Bayesian Analysis of Varying Coefficient Models and Applications

Zhaowei Hua

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2011

Approved by:

Hongtu Zhu

David B. Dunson

Fei Zou

Wei Sun

Sy-Miin Chow

Abstract

ZHAOWEI HUA: Bayesian Analysis of Varying Coefficient Models and Applications.
(Under the direction of Hongtu Zhu and David B. Dunson.)

The varying coefficient models have been very important analytic tools to study the dynamic pattern in biomedicine fields. Since nonparametric varying coefficient models make few assumptions on the specification of the model, the 'curse of dimensionality' is an very important issue. Nonparametric Bayesian methods combat the curse of dimensionality through specifying a sparseness-favoring structure. This is accomplished through the Bayesian penalty for model complexity (Jeffreys and Berger, 1992) and is aided through centering on a base Bayesian parametric model. This dissertation presents three novel semiparametric Bayesian methods for the analysis of longitudinal data, diffusion tensor imaging data, and longitudinal circumplex data.

In longitudinal data analysis, we propose a semiparametric Bayes approach to allow the impact of the predictors to vary across subjects, which allows flexibly local borrowing of information across subjects. Local hypothesis testing and confidence bands are developed for the identification of time windows for significant predictor impact, adjusting for multiple comparisons. The methods are assessed using simulation studies and applied to a yeast cell-cycle gene expression data set.

In analyzing diffusion tensor imaging data, we propose a semiparametric Bayesian local functional model to connect multiple diffusion properties along white matter fiber bundles with a set of covariates of interest. An LPP2 prior facilitates global and local borrowing of information among subjects, while an infinite factor model flexibly represents low-dimensional structure. Local hypothesis testing and confidence bands are

developed to identify fiber segments for significant association of covariates with multiple diffusion properties, controlling for multiple comparisons. The method is assessed by a simulation study and illustrated via two fiber tract data sets for neurodevelopment.

In analyzing longitudinal circumplex data, we propose a semiparametric Bayesian infinite state-space circumplex model to capture the dynamic transition pattern of affective experience, where affects are characterized as an ordering on the circumference of a circle. A sticky infinite state hidden Markov model via hierarchical Dirichlet proces is used to address the time related state-switching structure and the self-transition feature. The method is assessed by a simulation study and an emotion data set for the dynamics of emotion regulation.

Acknowledgments

The period of my dissertation is a precious component in my life that I benefit from many people in completing this dissertation and get inspired in problem solving.

I want to express my deepest gratitude to my advisors, Dr. Hongtu Zhu and Dr. David B. Dunson. Both of them are the shining models as world-class researchers. Their broad visions, deep insights, and high research standards are the leading sources in my growth throughout my dissertation research. I particularly appreciate the generous financial support from Dr. Hongtu Zhu, as well as the support from Dr. David B. Dunson in funding application.

I want to sincerely thank my committee members Dr. Sy-Miin Chow, Dr. Fei Zou and Dr. Wei Sun. I appreciate Dr. Sy-Miin Chow for her suggestions in improving my 3rd topic, providing the circumplex data, and her financial support. I thank Dr. Fei Zou and Dr. Wei Sun for commenting my views, reading my dissertation, and suggestions in improving my dissertation. I also owe my thanks to Dr. Wei Sun for his friendship and his personal support.

I am greatly thankful to Dr. Joseph Ibrahim for his generous financial support with NCI training grant program in genomics and cancer. I particularly appreciate Dr. Michael Kosorok for offering me a graduate fellowship at SAMSI and Dr. Amy Herring for her guidance at SAMSI.

I am deeply indebted to Dr. Jianwen Cai for introducing me into the Biostatistics research area in advising my master paper and her encouragements.

I owe my special thanks to Dr. Todd Schwartz for his guidance in diverse exciting projects at School of Nursing and his encouragement in my study at UNC with his professional and personal support. I am heartily thankful to Dr. William Kalsbeek for offering me an opportunity to work with him on an interesting sampling design project and his professional support. I am grateful to Dr. Diane holditch-davis and Dr. Mark Weaver for the initial opportunity at School of Nursing as a research assistant.

My sincere thanks also goes to Dr. Suchindran for offering me the admission opportunity to this prestigious department and his professional support.

Many friends and schoolmates have helped me weather the up and down days of my graduate student life. I greatly value their friendships. In particular, I want to thank all the group members in our BIAS lab where we are encouraged to pursue deep knowledge and understanding.

Finally and most importantly, my lovely words go to my husband Jiansuo. His persistent love and patience has been the pillar throughout my graduate life. I particularly thank my fellow brothers and sisters in God who have made home away from home. I thank my parents and sister's family back home for their unconditional love and support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CRF	Chinese restaurant franchise
DP	Dirichlet process
DTI	Diffusion tensor imaging
FA	Fractional anisotropy
FDR	False discovery rate
FNR	False negative rate
HDP	Hierarchical Dirichlet process
iHMM	Infinite state hidden Markov model
LPP	Local partition process
MCMC	Markov chain Monte Carlo
MD	Mean diffusivity
MDS	Multidimensional scaling
MRI	Magnetic resonance imaging
MSE	Mean squared error
PANAS	Positive and negative affect schedule
SBLFM	Semiparametric Bayesian local functional models
TFs	Transcription factors
VM	Von Mises distribution

Chapter 1

Introduction

Varying coefficient models are a class of widely developed models. It was initially introduced by Cleveland et al. (1991) to extend the applications of local regression techniques from one-dimensional to multidimensional setting. Local regression models were further extended by Hastie and Tibshirani (1993) to be formulated as varying coefficient models, which also tied generalized additive models and dynamic generalized linear model into one framework. Semivarying coefficient models (Zhang et al., 2002), mix-effect varying coefficient models (Liang et al., 2003), and random varying-coefficient models (Wu and Liang, 2004) were later developed to address different situations. Semivarying coefficient models allow the linear effects of some predictors on the outcomes. Random varying-coefficient models assumes the subject-specific varying coefficient effects, while mix-effect varying coefficient models allow both fixed varying coefficient effects and random varying coefficient effects.

Varying coefficient models arise from various statistical contexts. The vast volume of literature includes, among many others, Nicholls and Quinn (1982), Chen and Tsay (1993) and Cai et al. (2000b) on non-linear time series, Hoover et al. (1998), Wu et al. (1998) and Fan and Zhang (2000a) on longitudinal data analysis, Ramsay and Silverman (1997) on functional data analysis, Gelfand and Vounatsou (2003), Baladandayuthapani

et al. (2008) on spatial data analysis, Lu et al. (2009) and Brezger et al. (2007a) on varying coefficient spatiotemporal model, Cleveland et al. (1991), Hastie and Tibshirani (1993), Carroll et al. (1998), Kauermann and Tutz (1999), Xia and Li (1999), Zhang and Lee (2000) and Fan and Zhang (1999, 2000b) on local multidimensional regression, and Cai et al. (2000a) on generalized linear models with varying coefficients.

In this dissertation, we first review the existing literature in the rest of this chapter. In Chapter 2, we propose semiparametric Bayes local additive models for the analysis of longitudinal data. In Chapter 3, we develop a multivariate semiparametric Bayesian local factor functional regression framework to analyze fiber tract data. In Chapter 4, a semiparametric Bayesian state-space time-varying circumplex model is introduced to analyze longitudinal circumplex data.

1.1 Frequentist Analysis of Varying Coefficient Models

There are currently two major approaches to estimate the varying coefficients, which are kernel-local polynomial smoothing (Wu et al., 1998; Hoover et al., 1998; Fan and Zhang, 1999; Kauermann and Tutz, 1999), and spline smoothing (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001; Huang et al., 2002, 2004; Huang and Shen, 2004). We are going to outline these two approaches in the following subsections.

1.1.1 Kernel-local Polynomial Smoothing

Kernel local polynomial smoothing is a richly developed methodology associated with a large amount of contributing literature; see for example Fan and Gijbels (1992, 1995, 1996), Fan (1993), Ruppert and Wand (1994), Fan et al. (1995), and Ruppert (1997).

Its main idea is based on Taylor expansion to locally approximate a smooth function by a polynomial of some degree. Assume $f(t)$ is a smooth function with a $(p+1)$-st continuous derivative for some positive integer p at an arbitrary fixed point t_0. The

$f(t)$ can be locally approximated by a polynomial of degree p around a neighborhood of t_0, as in the form of Taylor expansion:

$$f(t) \approx f(t_0) + (t - t_0)f^{(1)}(t_0) + \cdots + (t - t_0)^p f^{(p)}(t_0)/p!,$$

where $f^{(r)}(t_0)$ is the r-th derivative of $f(t)$ at t_0.

Suppose that we have a set of data $\{U_i, X_i^T, y_i : i = 1, \ldots, n\}$ sampled from (U, X^T, y), we fit a varying-coefficient model of the form

$$y = X^T \boldsymbol{a}(U) + \epsilon,$$

where $E(\epsilon) = 0$, and $Var(\epsilon) = \sigma^2(U)$. The kernel local polynomial estimator $\hat{\boldsymbol{a}}(U)$ of $\boldsymbol{a}(U)$ is obtained by minimizing

$$L(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n} \{y_i - X_i^T \boldsymbol{a} - X_i^T \boldsymbol{b_1}(U_i - u) - \cdots - X_i^T \boldsymbol{b_p}(U_i - u)\}^2 K_h(U_i - u),$$

where $K_h(t) = K(t/h)/h$, $K(t)$ is a kernel function, such as Epanechnikov kernel $K(t) = 0.75(1 - t^2)_+$ and h is bandwidth.

Denote $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, $\boldsymbol{U}_u = (U_1 - u, \ldots, U_n - u)$, $\Gamma_u = (\boldsymbol{X}, \boldsymbol{U}_u\boldsymbol{X}, \ldots, \boldsymbol{U}_u^p\boldsymbol{X})$, $Y = (y_1, \ldots, y_n)^T$, and $W_u = \text{diag}(K_h(U_1 - u), \ldots, K_h(U_n - u))$. The one-stage kernel local polynomial estimator $\hat{\boldsymbol{a}}(U)$ is given by

$$\hat{\boldsymbol{a}}(U) = (1_p^T \otimes I_q, \boldsymbol{0}_q)(\Gamma_u^T W_u \Gamma_u)^{-1} \Gamma_u^T W_u Y.$$

Under some conditions, $\hat{\boldsymbol{a}}(U)$ is asymptotically normally distributed:

$$cov^{-1/2}(\hat{a}(u))\{\hat{a}(u) - a(u) - bias(\hat{a}(u))\} \xrightarrow{D} N(0, I_q),$$

where $bias(\hat{a}(u)) = 2^{-1}\mu_2 \boldsymbol{a}^{(2)}(u)h^2$, $cov(\hat{a}(u)) = \{nhf(u)E(XX^T|U = u)\}^{-1}\nu_0\sigma^2(u)$, and $\xrightarrow{D}$ denotes convergence in distribution. The bias of $\hat{\boldsymbol{a}}(U)$ comes from the approximation error of the linear approximation of $\boldsymbol{a}(u)$. In the asymptotic covariance matrix of $\hat{\boldsymbol{a}}(U)$, the $2hf(u)$ is approximately the probability of U falling into the neighbourhood of u with radius h, and $2nhf(u)$ is approximately the expected number of U_i in the neighbourhood of u.

