Bayesian Gaussian Process Modeling of Large Scale Longitudinal Neuroimaging Data

Qing He
AI R&D Lab
Samsung SDS America
San Jose, CA 95134
qing.h@samsung.com

Jian Kang
Department of Biostatistics
University of Michigan
Ann Arbor, MI 48109
jiankang@umich.edu

Abstract

We developed a spatial-temporal Gaussian process regression (STGPR) model for Bayesian inference of longitudinal imaging data. Our goal is to study progressions of the brain activities in different brain regions and how they are associated with time-independent predictors (disease status, gender, etc.) and time-varying predictors (age, weight, etc.). We assign Gaussian processes priors to spatial-temporal varying coefficients in the model. To cope with the large-scale dataset, we develop three fast posterior computation algorithms for Bayesian inference approximation based on the Karhunen–Loève expansions on the Gaussian processes. Compared with a voxel-wise linear model approach, we demonstrate the advantages of the proposed method in a simulation study, where we propose two metrics: relative L1 loss and gradients relative L1 loss for measuring coefficient estimation accuracy. We apply the proposed method to the analysis of the longitudinal positron emission tomography (PET) data in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study and obtain some meaningful results.

1 Introduction

Advanced biomedical technology has made more longitudinal neuroimaging data available, where brain images are collected at several time points. Studying brain activities at repeated occasions gains increased attention in the neuroimaging community because it creates an opportunity to further understand the structural and functional development of healthy or pathological brains. The datasets for longitudinal analysis are usually large and complicated, where the response data contains multiple measurements of three-dimensional (3D) brain images for each subject with some of the covariates to be time-varying as well.

The most common methods to fit the longitudinal model is to use the generalized estimating equations (GEE), where a mean function and a working correlation structure between a subject's repeated measurements are assumed, and generalized method of moments (GMM) approaches that model the moments of the statistics of interest [Zeger and Liang, 1986, Lai and Small, 2007, Skup et al., 2012]. However, traditional GEE and GMM are only designed for problems with non-high dimensional spatial correlations, and there are limitations to incorporate time dependent covariates into the model [Sullivan Pepe and Anderson, 1994, Fitzmaurice, 1995]. Machine learning tools have been widely used for brain imaging classification and prediction. Due to partial selection of seed voxels or ignoring the between-region variation, these methods may lead to substantial information loss and subject to misleading understanding in brain functions. Bayesian spatial modeling approaches have been widely proposed to model the correlation between neighboring voxels [Bowman, 2005]. Marquand et al. [2010] evaluated the predictive capability of Gaussian process (GP) models for two types of quantitative prediction: multivariate regression and probabilistic classification, using whole-

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

brain fMRI volumes from a study investigating subjective responses to thermal pain. They show that GP regression outperform support vector and relevance vector regression. However, the difficulties arise for high dimensional problems for which inversion of the correlation matrix is computationally infeasible. In this paper, we proposed a Bayesian nonparametric method via Gaussian processes and utilized approximation techniques through Karhunen–Loève representation of GP kernels to handle the high dimensional problem.

2 Model

Suppose we collect longitudinal imaging data from n subjects at m occasions over the whole three-dimensional (3D) brain $\mathcal{B} \subset \mathbb{R}^d$, where $\mathbb{R}^d$ denotes the $d-$dimensional Euclidean space and $d = 3$. For any given subject $i = 1, \ldots, n$ and for any given occasion $t = 0, 1, \ldots, m$, denote by $y_{i,t}(\mathbf{v}) \in \mathbb{R}$ the longitudinal imaging outcome at voxel $\mathbf{v} \in \mathcal{B}$, $\mathbf{x}_{i,t} = (x_{i,t,1}, \ldots, x_{i,t,p})^{\mathrm{T}} \in \mathbb{R}^p$ a $p-$dimensional vector of time varying covariates, and $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,q})^{\mathrm{T}} \in \mathbb{R}^q$ a $q-$dimensional vector of time independent covariates. For each voxel $\mathbf{v}$ and each occasion t, we assume the longitudinal imaging outcome as a linear function of both time varying and time independent covariates:

$$y_{i,t}(\mathbf{v}) \quad = \quad \alpha_t(\mathbf{v}) + \sum_{j=1}^{p} x_{i,t,j}\beta_j(\mathbf{v}) + \sum_{k=1}^{q} z_{i,k}\eta_{t,k}(\mathbf{v}) + \epsilon_{i,t}(\mathbf{v}). \tag{1}$$

