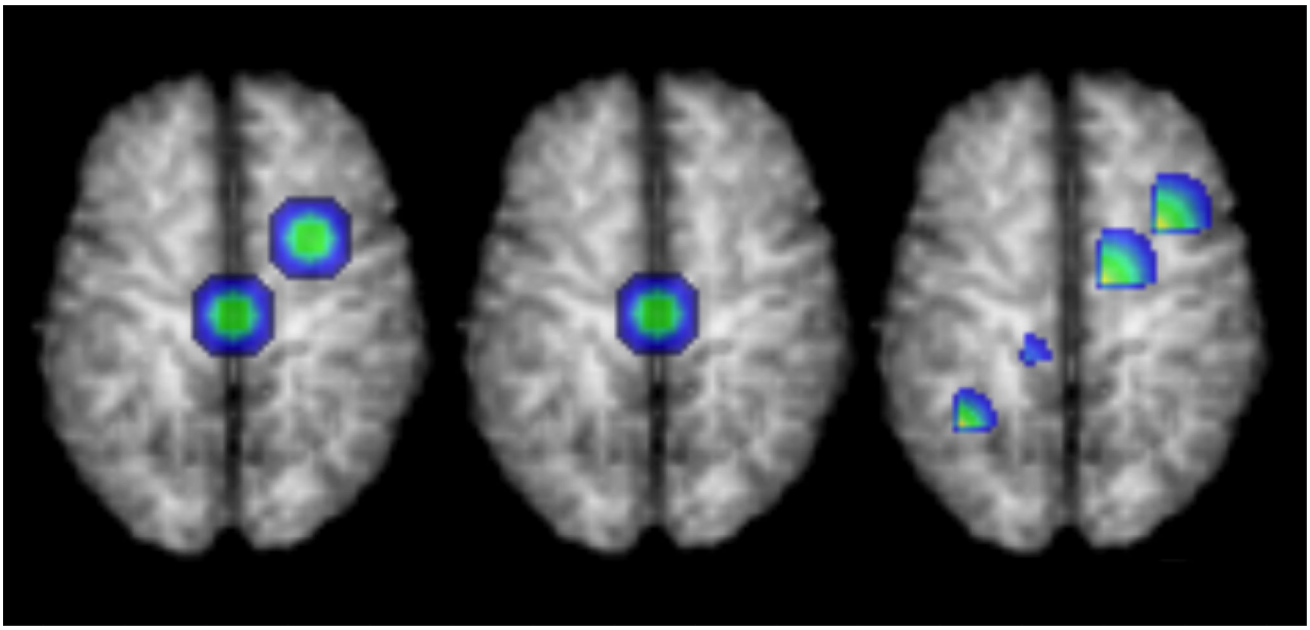

Fig 3.

The 3D rendering of signal $\mathscr{X}_0$ overlaid on the template $\mathscr{G}_0$ for scenarios 1, 2, and 3, respectively. $\mathscr{X}_0$ is equivalent for scenarios 3 and 4.