Bandwidth selection is an important issue in kernel smoothing. Larger bandwidth may gain on variance side, but loses on bias side. Smaller bandwidth may gain on bias side, but loses on variance side. The basic idea of a data driven bandwidth selection procedure is to find an estimator of mean squared error (MSE) of $\hat{\boldsymbol{a}}(U)$ first, and then minimizes MSE with respect to bandwidth. The optimal bandwidth is the one minimizing the MSE. Wu et al. (1998) and Hoover et al. (1998) proposed to use cross-validation to select the bandwidth. Zhang and Lee (2000) systematically investigated both variable bandwidth and constant bandwidth selection. For longitudinal data, it is better to delete a whole subject rather than just a single observation when estimating MSE(h).

1.1.2 Spline Smoothing

Three types of spline approaches have been developed for varying coefficient models. Such approaches include regression spline (Shi et al., 1996; Rice and Wu, 2001; Wu and Zhang, 2002; Liang et al., 2003), penalized spline (Eilers and Marx, 1996; Kauermann, 2005) and smoothing spline (Hoover et al., 1998; Brumback and Rice, 1998; Chiang et al., 2001; Eubank et al., 2004). The key idea of these approaches is to express varying coefficient functional effect as a linear combination of some spline bases. However, they differ from each other in the number of knots and penalization. Regression spline methods and penalized spline methods usually use fewer knots than smoothing spline

methods. The smoothness of regression spline methods is controlled by the number of the associated basis functions. When the number of pre-specified basis functions is large, it is very challenging to fit the regression spline models due to high-dimensionality and roughness. The penalized spline method uses two truncated power bases and penalizes the high order derivative of the associated regression splines. In contrast, the smoothing spline method penalizes the roughness of the resultant functions. For the sake of space, we are going to briefly review the smoothing spline method as follows.

Suppose that we have a set of longitudinal sample $(y_{ij}, x_{ij}, z_{ij}, t_{ij})$, for $i = 1, \ldots, n$, and $j = 1, ..., n_i$, where $x_i(t_{ij}) = (x_{ij0}, \ldots, x_{ijK})^T$, $z_i(t_{ij}) = (z_{ij0}, \ldots, z_{ijM})^T$ and $(y_{ij}, x_{ij}, z_{ij}, t_{ij})$ denote the jth outcome, two set of covariates and time design points, respectively, of the ith subject. We fit a nonparametric mixed effects time-varying coefficient model:

$$y_{ij} = x_i(t_{ij})^T \beta(t_{ij}) + z_i(t_{ij})^T v_i(t_{ij}) + \epsilon_i(t_{ij}), \tag{1.1}$$

where $\beta(t) = (\beta_0(t), \ldots, \beta_K(t))^T$ are smooth fixed effects of interest, $v_i(t) = (v_{i0}(t), \ldots, v_{iM}(t))^T$ are random effects sampled from a Gaussian process with zero mean and Γ covariance function, $\epsilon_i(t)$ is a zero-mean stochastic process, and $v_i(t)$ and $\epsilon_i(t)$ are independent. We use the cubic smoothing spline method to estimate the fixed effects and the random effects:

$$\beta_k(t_{ij}) = \mathbf{h}_{ij}^T \beta_k, \quad k = 0, 1, \ldots, K,$$
$$v_{im}(t_{ij}) = \mathbf{h}_{ij}^T \mathbf{v}_{im}, \quad m = 0, 1, \ldots, M,$$

where $\beta_k = (\beta_k(\tau_1), \ldots, \beta_k(\tau_L))^T$ denotes the values of $\beta_k(t)$ at a set of design knots $\tau_1, \ldots, \tau_L$, $\mathbf{v}_{im} = (\mathbf{v}_{im}(\tau_1), \ldots, \mathbf{v}_{im}(\tau_L))^T$ denotes the values of $v_{im}(t)$ at $\tau_1, \ldots, \tau_L$, and $\mathbf{h}_{ij}$ denote a L-dimensional unit vector whose l-th entry is 1 if $t_{ij} = \tau_l$ and 0 otherwise.

Let $\boldsymbol{\beta} = (\beta_0^T, \ldots, \beta_K^T)^T$, $\mathbf{v}_i = (\mathbf{v}_{i0}^T, \ldots, \mathbf{v}_{iM}^T)^T$, $\mathbf{x}_{ij} = x_i(t_{ij}) \otimes \mathbf{h}_{ij}$, $\mathbf{z}_{ij} = z_i(t_{ij}) \otimes \mathbf{h}_{ij}$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^T$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})^T$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$. Then the nonparametric mixed effect time-varying coefficient model (1.1) can be expressed as

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{v}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(0, \mathbf{R}_i), \quad \mathbf{v}_i \sim N_{(M+1)L}(0, \mathbf{D}).$$

If the components of $\beta(t)$ and $v_i(t)$s are twice continuously differentiable and their second derivatives are square integrable, the cubic spline smoothing estimator $\hat{\beta}_k(t)$ of $\beta_k(t)$ is the least square minimizer of the following penalized generalized log-likelihood:

$$\begin{aligned}\mathbf{J}(\beta, \{\mathbf{v}_i\}_{i=1}^n, \lambda) &= \sum_{i=1}^{n}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{v}_i)^T\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{v}_i) \\ &\quad + \sum_{i=1}^{n}\sum_{m=0}^{M}\lambda_{im}\int_a^b [v_{im}''(t)]^2 dt + \sum_{k=0}^{K}\lambda_k\int_a^b [\beta_k''(t)]^2 dt,\end{aligned}$$

where λ_k, $k = 0, \ldots, K$, λ_{im}, $i = 1, \ldots, n$, $m = 0, \ldots, M$, are the positive smoothing parameters. The first summation term being proportional to the twice negative logarithm of the generalized likelihood represents the goodness of fit. The second term is the weighted sum of the roughness of all the random-effect coefficient functions, and the third term is the weighted sum of the roughness of all the fix-effect coefficient functions. The smoothing parameters λ_k, $k = 0, \ldots, K$, and λ_{im}, $i = 1, \ldots, n$, $m = 0, \ldots, M$, are used to trade off the goodness of fit with the roughness of the smoothing spline estimators.

A major problem of smoothing spline technique is the choice of the optimal smoothing parameters to achieve the best performance of the resulting estimators. The "leave-one-subject-out" cross-validation (SCV) rule (Rice and Silverman, 1991; Hoover et al., 1998) is the most popular smoothing parameter selector. The idea is to minimize the

weighted sum of mean square errors associated with deletion of each subject, called SCV rule, with respect to the smoothing parameters. The advantages of SCV include: (i) deletion of the entire data for one subject at a time preserves the within-subject correlation of the data; (ii) SCV does not require to specify the within-subject correlation structure. However, it is challenging to minimize SCV rule when the number of smoothing parameters is large. Alternatively, the "leave-one-point-out" cross-validation (PCV) rule is used in Wang and Taylor (1995) and Eubank et al. (2004) among many others. The key advantage of PCV is less computationally intensive with no need to repeatedly compute the estimators. The drawback of PCV is the lack of accounting for the within-subject correlation.

1.2 Bayesian Analysis of Varying Coefficient Models

There are two major approaches to model varying coefficients from nonparametric Bayesian perspective, which are basis expansion methods (Rajan and Rayner, 1996; Lee and Shaddick, 2007; Brezger et al., 2007b; Huang et al., 2008; Malloy et al., 2010) and stochastic process mothods (Marina et al., 2008; He et al., 2010; Berrocal et al., 2010; Reich et al., 2010). Bayesian basis expansion methods approximately express any function g as $g(\cdot) = \sum_h b_h f_h(\cdot)$ by some chose basis $f = \{f_1, f_2, f_3, \ldots\}$, such as spline, wavelet and Fourier bases. The random function g can thus be parametrized by the basis coefficient vector $b = (b_1, b_2, \ldots)$. Specifying a prior probability model on b implicitly induces a prior probability model on the random function. Alternatively, stochastic process priors have been proposed as nonparametric prior distributions to model varying coefficient functions, including Gaussian process and random walk. In the following subsections, we are going to focus on reviewing Bayesian spline models and random effects models with Dirichlet Process (DP).

1.2.1 Bayesian Spline Models

Bayesian spline models have been studied in the literature for Bayesian regression splines and Bayesian penalized splines. For the literature of Bayesian regression splines, please refer to Denison et al. (1998b) and Hastie and Tibshirani (2000) among many others. For the work related to Bayesian penalized splines, see Biller and Fahrmeir (1997), Fahrmeir and Lang (2001), Brezger and Lang (2006), and Fahrmeir et al. (2004) for a brief view.

To estimate $\beta_k(t)$ in the model (1.1), a common cubic regression splines is to express $\beta_k(t)$ as

$$\beta_k(t) = \sum b_h f_h(t),$$

where $f(t) = (1, x, x^2, x^3, (t-\xi_1)_+^3, \ldots, (t-\xi_T)_+^3)$ is the set of basis functions with $(x)_+ = \max(x, 0)$ and $\xi = (\xi_1, \ldots, \xi_T)$ is a set of knots . The Bayesian specification is completed by placing a prior $p(\xi, b, \sigma)$ on the sets of knots and basis function coefficients. Relevant work can be found at Smith and Kohn (1996), Denison et al. (1998b) and Dimatteo et al. (2001). The strategy is to factor the prior $p(\xi, b, \sigma) = p(\xi)p(\sigma)p(b|\sigma)$. Zellner g-prior (Zellner, 1986) is used by Smith and Kohn (1996) for $p(b)$. Assuming a conjugate normal prior $b \sim N(0, c\sigma(B^TB)^{-1})$ with B is the design matrix in the sampled data set, the condition posterior mean $E(b|\xi, \sigma)$ is the linear shrinkage of the least square estimator $\hat{b}$. Dimatteo et al. (2001) chose a scalar c such that the prior variance in the Zellner g-prior equals to the information from one observation. Denison et al. (1998b) uses least square estimates of b conditioning on y and other parameters, which is equivalent to assigning a ridge prior $b \sim N(0, V)$ with $V = \text{diag}(\infty, v, \ldots, v)$.

Posterior computation typically proceeds via reversible jump MCMC (Green, 1995). Denison et al. (1998a) use "birth," "death" and "move" decisions to add, delete and

change knots in each iteration. It is important in updating ξ posteriori to marginalize ξ with respect to the coefficients b. With the conjugate normal prior for b, the marginal posterior $p(\xi|\sigma, y)$ can be evaluate analytically. Smith and Kohn (1998) proposed an interesting alternative, called focused sampling, for posterior computation.

1.2.2 Random Effects Models with Dirichlet Process

To estimate the random effects $v_{im}(t)$ in model (1.1), a popular option is to use basis expansion to express $v_{im}(t)$ as a linear combination of a set of pre-specified basis functions:

$$v_{im}(t) = \sum_{h=1}^{p} \theta_{ih} f_h(t),$$

where $\theta_i = (\theta_{i1}, \ldots, \theta_{ip})^T$ are basis coefficients specific to the ith subject, and $f(t) = \{f_1(t), \ldots, f_p(t)\}$ is a set of basis functions. Similar to the previous subsection, we specify a set of cubic spline basis functions for $f = (1, x, x^2, x^3, (t-\xi_1)^3_+, \ldots, (t-\xi_T)^3_+)$, in which $\xi = (\xi_1, \ldots, \xi_T)$ is a set of knots and $(x)_+$ returns 0 for negative x and x for positive x.