The random error processes is assumed $\epsilon_{i,t}(\mathbf{v}) \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, so that mutually independent over subjects, occasions and voxels. In Equation (1), the spatially-temporally varying intercept $\alpha_t(\mathbf{v}) \in \mathbb{R}$ is the population-level baseline spatial-temporal effects; the spatially varying coefficient $\beta_j(\mathbf{v})$ is the spatial effects of time varying covariate; the spatially-temporally varying coefficient $\eta_{t,k}(\mathbf{v})$ represents the spatial-temporal effects of time independent covariates.

Gaussian processes (GPs) are employed to serve as priors for the spatial-temporal effects in Model (1). For $t = 1, \ldots, m$, $j = 1, \ldots, p$, and $k = 1, \ldots, q$, we assume

$$\begin{aligned}
[\alpha_t \mid \alpha_{t-1}\,] &\overset{\text{iid}}{\sim} \mathcal{GP}(\alpha_{t-1}, \tau_\alpha^2\kappa), \text{and } \alpha_0 \sim \mathcal{GP}(0, \tau_{\alpha,0}^2\kappa), \\
\beta_j &\overset{\text{iid}}{\sim} \mathcal{GP}[0, \tau_\beta^2\kappa], \\
[\eta_{t,k} \mid \eta_{t-1,k}\,] &\overset{\text{iid}}{\sim} \mathcal{GP}[\eta_{t-1,k}, \tau_\eta^2\kappa], \text{and } \eta_{0,k} \sim \mathcal{GP}(0, \tau_{\eta,0}^2\kappa),
\end{aligned} \tag{2}$$

where $\kappa(\mathbf{v}, \mathbf{v}')$ is the covariance kernel function, and $\boldsymbol{\tau}^2 = (\tau_{\alpha,0}^2, \tau_\alpha^2, \tau_\beta^2, \tau_{\eta,0}^2, \tau_\eta^2)$ can be considered as scale parameters that invariant across location. We assume inverse-gamma distribution as priors of $\boldsymbol{\tau}^2$ where the shape and scale parameters of the distribution are hyperparameters.

The standard human brain template usually contains around 200,000 voxels. Thus, posterior inference on the proposed model for a whole brain analysis involves the extremely challenging ultra-high dimensional GP fitting. To mitigate this problem, we adopt the model representation using the Karhunen-Loéve (K-L) expansion. For example, the correlation kernel κ in Equation (2) have the following approximation based on K-L expansion approximation:

$$\kappa(\mathbf{v}, \mathbf{v}') = \sum_{l=0}^{\infty} \zeta_l\psi_l(\mathbf{v})\psi_l(\mathbf{v}') \approx \sum_{l=0}^{L} \zeta_l\psi_l(\mathbf{v})\psi_l(\mathbf{v}'), \tag{3}$$

where ζ_l's are eigen values and $\psi_l(\cdot)$'s are eigen vectors satisfying $\int_{\mathrm{R}^d} \phi_l(\mathbf{v})\phi_{l'}(\mathbf{v})\mathrm{d}\mathbf{v} = \delta_{ll'}$ where $\delta_{ll'} = 1$ if $l = l'$ and $\delta_{ll'} = 0$, otherwise. The prior models on α_t, β_j, and $\eta_{t,k}$ can be respectively represented using the K-L expansion as

$$\alpha_t(\mathbf{v}) \approx \sum_{l=0}^{L} \theta_{\alpha,t,l}\psi_l(\mathbf{v}), \qquad \beta_j(\mathbf{v}) \approx \sum_{l=0}^{L} \theta_{\beta,j,l}\psi_l(\mathbf{v}), \qquad \eta_{t,k}(\mathbf{v}) \approx \sum_{l=0}^{L} \theta_{\eta,t,k,l}\psi_l(\mathbf{v}).$$