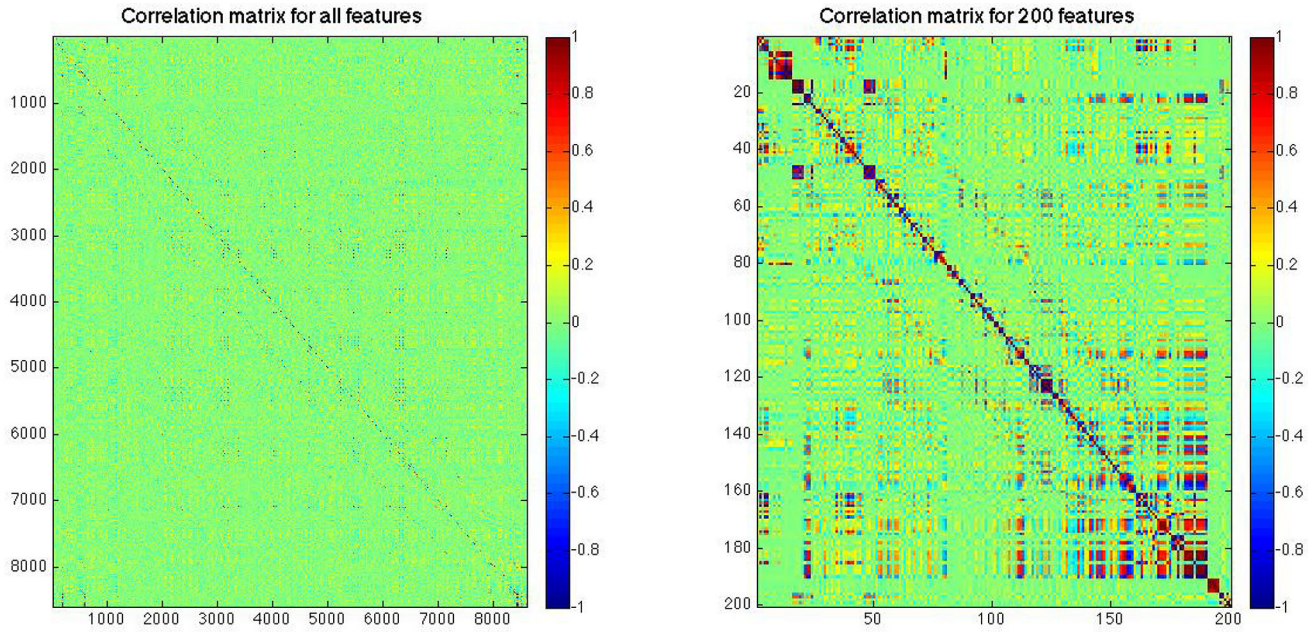

Fig 4.

ADNI data analysis results. The left panel shows the results for the correlation between the columns of the entire feature matrix $\boldsymbol{L}$ obtained in the first screening procedure; the right panel zooms in to shows the same figure for the first 200 features. We observe a high correlation among features in the same partition and among features in neighboring partitions.