A Bayesian nonparametric specification is completed by placing a DP prior for the distribution of the random basis coefficients vector θ_i. Assume $\theta_i \sim Q$ with Q unknown, DP(αQ_0) models Q as infinite mixtures of point masses through a stick-breaking representation:

$$\begin{aligned} Q &= \sum_{h=1}^{\infty} w_h \delta_{\theta_h}(\cdot), \quad \theta_h \overset{i.i.d}{\sim} Q_0, \\ w_h &= U_h \prod_{j<h} (1 - U_j), \quad U_h \overset{i.i.d}{\sim} \text{Beta}(1, \alpha), \end{aligned}$$

where w_h is a probability weight formulated from a stick-breaking procedure, $\delta_\theta(\cdot)$

denotes a point mass at θ which is sampled from Q_0, Q_0 is a p-dimensional central probability measure as a prior guess, and α is a precision parameter expressing confidence in the prior guess. Under this formulation, the probability of θ_i allocated to $\delta_\theta(\cdot)$ is equivalent to w_h.

An important property of DP prior for Q is the discrete nature of Q. This specification creates ties among θ_is, $i = 1, \ldots, n$, such that each subject is allocated into one of $k \leq n$ clusters. Subjects allocated to the same cluster have the same random effects value. Denote $S_i = j$ for the ith subject assigned to cluster j. All subjects having $S_i = j$ have $\theta_i = \theta_h^*$, where θ_h^* denotes the value taken by the random basis coefficients vector of all subjects in cluster j. Hence, all subject in the same cluster have identical random effects $v_{im}(t)$ which equals to $f(t)\theta_h^*$.

The standard approach for posterior computation is based on Markov chain Monte Carlo (MCMC) algorithms. Three major types of MCMC algorithms have been proposed, which are collapsed Gibbs sampler (Maceachern, 1994), the blocked Gibbs sampler (Ishwaran and James, 2001), and reversible jump-type approaches (Jain and Neal, 2004; Dahl, 2007). The collapsed Gibbs sampler avoids updating the infinitely many parameters through marginalizing out Q, which is based on the Pólya urn scheme of Blackwell and MacQueen (1973). The blocked Gibbs sampler uses truncation of the stick-breaking procedure to approximate Q. To avoid truncation, Walker (2007) proposed a slice sampling approach and Papaspiliopoulos and Roberts (2008) proposed a retrospective MCMC algorithm. Further, Papaspiliopoulos (2008) proposed an efficient exact block Gibbs sampler which combines the advantages of the retrospective MCMC method and the slice sampling method.

1.3 Applications

The evolution of the varying coefficient models is deeply rooted in applications and reflects the increasingly large and complex problems that are arising in science and industry. They have been successfully applied to many scientific areas, including economics, finance, politics, epidemiology, medical science, ecology and so on. This proposal focuses on the applications of the varying coefficient models in biomedicine and epidemiology areas. They arise from different statistical contexts, such as the analysis of longitudinal data, survival data, spatial data, functional data, genetics data and imaging data. We are going to briefly review the time-varying coefficient models in the analysis of longitudinal data and survival data in the following subsections.

1.3.1 Analysis of Longitudinal Data with Time-Varying Coefficient Models

In the analysis of longitudinal data from biomedicine and epidemiology studies, it is of highly interest to study the patterns of time-varying variables, such as disease progression or trends of health status. Longitudinal data usually consist of repeatedly measured outcomes and covariates from a random sample of subjects measured over time. One statistical focus in analyzing longitudinal data is to assess the time-varying impacts of covariates on outcome trajectories, where the covariates might or might not be time dependent. Thus nonparametric varying-coefficient models become particularly useful in longitudinal analyses.

A longitudinal sample is typically denoted by $(y_{ij}, x_{ij}, z_{ij}, t_{ij})$, for $i = 1, \ldots, n$, and $j = 1, ..., n_i$, where $x_i(t_{ij}) = (x_{ij0}, \ldots, x_{ijK})^T$, $z_i(t_{ij}) = (z_{ij0}, \ldots, z_{ijM})^T$ and $(y_{ij}, x_{ij}, z_{ij}, t_{ij})$ denote the jth outcome, two set of covariates and time design points, respectively, of the ith subject. To examine whether the associations of covariates with

outcomes change over time, we can fit a fixed-effect varying coefficient model (Brumback and Rice, 1998; Hoover et al., 1998), or a mix-effect varying coefficient model (Liang et al., 2003) as the above model (1.1), or a random varying-coeffcient model (Wu and Liang, 2004). The time-varying coefficients can be estimated by the kernel, polynomial and smoothing spline methods.

The issue of incorporating the within subject correlation structure into the estimation procedure is important in longitudinal data analysis. Lin and Carroll (2001) addressed this situation in the nonparametric fixed effects time-varying coefficient models. The studies of the estimation of the within subject correlation structure were systematically considered by Fan et al. (2007) and Sun et al. (2007).

Missing data has been an important topic in long term longitudinal studies. Hogan et al. (2004) studied the mixtures of varying coefficient models for longitudinal data having discrete or continuous nonignorable dropout.

1.3.2 Analysis of Survival Data with Time-Varying Coefficient Models

Survival analysis concerns the failure time to some event, such as failure of a machine component, death of a patient, or recurrence of prostate cancer of a male patient. The Cox proportional hazard (PH) model (Cox, 1972) is the most popular model in the analysis of survival data. However, it doesn't address any potential possible dynamic feature in the data set. Varying coefficient proportional hazard function models first introduced by Hastie and Tibshirani (1993) can be used to address any dynamic structure.

Suppose that we have a survival sample $(U_i, X_i^T, y_i, \delta_i)$, $i = 1, \ldots, n$, where $y_i = \min(T_i, C_i)$, $\delta_i = I(T_i > C_i)$, T_i and C_i are respectively the survival time and censoring

time of the ith subject, and U_i is a scalar covariate. We assume a noninformative censoring mechanism. The distinct failure times are denoted by $t_{(1)} < \ldots < t_{(L)}$, with d_l individuals failing at $t_{(l)}$ for $l = 1, \ldots, L$. We fit a varying coefficient proportional hazard function model (Zhang and Steele, 2004) as follows:

$$h(t|X, U) = h_0(t) \exp\{X^T \boldsymbol{a}(U)\}.$$

For any give u, the local partial likelihood estimator (Fan et al., 2006) of $\boldsymbol{a}(u)$ is obtained by maximizing the following local partial log-likelihood function

$$\sum_{l=1}^{L} \left(\sum_{j \in D_l} K_h(U_j - u) \left[X_j^T \{\boldsymbol{a} + \boldsymbol{b}(U_j - u)\} - \log \left(\sum_{k \in R_l} \exp[X_k^T \{\boldsymbol{a} + \boldsymbol{b}(U_j - u)\}] K_h(U_k - u) \right) \right] \right),$$

where R_l is the set of indices for the individuals at risk up to time $t_{(l)}$, and D_l is the set of indices for the events at $t_{(l)}$.

For independent data, kernel estimation of the semiparametric varying-coefficient proportional hazard function models have been studied by Cai and Sun (2003), Tian et al. (2005), and Fan et al. (2006), and spline estimation was studied by Ahmad et al. (2005). Other related works include Zucker and Karr (1990), and Winnett and Sasieni (2003). There are several nonparametric regression models for clustered failure time data. Under the assumption of independence, Cai et al. (2007) investigated kernel smoothing properties for varying coefficient models for multivariate survival data with stratified baseline hazards. Yu and Lin (2008) studied weighted local polynomial kernel estimating equations for clustered failure time data. They concluded the most efficient local polynomial kernel estimator can be attained by silencing the within-cluster correlation. Yu and Lin (2010) studied the performance of profile-kernel estimators of the nonparametric time-varying coefficients in a semiparametric time-varying coefficient model for clustered survival data under working independence, and showed that they

are semiparametric efficient for independent data.

Chapter 2

Semiparametric Bayes local additive models for longitudinal data

2.1 Introduction

Longitudinal data arise frequently in biomedical fields when subjects are repeatedly measured over time. To study the time pattern of covariate effects on an outcome variable, time-varying coefficient models are commonly used since their formal introduction to the statistical literature by Hastie and Tibshirani (1993). Nonparametric regression methods to estimate time-varying coefficient functions have been developed by Hoover et al. (1998), and Lin and Ying (2001) among many others. Please refer to Wu and Zhang (2006) and Fan and Zhang (2008) for comprehensive reviews of statistical procedures for varying coefficient models.

We focus on a random time-varying coefficient model:

$$Y_{ij} = \beta_{i0}(t_{ij}) + X_i'\beta_i(t_{ij}) + \epsilon_i(t_{ij}), \quad i = 1, \ldots, n,\ j = 1, \ldots, m_i,\ t_{ij} \in [0, T], \quad (2.1)$$

where Y_{ij} is the observed response variable, X_i is a $K \times 1$ covariate vector for the i-th subject, $\beta_{i0}(t_{ij})$ is a subject-specific baseline and $\beta_i(t_{ij}) = (\beta_{i1}(t_{ij}), \ldots, \beta_{iK}(t_{ij}))'$ is a

$K \times 1$ covariate effect vector for the i-th subject at time t_{ij}. To estimate the unknown functions $\beta_{i0}(t)$ and $\beta_i(t)$ in model (2.1), a common approach is to express them as linear combinations of some pre-specified basis functions. Heterogeneity in the curves is then controlled by the variation of their basis coefficients which are usually treated as random effects and follow a specific random effects distribution. A concern rises about the sensitivity of associated inferences to the choice of the random effects distribution on the basis coefficients.

There is a rich Bayesian nonparametric literature associated with modelling random effects. It has now become routine to use Dirichlet process (DP) priors (Ferguson, 1973, 1974) and DP mixtures (DPM) (LO, 1984; Escobar and West, 1995) for random effects distributions in Bayesian hierarchical models. For important early references, refer to Bush and MacEachern (1996), Müller and Rosner (1997) and Kleinman and Ibrahim (1998). Cruz-Mesía et al. (2007) adopted an analysis-of-variance dependent DP (De Iorio et al., 2004) for the problem of sequential classification analysis. Ohlssen et al. (2007) provide a review and tutorial on the practical use of such approaches, and the DPpackage in R is now available for routine use (http://cran.r-project.org/web/packages/DPpackage/index.html).

It is important to capture the local structure of similarity and deviation among subjects in detecting the signals for model (2.1). The DP prior assumes the random effects distribution $P \sim \mathrm{DP}(\alpha P_0)$, where P_0 is the base distribution and α is a concentration parameter. Subjects are partitioned into clusters with the number of sample clusters being proportional to $\alpha \log n$. Small α favors few clusters to be occupied leading to substantial borrowing of information across subjects within a cluster in estimating the basis coefficients. A drawback of the DP prior is the assumption of global clustering. Two subjects are either allocated to a common cluster or two different clusters. It is common in reality that two subjects have similar trajectories in certain time periods

while having local deviations. In such situations, DP priors either inappropriately allocate two subjects to a common cluster, obscuring the local differences or assigns them to two separate clusters. Dunson (2009) proposed a local partition process (LPP) prior on random effects, which allows both global and local clustering. The LPP prior leads to borrowing of information in estimating the basis coefficients and can accommodate basis selection to address the curse of dimensionality.

Characterizing local features has been an important focus in functional data analysis. Representative Bayesian semiparametric approaches include Bayesian wavelet-based functional mixed modeling (Morris and Carroll, 2006), random effects models relying on adaptive basis function representations (Thompson and Rosen, 2008; Botts and Daniels, 2008) and hierarchical Gaussian processes (Behseta et al., 2005). A hybrid Dirichlet mixture model (Petrone et al., 2009) was proposed to distinguish functional local features by characterizing individual functions as a patchwork of segments which are locally drawn from a collection of global GP realizations. We propose to capture the local features through assigning an LPP prior on the random effects distribution of the basis function coefficients, where a pre-specified set of basis functions are assumed to linearly span the unknown time-varying coefficient functions $\beta_{i0}(t)$ and $\beta_{ik}(t)$, $k = 1, \ldots, K$.