When κ_d, the $d-$dimensional κ, has the following special form:

$$\kappa_d(\mathbf{x}, \mathbf{x}'; a, b) = \prod_{i=1}^{d} \exp(-ax_i^2 - b(x_i - x_i')^2 - ax_i'^2),$$

the corresponding ψ_l and ζ_l have explicit forms and are able to be pre-computed. Thus, the new model representation based on K-L expansion have new coefficient parameters $\mathbf{\Theta}_l = (\boldsymbol{\theta}^{\mathrm{T}}_{\alpha,\cdot,l}, \boldsymbol{\theta}^{\mathrm{T}}_{\beta,\cdot,l}, \boldsymbol{\theta}^{\mathrm{T}}_{\eta,\cdot,\cdot,l})^{\mathrm{T}}$, $l = 1, \ldots, L$, each of length $Q = m+p+(m+1)q+1$, where $\boldsymbol{\theta}_{\alpha,\cdot,l} = (\theta_{\alpha,0,l}, \ldots, \theta_{\alpha,m,l})^{\mathrm{T}}$, $\boldsymbol{\theta}_{\beta,\cdot,l} = (\theta_{\beta,1,l}, \ldots, \theta_{\beta,p,l})^{\mathrm{T}}$, $\boldsymbol{\theta}_{\eta,\cdot,\cdot,l} = (\boldsymbol{\theta}^{\mathrm{T}}_{\eta,\cdot,1,l}, \ldots, \boldsymbol{\theta}^{\mathrm{T}}_{\eta,\cdot,q,l})^{\mathrm{T}}$ with $\boldsymbol{\theta}_{\eta,\cdot,k,l} = (\theta_{\eta,0,k,l}, \ldots, \theta_{\eta,m,k,l})^{\mathrm{T}}$. The priors of $\boldsymbol{\theta}_l$'s based on K-L expansion are:

$$\begin{aligned}
\boldsymbol{\theta}_{\alpha,\cdot,l} &\sim \mathrm{N}[\mathbf{0}_{m+1}, \zeta_l(\tau^2_{\alpha,0}\mathbf{J}_{m+1} + \tau^2_{\alpha}\mathbf{K}_{m+1})], \\
\boldsymbol{\theta}_{\beta,\cdot,l} &\sim \mathrm{N}(\mathbf{0}_p, \zeta_l\tau^2_{\beta}\mathbf{I}_p), \\
\boldsymbol{\theta}_{\eta,\cdot,l} &\sim \mathrm{N}[\mathbf{0}_{(m+1)q}, \zeta_l\mathbf{I}_q \otimes (\tau^2_{\eta,0}\mathbf{J}_{m+1} + \tau^2_{\eta}\mathbf{K}_{m+1})],
\end{aligned}$$

where $\mathbf{0}_d = (\underbrace{0, \ldots, 0}_{d})^{\mathrm{T}}$, $\mathbf{1}_d = (\underbrace{1, \ldots, 1}_{d})^{\mathrm{T}}$, $\mathbf{I}_d = \mathrm{diag}(\mathbf{1}_d)$, $\mathbf{J}_d = \mathbf{1}_d\mathbf{1}^{\mathrm{T}}_d$ and $\mathbf{K}_d = (k_{ij})_{1\le i,j\le d}$ with $k_{ij} = \min\{i, j\} - 1$. Future more, defined by $\mathbf{y}_i(\mathbf{v}) = [y_{i,0}(\mathbf{v}), \ldots, y_{i,m}(\mathbf{v})]^{\mathrm{T}}$ an $m+1$ vector of the outcome variable, $\mathbf{D}_i(\mathbf{v}) = [\mathbf{I}_{m+1}, \mathbf{X}_i, \mathbf{z}^{\mathrm{T}}_i \otimes \mathbf{I}_{m+1}]$ the design matrix of dimension $(m+1) \times Q$, where "$\otimes$" is the kronecker product, $\mathbf{X}_i = (\mathbf{x}_{i,0}, \ldots, \mathbf{x}_{i,m})^{\mathrm{T}}$ of dimension $(m+1) \times p$ with $\mathbf{x}_{i,t} = (x_{i,t,1}, \ldots, x_{i,t,p})^{\mathrm{T}}$, $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,q})^{\mathrm{T}}$, Model (1) can be written as the following matrix form under new representation:

$$\mathbf{y}_i(\mathbf{v}) \quad = \quad \mathbf{D}_i(\mathbf{v}) \sum_{l=1}^{\infty} \psi_l(\mathbf{v})\mathbf{\Theta}_l + \boldsymbol{\epsilon}_i(\mathbf{v}), \tag{4}$$

where $\boldsymbol{\epsilon}_i(\mathbf{v}) \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}_{m+1}, \sigma^2\mathbf{I}_{m+1})$ with $\boldsymbol{\epsilon}_i(\mathbf{v}) = [\epsilon_{i,0}(\mathbf{v}), \ldots, \epsilon_{i,m}(\mathbf{v})]^T$, and $\mathbf{\Theta}_l \overset{\text{iid}}{\sim} \mathrm{N}(\mathbf{0}_Q, \mathbf{\Sigma}_l)$ with

$$\mathbf{\Sigma}_l = \zeta_l \mathrm{diag}\{(\tau^2_{\alpha,0}\mathbf{J}_{m+1} + \tau^2_{\alpha}\mathbf{K}_{m+1}), \tau^2_{\rho}, \tau^2_{\beta}\mathbf{I}_p, \mathbf{I}_q \otimes (\tau^2_{\eta,0}\mathbf{J}_{m+1} + \tau^2_{\eta}\mathbf{K}_{m+1})\}.$$

3 Posterior Computation

Gibbs sampling are used for posterior computation. Three algorithms are developed to achieve computation efficiency by utilizing properties of K-L expansion.

Algorithm I uses Gibbs sampling to sample $\mathbf{\Theta} = (\mathbf{\Theta}^T_1, \ldots, \mathbf{\Theta}^T_L)^T$ from conditional full distribution $\pi(\mathbf{\Theta}|\mathbf{y}, \boldsymbol{\tau}^2_{\cdot}, \sigma^2)$ directly, which turns out to be Gaussian. Computing the mean and variance of the posterior Gaussian distribution involves calculating matrix inverse of scale $Q \times L$.

Algorithm II iteratively draws $\mathbf{\Theta}_l$ from its full conditional distribution $\pi(\mathbf{\Theta}_l \mid \mathbf{y}, \mathbf{\Theta}_{-l}, \boldsymbol{\tau}^2, \sigma^2)$, where $\mathbf{\Theta}_{-l} = \{\mathbf{\Theta}_r, r \in [1, \ldots, L], r \neq l\}$. Computing the mean and variance of the posterior Gaussian distribution involves calculating matrix inverse of scale Q.

Algorithm III takes advantage of the orthogonality of ψ_l to reduce computation complexity of calculating $\pi(\mathbf{\Theta}_l \mid \mathbf{y}, \mathbf{\Theta}_{-l}, \boldsymbol{\tau}^2, \sigma^2)$ to however at a cost of applying additional approximation.

4 Simulation and Real Data Analysis

To show the performance of our model, we conducted Monte Carlo simulation studies and compared the proposed spatial temporal Gaussian process regression method (STGPR) with a basic linear regression model (LM). The LM is also based on Model (1) but assumes mutual independence for $\alpha_t(\mathbf{v})$, $\beta(\mathbf{v})$ and $\eta_t(\mathbf{v})$ over different occasions and voxels. Therefore, those coefficients are estimated independently by LM at each voxel and at different time occasion. Point and interval estimates of LM are provided by least square estimators. Two time-dependent covariates are generated by $x_{i,1}(t) = 1.5 \times t + \epsilon_{i,1,t}$ and $x_{i,2}(t) = 2.5 \times t + \epsilon_{i,2,t}$ respectively, where $\epsilon_{i,j,t} \sim_{i,t} \mathcal{N}(0, 1), j = 1, \ldots, p$. These time-dependent covariates are considered as continuous variables and are standardized over subjects and over time before being incorporated into the model. In addition, two time-independent covariates, $z_{i,t,1}$ and $z_{i,t,2}$ are sampled independently from Bernoulli distribution with success probability to be 0.5 and 0.1. The coefficients, $\alpha_t, \beta_j, \eta_{t,k}$ are generated based on equation (2) and $\kappa(\mathbf{v}, \mathbf{v}') = e^{-\frac{d(\mathbf{v},\mathbf{v}')}{s}}$ with $d(\mathbf{v}, \mathbf{v}')$ calculating the l^2 distance between in $\mathcal{B}$. The spatial data are equally spread on two dimensional space of $[-1.8, 1.8]^2$. 20 samples are simulated and each measured at four sequential time occasions. The experiments are repeated 50 times. Two metrics