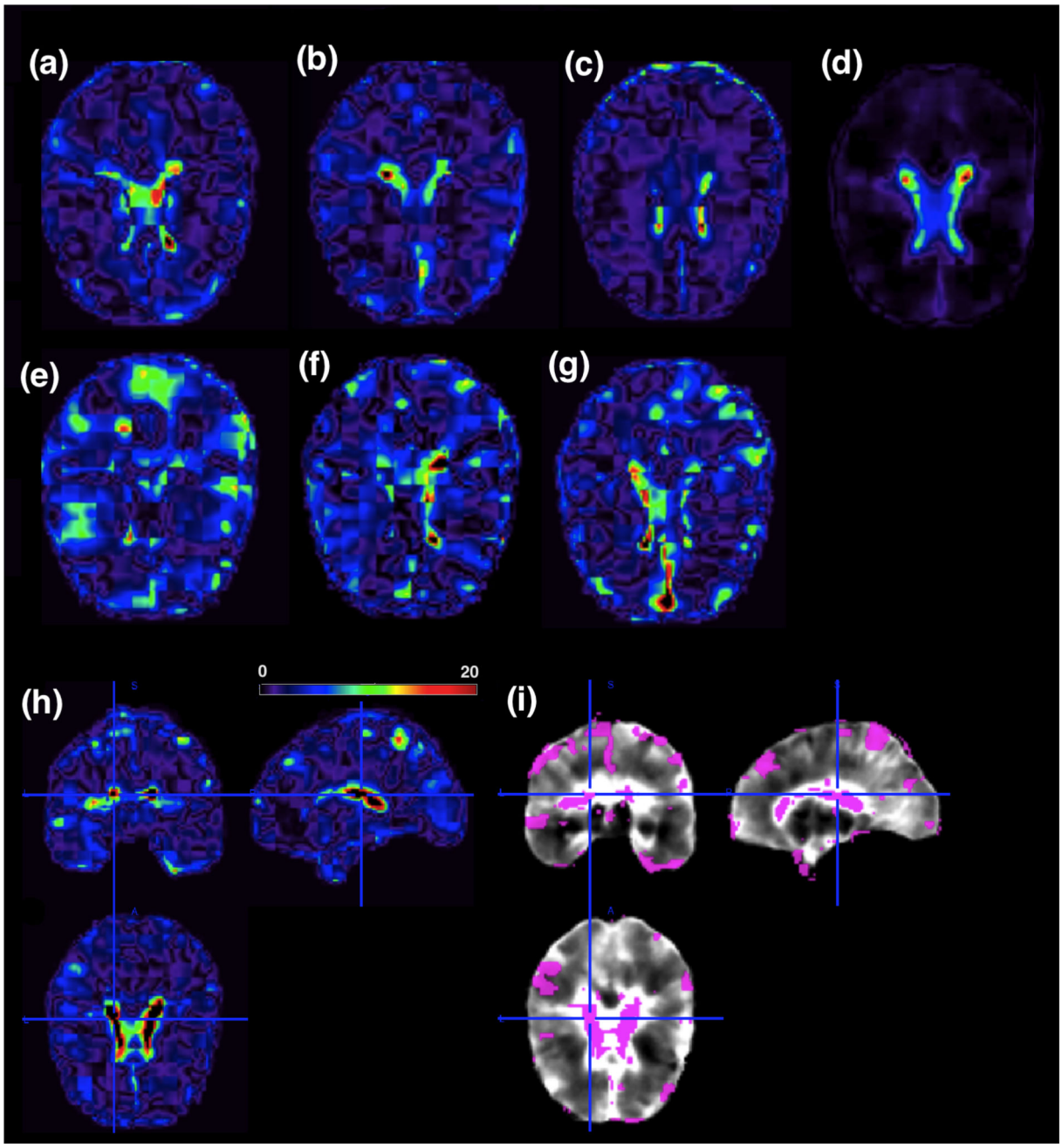

Fig 5.

ADNI data analysis results: panel (a)–(g) show an axial slice of the most important bases projected into the image space. The importance is given by the absolute value of the posterior mean in each one of the 7 selected features. Panel (h) shows the results for the absolute value of the projection $\mathscr{P}$ for the ADNI dataset. Colors on the right side of the colorbar indicate regions where differences are higher between the control group and the Alzheimer's group. Panel (i) shows a threshold of $\mathscr{P}$ with colored parts indicating the biomarkers used to predict the onset of AD.

Table 1

A 3D simulation study results for the average prediction accuracy in multiple runs (Monte Carlo Accuracy), follow by the results of a 10-fold cross validation procedure: prediction accuracy (10-fold Accuracy), false positive rate (FPR), and false negative rate (FNR). The partition model TPRM outperforms TALS and FPCA in all scenarios. For Scenario 3, the models are almost equivalent but TPRM is slightly favored.

		FPCA	TALS	TPRM
Scenario 1	Monte Carlo Accuracy	0.5615	0.5510	**0.8705**
	10-fold Accuracy	0.5750	0.5750	**0.8800**
	10-fold FPR	0	0.3750	0.1496
	10-fold FNR	1.0000	0.5081	0.0322
Scenario 2	Monte Carlo Accuracy	0.5795	0.5830	**0.8925**
	10-fold Accuracy	0.5700	0.6150	**0.9150**
	10-fold FPR	0.0063	0.3817	0.0919
	10-fold FNR	1.0000	0.4494	0.0497
Scenario 3	Monte Carlo Accuracy	0.5095	0.5330	**0.5710**
	10-fold Accuracy	0.5750	0.5700	**0.6100**
	10-fold FPR	0	0.2681	0.4068
	10-fold FNR	1.0000	0.6639	0.3533
Scenario 4	Monte Carlo Accuracy	0.5030	0.5275	**0.6870**
	10-fold Accuracy	0.5750	0.5350	**0.7150**
	10-fold FPR	0	0.3717	0.2764
	10-fold FNR	1.0000	0.5543	0.2724
Scenario 5	Monte Carlo Accuracy	0.7900	0.8220	**0.9415**
	10-fold Accuracy	0.7600	0.8000	**0.9250**
	10-fold FPR	0	0.1597	0.0918
	10-fold FNR	0.5357	0.2273	0.0245
Scenario 6	Monte Carlo Accuracy	0.6455	0.6950	**0.8340**
	10-fold Accuracy	0.5850	0.7450	**0.8250**
	10-fold FPR	0	0.1457	0.1936
	10-fold FNR	0.9667	0.3730	0.1151
Scenario 7	Monte Carlo Accuracy	0.5480	0.5480	**0.6635**
	10-fold Accuracy	0.5750	0.5200	**0.6750**
	10-fold FPR	0	0.3249	0.3427
	10-fold FNR	1.0000	0.6930	0.2651
Scenario 8	Monte Carlo Accuracy	0.6330	0.6260	**0.7430**
	10-fold Accuracy	0.5800	0.6550	**0.7350**
	10-fold FPR	0	0.2691	0.2421
	10-fold FNR	0.9833	0.4152	0.2400

Table 2

ADNI data analysis results: mean prediction accuracy based on a 10-fold cross-validation procedure. The prediction accuracy does not increase as rank increases, and it is bigger for smaller partitions.

Partition size	**96 × 96 × 96**	**24 × 24 × 24**	**12 × 12 × 12**	**6 × 6 × 6**
$R = 5$	0.5320	0.7040	0.6801	**0.9377**
$R = 10$	0.6645	0.6670	0.8108	0.8952
$R = 20$	0.6665	0.6791	0.8231	0.8630
$R = 30$	0.6492	0.7418	0.8487	0.8230