The motivating application for this chapter is a yeast cell-cycle gene expression data set. The expression profiles of 297 genes identified as cell cycle-regulated genes in the genome of *Saccharomyces cerevisiae* were measured over 2 cell cycles. Since transcription factors (TFs) are key elements controlling the movement of genetic information from DNA to mRNA at cell-cycle level transcription, it is important to capture the dynamic behavior of gene expression regulated by TFs. In addition, cell cycle-regulated genes are involved in different processes such as DNA synthesis, budding, and cytokinesis (Spellman et al., 1998) in each cell cycle, implying different expression behaviors

of different genes. Typical analyses assume a common time-varying behavior for all the genes (Wang et al., 2007, 2008). We propose to use our random time-varying coefficient model with LPP prior to determine the gene-specific time-varying TF regulating course.

Considering TFs regulating cell cycle-regulated genes which are active at different biological processes, it is important to develop a local inference approach to determine TFs-regulated time windows. Current Bayesian varying-coefficient models allow regression coefficient to vary with time and/or spatial location (refer to Lee and Shaddick (2007) for a recent reference). Such approaches can be used to obtain a posterior mean estimate of the varying coefficient curve, as well as pointwise 95% credible intervals. However, formal Bayesian hypothesis testing of local significance of a predictor within a given time window is not considered. Although one can potentially identify time regions across which the pointwise credible bands do not include zero as potentially significant, this may lead to an inflated type I error rate.

The goals of this chapter cover three aspects. The first applies the LPP prior to the random effects distributions in the random time-varying coefficient model (2.1) to facilitate global and local borrowing of information among subjects. The second constructs a Bayesian confidence band for the mean time-varying effect of the predictor. The third develops a Bayesian local hypothesis testing approach to examine if and where the predictor has significant impact on an outcome trajectory. Multiplicity issues arising from pointwise inference is accounted for by controlling false discovery rate (FDR).

In this chapter, section 2.2 introduces the model and associated inferences motivated by the above considerations. Section 2.3 introduces an MCMC algorithm for posterior computation. Subsection 2.4 is for an illustrative simulation. Section 2.5 applies the model to a yeast cell-cycle gene data set data set. Section 2.6 makes some further discussion.

2.2 Semiparametric Bayes Local Additive Model

2.2.1 LPP2 prior

Dunson (2009) proposed a class of local partition process priors (LPP) for unknown random effects distributions to facilitate both global and local clustering of random effects, providing a generalization of the widely-used Dirichlet process prior that avoids the global clustering assumption. Our focus is on the simpler LPP2 prior, which is reviewed below.

Sample two collections of independent and identically distributed p-dimensional random elements $\Theta_{gh} = (\Theta_{g,h,1}, \ldots, \Theta_{g,h,p})'$, $g = 0, 1$, $h = 1, \ldots, \infty$, from a base distribution P_0. Denote $\Xi_0 = \{\Theta_{0h}\}_{h=1}^{\infty}$ for the global family of coefficient vectors and $\Xi_1 = \{\Theta_{1h}\}_{h=1}^{\infty}$ for the local family. Introduce a p-dimensional indices vector $z = (z_1, \ldots, z_p)'$ for the allocation to global clustering or local clustering, with $z_j \sim \text{Ber}(\nu_j)$ (Ber: Bernoulli) independently, taking the value of 1 for global clustering and 0 for local clustering, $j = 1, \ldots, p$. Define a $p \times 2$ local cluster indices matrix $\psi = (\psi_1', \ldots, \psi_p')'$, with $\psi_j = (1 - z_j, \phi_j)$ and $\phi_j \in \{1, \ldots, \infty\}$, $j = 1, \ldots, p$. A corresponding hybrid atom $\Theta_\psi = (\Theta_{\psi_1,1}, \ldots, \Theta_{\psi_p,p})'$ is obtained by setting the jth element of Θ_ψ equal to $\Theta_{\psi_j,j}$.

Considering n vectors of p-dimensional random effects for n subjects, $\theta_i = (\theta_{i1}, \ldots, \theta_{ip})' \sim P$ with P unknown, $i = 1, \ldots, n$, LPP2 prior models P as a hybrid mixture distribution:

$$P = \sum_{z_1=0}^{1} \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_p=0}^{1} \sum_{\psi_p=(1-z_p,1)}^{(1-z_p,\infty)} \pi_{\psi_1,\ldots,\psi_p} \delta_{\Theta_\psi}, \tag{2.2}$$

where δ_x denotes a degenerate distribution with all its mass at x, $\pi_{\psi_1,\ldots,\psi_p}$ is the probability of $\theta_i = \Theta_\psi$ that $\pi_{\psi_1,\ldots,\psi_p} \geq 0$ and $\sum_{z_1=0}^{1} \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_p=0}^{1} \sum_{\psi_p=(1-z_p,1)}^{(1-z_p,\infty)} \pi_{\psi_1,\ldots,\psi_p} = 1$.

Let π_h denote $\Pr(\theta_{ij} = \Theta_{g,h,j})$, $g = 0, 1$, $h = 1, \ldots, \infty$, $j = 1, \ldots, p$. Those random effects having $z_j = 1$ are specified to be allocated together to a component in the

global family Ξ_0, while others having $z_j = 0$ are specified to be allocated to their own component in the local family Ξ_1. Let $J_0 = \{j : z_j = 1\}$ and $J_1 = \{j : z_j = 0\}$. Conditional on the values of z,

$$\Pr(\psi_j = (0, h), j \in J_0 | z_1, \ldots, z_p) = \pi_h, \ h = 1, \ldots, \infty,$$
$$\Pr(\psi_j = (1, h_j), j \in J_1 | z_1, \ldots, z_p) = \prod_{j \in J_1} \pi_{h_j}, \ h_j = 1, \ldots, \infty.$$

The allocation probability of $\theta_i = \Theta_\psi$ in (3.8) is then simply

$$\pi_{\psi_1, \ldots, \psi_p} = \Pr(\psi_1 = (1 - z_1, h_1), \ldots, \psi_p = (1 - z_p, h_p)) = \pi_h \left\{ \prod_{j \in J_1} \pi_{h_j} \right\} \left\{ \prod_{j=1}^{p} \nu_j^{z_j} (1 - \nu_j)^{1 - z_j} \right\}.$$

The specification is completed by choosing the hyperpriors

$$\nu_j \sim \text{Beta}(1, \gamma), \quad j = 1, \ldots, p,$$
$$\pi_h = \pi_h^* \prod_{l<h} (1 - \pi_l^*), \quad \pi_h^* \sim \text{Beta}(1, \alpha), \quad h = 1, \ldots, \infty, \quad j = 0, 1, \ldots, p,$$

where γ controls the overall weight on the local family, and α controls the overall number of clusters. As shorthand the LPP prior is denoted $P \sim \text{LPP2}(\alpha, \gamma, P_0)$.

The LPP2 prior specification (3.8) can also be viewed as a hybrid mixture model of infinitely many components drawn from P_0 via 2-stage clustering. Stage 1 determines the membership of global or local clustering for each of the p components. Those components allocated to the global clustering membership will be clustered together to an atom in the global family at stage 2, while those allocated to the local family will be allocated individually to their own clusters. The joint cluster membership probability at stage 1 corresponds to $\Pr(z_1, \ldots, z_p) = \prod_{j=1}^{p} \nu_j{}^{z_j} (1 - \nu_j)^{1 - z_j}$. The joint cluster allocation weight at stage 2 conditional on stage 1 corresponds to $\Pr(\psi_1 = (1 - z_1, h_1), \ldots, \psi_p = (1 - z_p, h_p) | z_1, \ldots, z_p)$. The overall joint cluster allocation weight

corresponds to $\pi_{\psi_1,\ldots,\psi_p}$.

2.2.2 The Model

We propose a semiparametric Bayesian local additive model for the analysis of repeated measurements. Considering a functional response $Y_i(t)$ for the ith subject $i = 1, \ldots, n$, we rewrite model (2.1) as

$$Y_i(t) = \beta_{i0}(t) + \sum_{k=1}^{K} \beta_{ik}(t) x_{ik} + \epsilon_i(t), \quad \epsilon_i(t) \sim N(0, \tau^{-1}), \tag{2.3}$$

where $\beta_{i0}(t)$ is the ith subject's baseline curve, $\beta_{ik}(t)$ is the ith subject's time-varying coefficient for the kth predictor x_{ik}, x_{ik} is normalized, and the measurement error process $\epsilon_i(t)$ is a realization of an uncorrelated zero-mean Gaussian process with variance τ^{-1}.

Similar to Huang et al. (2002), we consider basis expansion to estimate the time-varying coefficient regression functions. Assume that $\{\beta_{ik}(t)\}_{k=0}^{K}$ can be expressed as a linear combination of basis functions as follows:

$$\beta_{ik}(t) = \sum_{l=1}^{p_k} \theta_{ikl} b_{kl}(t) = b_{ik}(t)' \theta_{ik}, \tag{2.4}$$

where $\theta_{ik} = (\theta_{ik1}, \ldots, \theta_{ikp_k})'$ and the basis functions $b_k(t) = (b_{k1}(t), \ldots, b_{kp_k}(t))'$. Various basis systems can be applied, including Fourier bases, polynomial bases, Wavelet bases, and B-spline bases such as cubic B-spline bases. For computational convenience, we let the basis function $b_k(t) \equiv b(t) = (b_1(t), \cdots, b_p(t))'$ and $p_k \equiv p$ for all k. We specify Gaussian kernel as the basis functions $b_1(t) = 1$, $b_{j+1}(t) = \exp(-\zeta ||t - \eta_j||^2)$, $j = 1, \ldots, p-1$, where $\eta_1, \ldots, \eta_{p-1}$ are equally spaced kernel locations and $\zeta = 25$. Let

$\theta_i = (\theta'_{i1}, \cdots, \theta'_{iK})'$ and $B_i(t) = (b(t)'x_{i1}, \ \cdots, \ b(t)'x_{iK})'$, model (2.3) can be written as

$$Y_i(t) = b(t)'\theta_{i0} + B_i(t)'\theta_i + \epsilon_i(t). \tag{2.5}$$

A Bayesian specification of our model is completed with LPP2 priors for the distributions of the random effects, $\{\theta_{ik}\}_{k=0}^{K}$. Assuming $\theta_{i0} \sim P$ and $\theta_i \sim \tilde{P}$ with P and $\tilde{P}$ unknown and independent, we let:

$$\begin{aligned} P &\sim \text{LPP2}(\alpha_0, \gamma_0, P_0) \\ \tilde{P} &\sim \text{LPP2}(\alpha_1, \gamma_1, \otimes\{P_{0k}\}_{k=1}^{K}), \ P_{0k} = P_0, \end{aligned} \tag{2.6}$$

where the prior guess is around a p-dimensional parametric base distribution P_0, and (α_0, α_1) and (γ_0, γ_1) are hyperparameters, with (α_0, α_1) characterizing concentration around the prior guess and (γ_0, γ_1) determining the overall allocation weight on the local family. These hyperparameters impact the induced global and local clustering structure, and by choosing hyperpriors we allow a high degree of data adaptivity.