are generated to show the performance of STGPR compared to LM in regards to both coefficients estimation and image outcome prediction. The relative L_1 loss represents the L_1 penalty of the relative values of estimates and the gradients' relative L loss, which is computed for the gradients' of estimates, quantifies the L_1 penalty of estimated smoothness. For this extend abstract, please find our preliminary results in Table (1). Both LM and STGPR provide reasonable parameter and outcome estimation and prediction. The LM model is more influenced by random noise, while STGPR model provides stabler estimation with smaller standard errors of RL_1 and GRL_4 than LM model. The STGPR model results in a much smaller GRL_4 than the LM model for all the parameters and outcome variables, indicating that the smoothness of the estimated curves by the proposed method are much closer to the truth. As a result, the proposed STGPR model provides stable outcome prediction and reliable estimates of the covariates' spatial and temporal effects.

Table 1: Simulation results comparing LM and STGPR in regards to estimation and prediction of both coefficients and outcome variables through RL_1 Loss(relative L_1 Loss) and GRL_1 (gradients' relative L_1 Loss) averaging over space and necessarily over time. For STGPR, Algorithm II is used on simulated data of size 60×60 and Algorithm III is used for simulated data of size 100×100. The values in the parenthesis are standard error of the quantities of interest.

Algorithm II	**Size**(60×60)			
	LM		STGPR	
	RL_1	$GRL_1(\times 10)$	RL_1	$GRL_1(\times 10)$
α	1.06(0.07)	244.66(434.68)	0.76(0.02)	9.52(1.10)
β_1	1.31(0.25)	10.91(7.68)	1.09(0.04)	1.13(0.39)
β_2	0.73(0.07)	12.00(2.53)	0.77(0.02)	1.99(0.65)
η_1	0.94(0.06)	7.52(6.55)	1.17(0.02)	0.97(0.54)
η_2	0.61(0.05)	3.15(0.69)	0.42(0.01)	0.45(0.01)
f	0.37(0.01)	3.92(1.89)	1.01(0.01)	1.84(0.56)
α_{pred}	5.15(0.36)	9.92(16.71)	4.85(0.06)	1.12(0.41)
$\eta_{1,\text{pred}}$	3.12(0.16)	24.40(20.02)	3.22(0.03)	5.60(4.08)
$\eta_{2,\text{pred}}$	5.11(0.96)	15.35(36.78)	4.31(0.10)	5.53(0.98)
y_{pred}	2.57(0.21)	7.39(26.32)	2.93(0.01)	1.51(0.76)
Algorithm III	**Size**(100×100)			
	LM		STGPR	
	RL_1	$GRL_1(\times 10)$	RL_1	$GRL_1(\times 10)$
α	1.20(0.15)	10.79(17.30)	0.60(0.02)	1.21(3.30)
β_1	1.02(0.08)	159.41(1078.06)	0.52(0.03)	1.28(4.83)
β_2	0.65(0.06)	6.70(7.88)	0.25(0.01)	0.60(0.41)
η_1	1.38(0.23)	2.53 (0.87)	0.41(0.03)	0.32(0.05)
η_2	0.59(0.05)	2.37(2.07)	0.43(0.01)	0.27(0.08)
f	1.17(0.64)	2.11(0.53)	0.81(0.15)	0.98(0.94)
α_{pred}	3.13(0.15)	3.37 (3.32)	2.73(0.05)	0.72(0.54)
$\eta_{1,\text{pred}}$	4.91(0.44)	15.60(7.61)	4.82(0.08)	7.36(25.39)
$\eta_{2,\text{pred}}$	4.06(0.73)	6.30(9.66)	3.87(0.13)	2.03(0.53)
y_{pred}	2.16(0.09)	1.56(0.68)	2.19(0.02)	0.69(0.57)