Considering the potentially varied shapes of the time-varying functions, we devise to automatically select among a large number of pre-specified potential basis functions and accommodate uncertainty in basis function selection by choosing a shrinkage prior for the basis coefficients. To achieve the shrinkage property, we specify the base distribution P_0 as:

$$P_0: \quad \theta_{ik} \sim N_p(0_{p\times 1}, \Lambda^{-1}), \quad i = 1, \ldots, n, \quad k = 0, 1, \ldots, K. \tag{2.7}$$

$\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_p)$ with $\lambda_l \sim G(0.5, 0.5)$. To allow borrowing information across coefficients over the degree of shrinkage, we prefer $\lambda_l = \lambda$, a common λ across coefficients. This specification leads to light-tails feature of P_0 a priori and approximately Gaussian feature a posteriori. The coefficients for unnecessary basis functions are shrunk to zero

a posteriori, while the coefficients for significant basis functions fall in the tails. If too many unnecessary basis functions are included, Λ will be over-estimated and the coefficients for important basis functions will be heavily shrunk.

2.2.3 Induced Properties

The LPP2(α, γ, P_0) prior on the random effects distribution constructs a nonparametric prior for the time-varying coefficient functions in the model specified through (2.3) – (2.7). Considering a generic unknown function

$$f(t) = \sum_{j=1}^{p} \vartheta_j b_j(t) = B(t)'\vartheta,$$

with $t \in [0, \mathrm{T}]$, $\vartheta = (\vartheta_1, \ldots, \vartheta_p)'$ and assuming $\vartheta \sim P$, $P \sim$ LPP2(α, γ, P_0), with P_0 defined as in (2.7), the induced prior presents the following properties.

Proposition 2.2.1. *Conditional on* $\Lambda^{-1} = diag\{\lambda_1^{-1}, \ldots, \lambda_p^{-1}\}$,

$$(f(t_1), f(t_2), \ldots, f(t_L))'|\Lambda^{-1} \sim N_L(0_{L\times 1}, \Sigma_{L\times L})$$

with $\Sigma_{L\times L} = [\sigma_{ij}]$, $i = 1, \ldots, L, j = 1, \ldots, L$, *and* $\sigma_{ij} = B(t_i)'\Lambda^{-1}B(t_j)$.

Further,

$$\mathrm{Corr}(f(t_1), f(t_2)|\Lambda^{-1}) = \frac{B(t_1)'\Lambda^{-1}B(t_2)}{(B(t_1)'\Lambda^{-1}B(t_1))^{1/2}(B(t_2)'\Lambda^{-1}B(t_2))^{1/2}}.$$

According to the Cauchy-Schwarz inequality, the correlation is bounded above by 1 and converges to 1 as $t_1 \to t_2$. An appealing property of the induced prior is, when $\lambda_1 = \ldots = \lambda_p = \lambda$,

$$\mathrm{Corr}(f(t_1), f(t_2)|\lambda) = \frac{B(t_1)'B(t_2)}{(B(t_1)'B(t_1))^{1/2}(B(t_2)'B(t_2))^{1/2}},$$

which depends only on the set of basis functions and not the base measure P_0. When a local basis is chosen, such as B-splines or Gaussian kernels, a local dependence structure will result with the correlation between $f(t_1)$ and $f(t_2)$ decaying to close to zero as t_1 and t_2 move further apart. We can extend this correlation structure for long-range dependencies by accommodating non-diagonal Λ or choosing Fourier bases.

From the above, it follows directly that the LPP2(α, γ, P_0) prior on the distribution of the basis coefficients leads to baseline and time-varying coefficient functions that are assigned Gaussian process priors *marginally*. In particular, the curves for individual subjects are drawn from a Gaussian process with mean function $m(t) = 0$ and covariance function $c(t_1, t_2) = B(t_1)'\Lambda^{-1}B(t_2)$. By specifying hyperpriors on the parameters in Λ, the data inform about the parameters in the covariance function. Although the marginal prior for the individual curves is a Gaussian process, the marginal distribution across subjects is nonparametric and highly flexible, so the variation among subjects is not restricted to be Gaussian.

2.2.4 Hypothesis Formulation & Testing

We are interested in testing the null hypotheses that a predictor has no effect at a particular time or interval of times, and that there are time intervals across which the predictor has no effect. We provide a local interval null hypothesis testing approach to make pointwise inference for the predictor effects over time. The local null and alternative hypotheses for the kth predictor effect specific to a time point $t \in [0, T]$ are formulated as:

$$H_{0k}(t) : \left|\overline{\beta}_k(t)\right| \leq \epsilon, \text{ versus } H_{1k}(t) : \left|\overline{\beta}_k(t)\right| > \epsilon,$$

where $\overline{\beta}_k(t)$ represents the mean of the subject-specific time-varying coefficients for the kth predictor at time t. To test whether the average time-varying coefficient curve for the kth predictor deviates from a small neighborhood of zero anywhere in time interval $[0, T]$, the global null and alternative hypotheses are formulated as

$$H_{0k} : \left|\overline{\beta}_k(t)\right| \leq \epsilon, \text{ for all } t \in [0, T]$$
$$H_{1k} : \left|\overline{\beta}_k(t)\right| > \epsilon, \text{ for any } t \in [0, T].$$

The global null hypothesis is the intersection of all the local null hypotheses. Any rejection of a local null hypothesis leads to the rejection of the global null hypothesis. We choose a suitable small value 0.05 for ϵ as our default, noting again that the data have been normalized.

To conduct local hypotheses testing, we use the Bayesian decision rule for multiple testing proposed by Müller et al (2004). Following the implementation in Wang & Dunson (2010), our strategy is to reject $H_{0k}(t)$ if the posterior alternative hypothesis probability $\upsilon_{kt} = \Pr(H_{1k}(t)|Data) \geq r$ for any $t \in [0, T]$, with r as a common threshold for all the local hypotheses. r is chosen to minimize the posterior expected false negative rate (FNR) under the constraint of the posterior expected false discovery rate (FDR) being no greater than α_T, where α_T is pre-specified (we focus on a value of 0.05). Denote $d_{kt} = 1(\upsilon_{kt} \geq r)$, an indicator of rejecting $H_{0k}(t)$. The posterior expected FNR and FDR are calculated as:

$$\overline{FNR} = \frac{\sum_{k=1}^{K} \int_0^T (1 - d_{kt})\upsilon_{kt} \mathrm{d}t}{K \int_0^T \mathrm{d}t - \sum_{k=1}^{K} \int_0^T d_{kt} \mathrm{d}t + \kappa_0} \approx \frac{\sum_{k=1}^{K} \sum_{t=t_1}^{t_m} (1 - d_{kt})\upsilon_{kt} T/m}{KT - \sum_{k=1}^{K} \sum_{t=t_1}^{t_m} d_{kt} T/m + \kappa_0},$$

$$\overline{FDR} = \frac{\sum_{k=1}^{K} \int_0^T d_{kt}(1 - \upsilon_{kt}) \mathrm{d}t}{\sum_{k=1}^{K} \int_0^T d_{kt} \mathrm{d}t + \kappa_0} \approx \frac{\sum_{k=1}^{K} \sum_{t=t_1}^{t_m} d_{kt}(1 - \upsilon_{kt}) T/m}{\sum_{k=1}^{K} \sum_{t=t_1}^{t_m} d_{kt} T/m + \kappa_0},$$

where $t_1, \ldots, t_m$ is a fine grid of times equally spaced along $[0, T]$ and κ_0 is a small

positive constant to avoid a zero denominator. In summary, our decision rule is to determine the optimal threshold r^* by $r^* = \text{argmin}\, \overline{FNR}\{r \in [0,1], \overline{FDR} \leq \alpha_T\}$. For each combination of k and $t \in \{t_1, \ldots, t_m\}$, we reject $H_{0k}(t)$ if $\upsilon_{kt} \geq r^*$. For a sufficiently fine grid of $m \geq 10$, the results are robust to m, with the optimal threshold appropriately adapting to the chosen m.

2.2.5 Bayesian Confidence Band

We focus on constructing a Bayesian simultaneous confidence band for the mean coefficient curve $\overline{\beta}_k(t)$, $k = 1, \ldots, K$, from its posterior MCMC samples. Assuming there is a collection of posterior sampled curves $\beta_k^s = (\beta_k^s(t_1), \ldots, \beta_k^s(t_L))'$, $s = 1, \ldots, S$ indexing posterior iterations after burn-in, our goal is to compute a simultaneous confidence band for $\overline{\beta}_k(t)$. The principle in constructing a Bayesian confidence band is to search for a region $R_\alpha = \{R_\alpha(t), t \in [0,T]\}$ such that $\Pr\{\overline{\beta}_k(t) \in R_\alpha(t), t \in [0,T]\} = 1 - \alpha$ a posteriori.

The most commonly-used method is to compute a confidence band based on pointwise confidence intervals, which we denote as CR. The constructed confidence band is obtained as $[\beta_k^L(t_l), \beta_k^U(t_l)]$, $l = 1, \ldots, L$, where $\beta_k^L(t_l)$ is the $\alpha/2$ empirical percentile of the posterior samples and $\beta_k^U(t_l)$ is the $1 - \alpha/2$ percentile. However, such pointwise confidence bands are not interpretable as joint confidence bands. A method proposed by Crainiceanu et al.(2007) assumes approximate posterior normality and derives the $1 - \alpha$ sample percentile $c_{1-\alpha}$ of

$$\max_{l=1,\ldots,L} \left| \frac{\beta_k^s(t_l) - \hat{\beta}_k(t_l)}{\sqrt{\widehat{\text{var}}(\hat{\beta}_k(t_l))}} \right|, \quad s = 1, \ldots, S,$$

where $\hat{\beta}_k(t_l)$ is the posterior mean at time t_l and $\sqrt{\widehat{\text{var}}(\hat{\beta}_k(t_l))}$ is its posterior standard

deviation. A simultaneous confidence region is given by the hyperrectangular

$$[\hat{\beta}_k(t_l) - c_{1-\alpha}\sqrt{\widehat{\text{var}}(\hat{\beta}_k(t_l))}, \hat{\beta}_k(t_l) + c_{1-\alpha}\sqrt{\widehat{\text{var}}(\hat{\beta}_k(t_l))}], \;\; l = 1, \ldots, L.$$

In our implementation, we replace $c_{1-\alpha}$ by c_b which is calculated by $c_b = \max(|c_{\alpha/2}|, |c_{1-\alpha/2}|)$ to account for skewness. We denote this method as Pnorm.

Alternatively, we propose two new approaches that avoid the posterior normality assumption. The strategy remains based on pointwise measures of uncertainty. The first method Mdev computes the posterior sample average curve $\hat{\beta}_k(t_l)$ and the pointwise $\alpha/2$ percentile $s_{\alpha/2}(t_l)$ and $1-\alpha/2$ percentile $s_{1-\alpha/2}(t_l)$, $l = 1, \ldots, L$. By deriving the maximal deviations $s_{\alpha/2} = \max_{l=1,\ldots,L} \;\; (\hat{\beta}_k(t_l) - s_{\alpha/2}(t_l))$ and $s_{1-\alpha/2} = \max_{l=1,\ldots,L} \;\; (s_{1-\alpha/2}(t_l) - \hat{\beta}_k(t_l))$ for the lower and upper $\alpha/2$ percentiles away from the posterior mean estimator $\hat{\beta}_k(t_l)$, we obtain the confidence band

$$[\hat{\beta}_k(t) - s_{\alpha/2}, \hat{\beta}_k(t) + s_{1-\alpha/2}].$$

This interval is potentially conservative. The second method Mdiff searches the q_{b_α} sample percentile of

$$\max_{l=1,\ldots,L} \;\; \left|\beta_k^s(t_l) - \hat{\beta}_k(t_l)\right| \;\; s = 1, \ldots, S,$$

such that $1 - (b_\alpha)^L = \alpha$. The confidence band is calculated by

$$[\hat{\beta}_k(t) - q_{b_\alpha}, \hat{\beta}_k(t) + q_{b_\alpha}].$$

2.3 Posterior Computation

We develop a Markov chain Monte Carlo (MCMC) algorithm for posterior computation by adapting an efficient exact block Gibbs sampler (Papaspiliopoulos, 2008) for

Dirichlet mixture models. The exact block Gibbs sampler combines the advantages of the retrospective MCMC method of Papaspiliopoulos & Roberts (2008) and the slice sampling method of Walker (2007). Allowing data to inform information about the hyperparameters involved in the model through (2.3) – (2.7), we specify $\alpha_0 \sim \text{Ga}(a_\alpha, b_\alpha)$, $\gamma_0 \sim \text{Ga}(a_\gamma, b_\gamma)$, $\alpha_1 \sim \text{Ga}(a_\alpha, b_\alpha)$, $\gamma_1 \sim \text{Ga}(a_\gamma, b_\gamma)$, $\tau \sim \text{Ga}(a_\tau, b_\tau)$. Introducing auxiliary variables $\{u_{ikl}\}$ to avoid truncated approximations, $k = 0, \ldots, K$, $l = 0, \ldots, p$, $i = 1, \ldots, n$, the complete data joint likelihood of y, u, and z is:

$$\prod_{i=1}^{n}\{g(y_i; \Theta_{0\psi_{i0}}, \Theta_{\psi_{i1},\ldots,\psi_{iK}}, \tau)1(u_{i00} < \pi_{0\phi_{i00}}) \prod_{l=1}^{p} 1(u_{i0l} < \pi_{0\phi_{i0l}})\nu_{0l}^{z_{i0l}}(1-\nu_{0l})^{(1-z_{i0l})}$$
$$\prod_{k=1}^{K} 1(u_{ik0} < \pi_{\phi_{ik0}}) \prod_{l=1}^{p} 1(u_{ikl} < \pi_{\phi_{ikl}})\nu_{kl}^{z_{ikl}}(1-\nu_{kl})^{(1-z_{ikl})}\},$$

where y_i is a $m_i \times 1$ outcome vector for the i-th subject, $g(\cdot)$ is the density function of the outcome vector y_i, $i = 1, \ldots, n$, and the $\{u_{ikl}\}$s are constrained to the interval $(0, 1)$.

Please refer to the appendix B for the sampling algorithm in details.

This algorithm is straightforward to implement and exhibits good performance in convergence and mixing. To perform hypothesis testing, we estimate the posterior hypothesis probability from the MCMC posterior samples. For any $t \in [0, T]$, $\Pr(H_{1k}(t)|\text{Data})$ is estimated as the proportion of iterations after burn-in satisfying $1/n \sum_{i=1}^{n} \hat{\beta}_{ik}(t) \notin [-\epsilon, \epsilon]$.

2.4 Simulation Study

We conducted a simulation study to assess the finite sample performance of our approach. In each simulation, we generate a sample of 100 curves from the following

model:

$$Y_i(t) = \beta_{i0}(t) + \beta_{i1}(t)x_i + \epsilon_i(t),\ \epsilon_i(t) \sim N(0, \tau^{-1}), \quad i = 1, \ldots, 100,\ t \in [0, 1].$$

Each curve has 40 observations equally spaced along the time interval $[0, 1]$ and we set $\tau = 1$. The covariate x_i's are sampled from Unif$(0, 1)$ and then standardized to have mean 0 and standard deviation 1. The baseline $\beta_{i0}(t)$ for $t \in [0, 1]$, $i = 1, \ldots, 100$ is generated from a Gaussian process with mean curve $m_{i0}(t) = a_{i0} + b_{i0} \sin(\pi t/60)$ and covariance function $c_0(t, t') = \sigma_0^2 \rho^{|t-t'|}$, where $a_{i0} \sim N(5, 1)$, $b_{i0} \sim N(20, 1)$, $\sigma_0 = 0.224$ and $\rho = 0.5$. The covariate effect $\beta_{i1}(t)$ is considered for three cases: (1) global null case: $\beta_{i1}(t) \sim \text{GP}(0, c_1(t, t'))$, (2) local alternative case: $\beta_{i1}(t) \sim \text{GP}((a_{i1} + b_{i1} \cos(\pi(t - 20)/20))1(t > 0.4), c_1(t, t'))$, and (3) global alternative case: $\beta_{i1}(t) \sim \text{GP}(a_{i1} + b_{i1} \cos(\pi(t - 20)/20), c_1(t, t'))$. The random coefficients $a_{i1} \sim N(3, 1)$ and $b_{i1} \sim N(2, 1)$. The covariance function $c_1(t, t')$ is set equivalent to $c_0(t, t')$.

We analyze each simulated scenario using 100 repeated simulations by our method. We set Ga$(1, 1)$ hyperpriors for α_0, γ_0, α_1, and γ_1, and a Ga$(0.1, 0.1)$ hyperprior for τ. The exact block Gibbs sampler is run 25,000 iterations, with the first 2500 samples discarded as burn-in. Every 20th sample is collected to thin the chain. For each case, our Gibbs sampler converges rapidly and exhibits efficient mixing based on the examination of the trace plots of α_0, β_0, α_1, β_1, τ, the elements of estimated functions, and the performance of hypothesis testing as scaling up the simulation iterations. It is not reliable to assess convergence and mixing based on the elements of the posterior sampled atoms Θ_h or Θ_{jh} due to label switching issues.

We compare the LPP2 prior, the standard DP prior and a fixed effects model with P_0 as prior in terms of posterior estimation and posterior hypothesis testing. Mean square error (MSE) is used to assess the accuracy of posterior estimation. Table 2.1 shows the LPP2 prior having smaller average mean square error than the DP prior

and the fixed effects model under each of the three cases. In performing the global hypothesis testing, we propose to reject the global null if any local null hypothesis is rejected. Table 2.2 summarizes type I error rates in the global null case and powers in the global and local alternative cases. LPP2 prior is shown to have a type I error closer to the significance level 0.05 in the null case (0.05 vs 0.19) and larger power in the local alternative case (0.72 vs 0.63) than DP prior. To perform the time-wise local inference, the proposed approach rejects the local null hypotheses if $\Pr(H_{1k}(t)|Data) \geq r$, where r is optimized to minimize FNR under the constraint of FDR ≤ 0.05. The averages of the observed FDRs and powers from 100 repeated data sets are calculated to evaluate the local hypothesis test approach. As seen in table 2.3, LPP2 prior has FDRs closer to 0.05 in the global null case (0.042 vs 0.034) and the local alternative case (0.032 vs 0.028), and larger power in the local alternative case (0.367 vs 0.332) and the global alternative case (0.998 vs 0.973).

We employ 4 approaches: Mdiff, Mdev, Pnorm and CR to construct simultaneous 95% confidence bands. The comparison results of the 95% confidence band coverage rates are listed in Table 2.4. As seen in Table 2.4, our proposed approach Mdiff gives the 95% confidence band coverage rate around 95% in the global null case, the local alternative case and the global alternative case. The proposed Mdev performs similarly to Pnorm in all the cases, while the CR has poor performance.

2.5 Yeast Cell-Cycle Gene Expression Application

2.5.1 Background and Motivation

The cell cycle is a regulated life process leading to the division and replication of a cell. In cells with a nucleus, it consists of four sequential phases: GAP 1 (G1) for cell size growth and preparation for DNA replication, Synthesis (S) for DNA replication, GAP 2

(G2) for continual cell growth and preparation for mitosis, and Mitosis (M) for division of chromosomes and cytoplasm into two daughter cells. Since 1985 (Nasmyth), there has been collective evidence that transcription factors (TFs) regulate the expression of the cell-cycle-regulated genes at different stages of cell cycle. As reported in the literature for yeast (Simon et al., 2001), MBF and SBF function in the G1/S transition, Mcm1p regulates actively in early G1, M and the M/G1 boundary, and Swi5p and Ace2p takes account of many gene expressions in M and M/G1. However, the function of these TFs over the full cell cycle remains unknown. Other potential active TFs are also waiting to be discovered to function in the cell cycle to regulate the gene expression.

We apply our approach to a yeast cell cycle gene expression data set, appended with a derivative ChIP binding data set. The yeast cell cycle gene expression data set, collected over 2 cell cycles for 297 genes, is a subset of the data set collected by Spellman et al. (1998) to identify a catalog of protein-encoding transcripts in the genome of *Saccharomyces cerevisiae*. 6,178 yeast ORFs were examined simultaneously for their expression at mRNA levels, which were synchronized by α pheromone for G1 arrest, cdc15 for mitosis, and so on. Genes captured by α pheromone over two cell cycle periods were measured at 18 time points, for every 7 minutes within a total of 119 minutes. 297 genes were identified as cell-cycle-regulated by a model-based approach (Luan and Li, 2003). The derivative ChIP binding data set contains the binding probabilities of 96 TFs to these 297 cell-cycle-regulated genes, where each TF has at least 1 non-zero binding probability. These binding probabilities were obtained by applying the mixture model approach (Chen, Jensen, and Stoeckert 2007; Wang, Chen, and Li 2007) to the ChIP data of Lee et al. (2002).

Our first analysis goal is to apply the random effects time-varying coefficient model (2.3)-(2.7) to connect the log gene expression measures with a set of TFs. Let $y_i(t)$ denote the log-expression level of the i-th gene at time point t, and x_{ik} denote the

binding probability of the k-th TF to the i-th gene, for $k = 1, \ldots, 96$ and $i = 1, \ldots, 297$. Wang et al. (2008) linked the binding probabilities to the gene expression levels using a fixed effect nonparametric time-varying coefficient model:

$$y_i(t) = \beta_0(t) + \sum_{k=1}^{96} \beta_k(t) x_{ik} + \epsilon_i(t), \tag{2.8}$$

where $\beta_k(t)$ is the population mean transcription effect of the k-th TF on the log expression level of the i-th gene at time point t, and $\epsilon_i(t)$'s were assumed to be independent within and across genes. However, considering that different cell-cycle-regulated genes function in different processes such as DNA synthesis, budding, and cytokinesis (Spellman et al., 1998), one needs to account for the differential behaviors of the TFs in regulating different genes in the network of transcript regulations. To address the issue of differential transcription behaviors, we fit our random effects time-varying coefficient model :

$$y_i(t) = \beta_{i0}(t) + \sum_{k=1}^{96} \beta_{ik}(t) x_{ik} + \epsilon_i(t).$$

where we assume the gene-specific time-varying regulating course $\beta_{ik}(t)$ for the k-th TF on the i-th gene expression. We additionally construct Bayesian 95% confidence bands for the overall mean time-varying coefficient curves for each TF to explain the uncertainty in estimation.

Our second analysis goal is to identify the key TFs in cell-cycle regulated gene expression network and profile their active/inactive phases. Wang et al. (2008) identified the significant TFs through zeroing small regression coefficients using the SCAD penalty. However, they didn't incorporate the statistical inference error from the probability perspective. We apply the point-wise Bayesian local null hypothesis testing procedure to assess if a TF plays a significant role in regulating gene expression while

identifying the active and inactive phases of TF regulation across cell cycles. We limit type I errors by controlling expected FDR to be no more than 0.05.