In order to show the usefulness of the proposed STGPR, we used the model to analyze PET images of patients from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study (http://www.loni.ucla.edu/ADNI/). The goal of this national multi-center project is to develop biomarkers of Alzheimer's Disease (AD) in elderly subjects. For more details about the ADNI, see Mueller et al. (2005). Participants are classified into typical controls (TC), mild cognitive impairment (MCI) patients, and Alzheimer's disease (AD) patients. PET scans are obtained for each participants at baseline screening, 6 months, and 12 months. The data we used contain one slice of PET scan images from 49 AD, 69 NORM, and 117 MCI participants and at three time occasions. The slice of brain images are chosen to cover most of the hippocampus regions, which are known by researchers to be one of the first brain area to suffer damage of AD so that have good indication of AD initiation and development. For predictors, we incorporated gender and diagnostic status (TC, AD, MCI) as time independent covariates and used age and weight to be time varying covariates. The proposed STGPR was used to investigate the spatial and temporal influence of different covariates and also make prediction on image outcomes given previous images and all covariates.

References

Anjishnu Banerjee, David B Dunson, and Surya T Tokdar. Efficient gaussian process regression for large datasets. *Biometrika*, page ass068, 2012.

F. D Bowman. Spatio-temporal modeling of localized brian activity. *Biostatistics*, 6:558–575, 2005.

Shuo Chen and F DuBois Bowman. A novel support vector classifier for longitudinal high-dimensional data and its application to neuroimaging data. *Statistical Analysis and Data Mining*, 4 (6):604–611, 2011.

Ciprian M Crainiceanu, Ana-Maria Staicu, and Chong-Zhi Di. Generalized multilevel functional regression. *Journal of the American Statistical Association*, 104(488):1550–1561, 2009.

Chong-Zhi Di, Ciprian M Crainiceanu, Brian S Caffo, and Naresh M Punjabi. Multilevel functional principal component analysis. *The annals of applied statistics*, 3(1):458, 2009.

Garrett M Fitzmaurice. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*, pages 309–317, 1995.

Karl J Friston, Jean-Baptiste Poline, Andrew P Holmes, Chris D Frith, and Richard SJ Frackowiak. A multivariate analysis of pet activation studies. *Human brain mapping*, 4(2):140–151, 1996.

Nacer Kerrouche, Karl Herholz, Rüdiger Mielke, Vjera Holthoff, and Jean-Claude Baron. 18fdg pet in vascular dementia: differentiation from alzheimer's disease using voxel-based multivariate analysis. *Journal of Cerebral Blood Flow & Metabolism*, 26(9):1213–1221, 2006.

Tze Leung Lai and Dylan Small. Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):79–99, 2007.

Yimei Li, Hongtu Zhu, Yasheng Chen, Hongyu An, John Gilmore, Weili Lin, and Dinggang Shen. Lstgee: Longitudinal analysis of neuroimaging data. In *SPIE Medical Imaging*, pages 72590F–72590F. International Society for Optics and Photonics, 2009.

Yisheng Li, Xihong Lin, and Peter Müller. Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics*, 66(1):70–78, 2010.

Andre Marquand, Matthew Howard, Michael Brammer, Carlton Chu, Steven Coen, and Janaina Mourão-Miranda. Quantitative prediction of subjective pain intensity from whole-brain fmri data using gaussian processes. *Neuroimage*, 49(3):2178–2189, 2010.

Philip T Reiss and R Todd Ogden. Functional generalized linear models with images as predictors. *Biometrics*, 66(1):61–69, 2010.

Martha Skup, Hongtu Zhu, and Heping Zhang. Multiscale adaptive marginal analysis of longitudinal neuroimaging data with time-varying covariates. *Biometrics*, 68(4):1083–1092, 2012.

Margaret Sullivan Pepe and Garnet L Anderson. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation*, 23(4):939–951, 1994.

Xuejing Wang, Bin Nan, Ji Zhu, and Robert Koeppe. Regularized 3d functional regression for brain image data via haar wavelets. *The annals of applied statistics*, 8(2):1045, 2014.

Shun Xu, Martin Styner, John Gilmore, Joseph Piven, and Guido Gerig. Multivariate nonlinear mixed model to analyze longitudinal image data: Mri study of early brain development. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008.

Scott L Zeger and Kung-Yee Liang. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pages 121–130, 1986.

Yihong Zhao, R Todd Ogden, and Philip T Reiss. Wavelet-based lasso in functional linear regression. *Journal of Computational and Graphical Statistics*, 21(3):600–617, 2012.