In analyzing the data, we assume that between-gene expression levels are independent. For a specific gene i, within-gene expression levels are assumed to be independent given its baseline effect $\beta_{i0}(t)$ and transcription effects $\{\beta_{ik}(t)\}_{k=1}^{96}$. We place Dirichlet process and LPP2 priors on the distributions of the basis coefficients. Given that many of the log gene expression trajectories have similar shapes with only local deviation, we expect that the LPP2 prior will produce a more parsimonious representation of the log gene expression trajectories.

2.5.2 Analysis and Results

We focus initially on reanalyzing the yeast cell cycle gene data with the proposed approach. We assign $\text{Ga}(1,1)$ hyper priors for the hyperparameters α_0, γ_0, α_1, and γ_1, and a $\text{Ga}(0.1, 0.1)$ prior for the precision parameter τ. We recommend these priors as default values in other analyses of standardized data, because these priors are weakly informative across a wide range of plausible values for general longitudinal data, allowing the data to inform strongly. Gaussian kernels as proposed in the model of Section (2.2.2) are used as basis functions. We summarize the 1125 posterior samples from the MCMC output with thinning of 20 iterations after the burn-in of 2500 iterations. Using multiple chains with widely-distributed points, the proposed exact block Gibbs sampler exhibits good rates of convergence and mixing. This formulation took 113 seconds per iteration in Matlab 2010b on a Lenovo X61 laptop. We repeated the analysis for a variety of hyperparameters values. We multiplied the mean and variance of α_0, γ_0, α_1, and γ_1 by 2 and 0.5, and multiplied the variance of P_0 by 2 and 0.5. There were no noticeable differences in the results.

Our approach has the inherent ability to identify latent clusters in the data, which

is not available in the model of (2.8). The estimated values of the hyperparameters α_0 and α_1 were $\hat{\alpha}_0 = 0.15$ and $\hat{\alpha}_1 = 0.18$, with 95% credible intervals of $[0.01, 0.53]$ and $[0.01, 0.55]$. These values suggest few global and local clusters are present in the data, which means that a sparse representation of the data is obtained. The posterior probability of $\phi_0^* = 1$ was 0.96, suggesting that subjects are allocated to 1 global cluster for the baseline effects. The number of local clusters tended to be small, with the estimated averages of 1.07 and 2.34 for the baseline effect and the covariate effects. The estimated values of the hyperparameters γ_0 and γ_1 were 6.23 and 54.46, with 95% credible intervals of $[2.48, 11.12]$ and $[40.96, 88.53]$. Since the values of γ_0 and γ_1 close to zeros favors pure global clustering, while large values favor pure local clustering, there is clear evidence in the data favoring our approach over the Dirichlet process.

Wang et al. (2008) used the estimated time-varying effect curve and the confidence band to visualize peak effects. We conducted a formal Bayesian local hypothesis testing procedure to characterize the activation/deactivation phase of TFs in regulating the expression of cell cycle-regulated genes. The proposed Bayesian multiple hypothesis testing procedure gives the optimal rejection threshold $r^* = 0.886$ by controlling $\overline{FDR} \leq 0.05$. The decision rule rejects H_{0k}, $k = 1, \ldots, 96$ if any $\upsilon_{kt} \geq 0.886$. Figure 2.4 presents the posterior probability curves in favor of time-wise local alternative hypotheses for eight experimentally verified TFs to regulate gene expression during cell cycle. Our procedure found some interesting features not discovered in Wang et al. (2008). For example, Wang et al. (2008) didn't identify the fact that FKH2 is associated with genes expressed in G1 and S (Simon et al., 2001), while our procedure showed FKH2 to be active during G1/S phase. MBP1, SWI4 and SWI6 work as a complex to bind predominantly to promoter regions of G1 genes (Simon et al., 2001). Although Wang et al. (2008) found approximately similar peak effect time points for SWI4 and SWI6, Figure 2.4 clearly displays that they have similar active time windows.

We use our proposed Mdiff method to construct the 95% Bayesian confidence bands for the time-varying transcription effects. As presented in Figure 2.2, MBP1, SWI4, SWI6, MCM1, NDD1, FKH2, ACE2, and SWI5 were shown to have periodic transcriptional effect patterns over the collected two cell cycles, while Wang et al. (2008) only showed certain periodic patterns for SWI4 and FKH2. Since the Mdiff method considers the multiplicity issue in a conservative way similar to the Bonferroni method, the constructed confidence bands are wide. Methodology for improved construction of simultaneous confidence bands is an important area for future research.

We repeated the analysis using DP priors for $P \sim DP(\alpha_0 P_0)$ and $\tilde{P} \sim DP(\alpha_1 P_0)$. The posterior means of α_0 and α_1 were 2.43 and 0.58, with 95% credible intervals of $[1.26, 3.99]$ and $[0.14, 1.42]$. The posterior means of the number of clusters were 14 for the baseline effect and 3.99 for the transcription factor effects, reflecting a larger number of clusters than in the local partition process. In addition, the point-wise local hypothesis testing procedure didn't profile the active phases of the verified TFs in the literature, as illustrated in the Appendix Figure 2 in the supplementary material. This is likely due to the tendency of the Dirichlet process to overly favor clustering of subjects. In the situation of sparse basis coefficients, we recommend the local partition process for gains in inference.

2.6 Discussion

This chapter has proposed a nonparametric Bayesian model for longitudinal data analysis, which allows a flexible time-varying baseline and covariate effects for each subject, while characterizing variability among subjects using a local partition process prior. Posterior computation under the proposed Bayesian model is straightforward to implement efficiently, with the steps involved consisting of sampling sequentially from standard distributions. We have proposed novel methods for hypothesis testing and

estimation of simultaneous confidence bounds on the mean time-varying coefficients curves, with preliminary simulation studies showing excellent frequentist operating characteristics for these methods. A standard Bayesian nonparametric formulation based on Dirichlet process priors for the random effects has substantially worse performance in all cases we have considered. We hope that this illustration of the potential of nonparametric Bayes methods for longitudinal data analysis stimulates more work on developing novel nonparametric methods motivated by biomedical applications. In such settings, it is often the case that "off the shelf" priors, such as Dirichlet and Gaussian processes, can be substantially improved upon, with the applied context motivating modifications having more biologically realistic properties.

Table 2.1: Comparison of the LLP2 prior, the DP prior and the fixed effects model: mean square errors obtained from the 3 simulation cases: (i) global null case, (ii) local alternative case, and (iii) global alternative case.

	LPP2		DP		Fixed Effects	
	$\beta_{i0}(t)$	$\beta_{i1}(t)$	$\beta_{i0}(t)$	$\beta_{i1}(t)$	$\beta_{i0}(t)$	$\beta_{i1}(t)$
Global Null	0.536	0.969	0.561	9.571	0.989	1.047
Local Alternative	0.560	0.836	0.588	19.561	1.148	1.191
Global Alternative	1.300	2.673	1.365	16.741	1.580	1.719

Table 2.2: Type I error rates and powers of the global hypothesis tests under both LPP2 prior and DP prior obtained from 100 repeated data sets for the 3 simulation cases: (i) global null case, (ii) local alternative case, and (iii) global alternative case.

	LPP2	DP
Global Null	0.05	0.19
Local Alternative	0.72	0.63
Global Alternative	0.92	0.92

Table 2.3: The average of observed FDRs and powers for multiple hypothesis testing procedure under the LPP2 prior and the DP prior obtain from 100 simulated data sets for the 3 simulation cases: (i) global null case, (ii) local alternative case, and (iii) global alternative case.

	LPP2		DP	
	FDR	Power	FDR	Power
Global Null	0.042	-	0.034	-
Local Alternative	0.032	0.367	0.028	0.332
Global Alternative	-	0.998	-	0.973

Table 2.4: The coverage rates of 95% confidence band for 4 methods: Mdiff, Mdev, Pnorm, CR under the LPP2 prior for the 3 simulation cases: (i) global null case, (ii) local alternative case, and (iii) global alternative case.

	Mdiff	Mdev	Pnorm	CR
Global Null	0.987	0.733	0.880	0
Local Alternative	0.967	0.680	0.847	0
Global Alternative	0.950	0.645	0.774	0

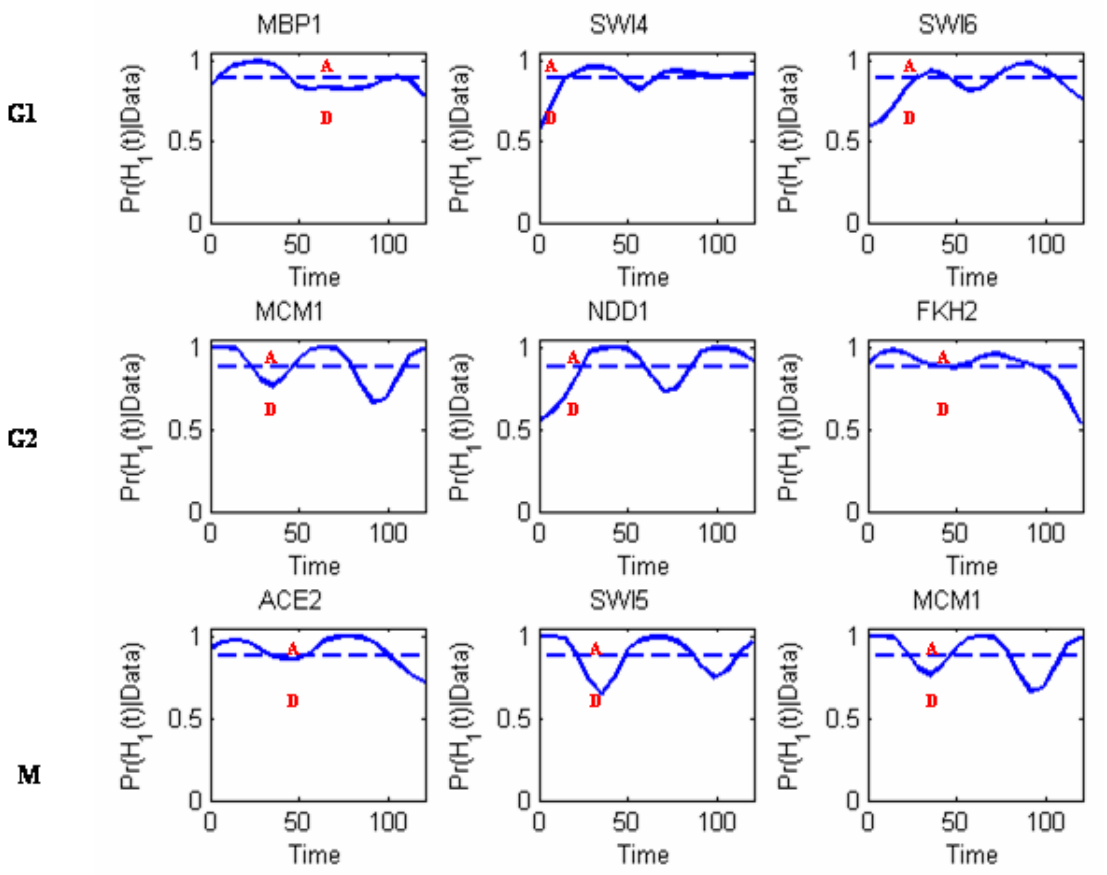

Fig. 2.1: Posterior probabilities of time-wise local alternative hypothesis for analyzing the yeast cell-cycle gene expression data using LPP2 prior. The dashed line corresponds to the optimal rejecting threshold $r^* = 0.886$. G1: TFs reported to regulate gene expression at the G1 phase; G2: TFs reported to regulate gene expression at the G2 phase; M: TFs reported to regulate gene expression at the M phase. A: active phase above the threshold level; D: deactive phase below the threshold level.

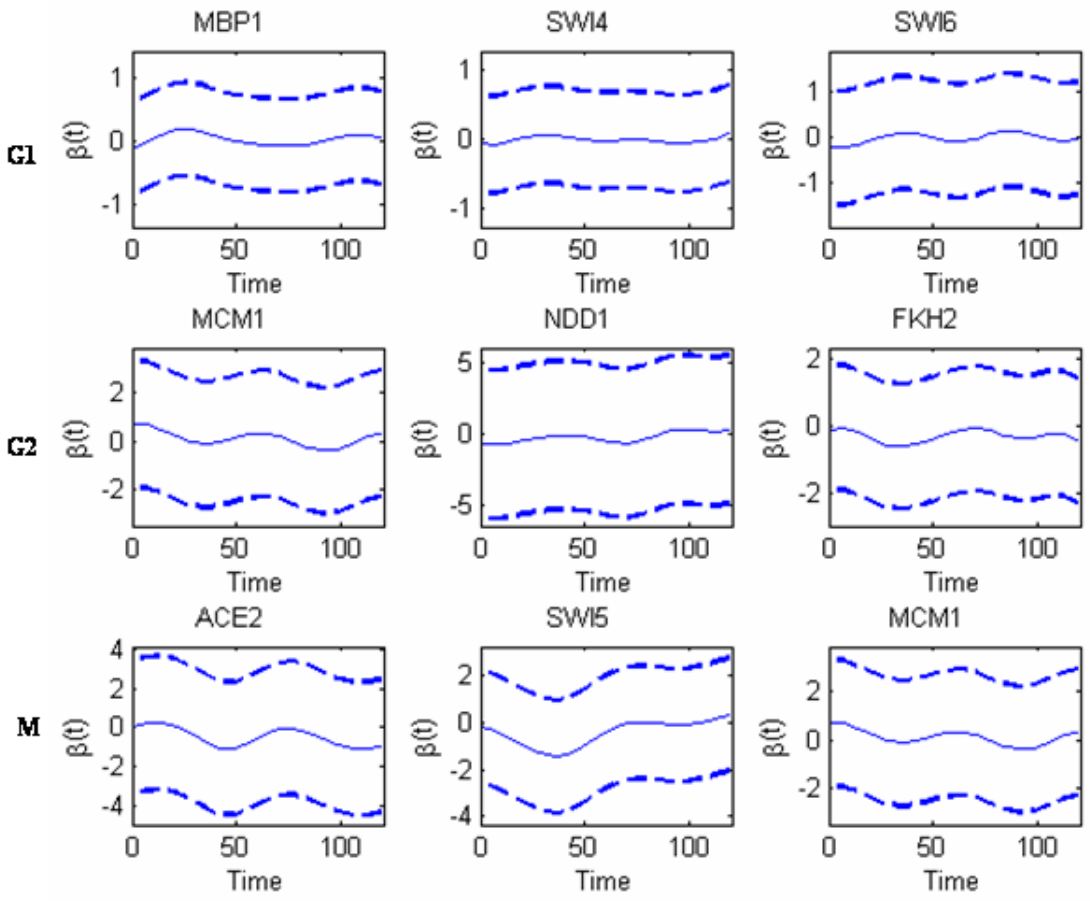

Fig. 2.2: Posterior estimated mean time-varying coefficients (solid line: —) and pointwise 95% confidence bands (dashed lines: - - -) for analyzing the yeast cell-cycle gene expression data. G1: TFs reported to regulate gene expression at the G1 phase; G2: TFs reported to regulate gene expression at the G2 phase; M: TFs reported to regulate gene expression at the M phase.

2.7 Appendix

2.7.1 Appendix A: Proof of Proposition 1

From the realization of the LPP2 prior through sticking-breaking process, regardless of which component a basis coefficient ϑ_j comes from, ϑ_j is drawn from the base Cauchy measure P_0. Then by the property of the linear combination of normal distributions remaining a normal distribution, conditional on $\Lambda^{-1} = \text{diag}\{\lambda_1^{-1}, \ldots, \lambda_p^{-1}\}$, the conditional distribution

$$f(t)|\Lambda^{-1} \sim N(0, B(t)'\Lambda^{-1}B(t)).$$

Furthermore, the covariance between $f(t_1)$ and $f(t_2)$ conditional on Λ^{-1} is easily derived as

$$Cov(f(t_1), f(t_2)|\Lambda^{-1}) = B(t_1)'\Lambda^{-1}B(t_2).$$

Then for a vector of such functions indexed by a finite time set, it follows that

$$[f(t_1), f(t_2), \ldots, f(t_L)]'|\Lambda^{-1} \sim N_L(0_{L\times 1}, \Sigma_{L\times L})$$

with $\Sigma_{L\times L} = [\sigma_{ij}]$, $i = 1, \ldots, L, j = 1, \ldots, L$, and $\sigma_{ij} = B(t_i)'\lambda_1^{-1}B(t_j)$.

2.7.2 Appendix B: the MCMC Posterior Computation Algorithm

We define the following notations for convenient use in the description of the sampler:

- $W_0 = (\text{w}'_{01}, \ldots, \text{w}'_{0n})'$, with $\text{w}_{0i} = (b(t_1), \ldots, b(t_{m_i}))'$, $i = 1, \ldots, n$,
- $Z_0 = \text{diag}(1(z_{101} = 1 - j), \ldots, 1(z_{10p} = 1 - j), \ldots, 1(z_{n01} = 1 - j), \ldots, 1(z_{n0p} = 1 - j))$,

- $\Phi_0 = \text{diag}(1(\phi_{101} = h), \ldots, 1(\phi_{10p} = h), \ldots, 1(\phi_{n01} = h), \ldots, 1(\phi_{n0p} = h))$,
- $R_0 = (r_0', \ldots, r_0')'_{np \times p}$ with $r_0 = \text{diag}(1, \ldots, 1)_{p \times p}$,
- $Y = (y_1', \ldots, y_n')'$,
- $Z_{0Y} = \text{diag}(1(\sum_{l=1}^{p} 1(z_{10l} = 1 - j) \geq 1), \ldots, 1(\sum_{l=1}^{p} 1(z_{n0l} = 1 - j) \geq 1))_{n \times n}$,
- $\Phi_{0Y} = \text{diag}(1(\sum_{l=1}^{p} 1(\phi_{10l} = h) \geq 1), \ldots, 1(\sum_{l=1}^{p} 1(\phi_{n0l} = h) \geq 1))_{n \times n}$.
- $W = (\text{w}_1', \ldots, \text{w}_n')'$, with $\text{w}_i = (B_i(t_1), \ldots, B_i(t_{m_i}))'$ for $i = 1, \ldots, n$,
- $Z = \text{diag}(1(z_{111} = 1 - j), \ldots, 1(z_{11p} = 1 - j), \ldots, 1(z_{1K1} = 1 - j), \ldots, 1(z_{1Kp} = 1 - j), \ldots, 1(z_{n11} = 1 - j), \ldots, 1(z_{n1p} = 1 - j), \ldots, 1(z_{nK1} = 1 - j), \ldots, 1(z_{nKp} = 1 - j))$,
- $\Phi = \text{diag}(1(\phi_{111} = h), \ldots, 1(\phi_{11p} = h), \ldots, 1(\phi_{1K1} = h), \ldots, 1(\phi_{1Kp} = h), \ldots, 1(\phi_{n11} = h), \ldots, 1(\phi_{n1p} = h), \ldots, 1(\phi_{nK1} = h), \ldots, 1(\phi_{nKp} = h))$,
- $R = (r', \ldots, r')'_{npK \times pK}$ with $r = \text{diag}(1, \ldots, 1)_{pK \times pK}$,
- $Y = (y_1', \ldots, y_n')'$,
- $Z_Y = \text{diag}(1(\sum_{k=1}^{K} \sum_{l=1}^{p} 1(z_{1kl} = 1 - j) \geq 1), \ldots, 1(\sum_{k=1}^{K} \sum_{l=1}^{p} 1(z_{nkl} = 1 - j) \geq 1))_{n \times n}$,
- $\Phi_Y = \text{diag}(1(\sum_{k=1}^{K} \sum_{l=1}^{p} 1(\phi_{1kl} = h) \geq 1), \ldots, 1(\sum_{k=1}^{K} \sum_{l=1}^{p} 1(\phi_{nkl} = h) \geq 1))_{n \times n}$,

Our sampling algorithm proceeds as follows:

Step 1. Update the subject-specific baseline $\beta_{i0}(t)$ part

1. Update the latent u_{i0l} from its conditional distribution $\text{Unif}(0, \pi_{0\phi_{i0l}})$ for $l = 0, 1, \ldots, p$.

2. Update the latent z_{i0l} from its conditional distribution $\text{Ber}(p_{i0l})$, with

$$p_{i0l} = \frac{\nu_{0l} N(y_i; \theta_{i0(z_{i0l}=1)}, \theta_i, \tau)}{\nu_{0l} N(y_i; \theta_{i0(z_{i0l}=1)}, \theta_i, \tau) + (1-\nu_{0l}) N(y_i; \theta_{i0(z_{i0l}=0)}, \theta_i, \tau)},$$

where $\theta_{i0(z_{i0l}=j)}$ refers to the current value of θ_{i0} with inserting $\Theta_{00\phi_{i0l}l}$ to the lth component for $j=1$, and $\Theta_{01\phi_{i0l}l}$ for $j=0$.

3. Update the stick-breaking variable π^*_{0h} with its conditional distribution

$$\text{Beta}(\sum_{l=0}^{p}\sum_{i=1}^{n} 1(\phi_{i0l}=h)+1, \sum_{l=0}^{p}\sum_{i=1}^{n} 1(\phi_{i0l}>h)+\alpha)$$

for $h \le \phi^*_0$, with $\phi^*_0 = \max\{\phi_{i0l}, i=1,\ldots,n, l=0,1,\ldots,p\}$; for $h > \phi^*_0$, sample it from $\text{Beta}(1,\alpha_0)$.

4. Update ϕ_{i0l} with its conditional probability $\Pr(\phi_{i0l}=h) \propto 1(h \in A_{i0l}) N(y_i; \theta_{i0(\phi_{i0l}=h)}, \theta_i, \tau)$, where $A_{i0l} = \{h : \pi_{0h} > u_{i0l}\} \subset \{1,2,\ldots,\infty\}$, which is obtained by sampling π^*_{0h} for $h=1,\ldots,\tilde{\phi}_0$ with $\tilde{\phi}_0$ the smallest value satisfying

$$\sum_{h=1}^{\tilde{\phi}_0} \pi^*_{0h} \prod_{l<h}(1-\pi^*_{0l}) \ge 1-u^*_0,$$

where $u^*_0 = \min\{u_{i0l}, i=1,\ldots,n, l=0,1,\ldots,p\}$.

5. Update Θ_{0jh} from its conditional distribution $N_p(\mu_{0jh}, \Sigma_{0jh})$, where

$$\Sigma_{0jh} = [\text{diag}(\lambda_{0jh},\ldots,\lambda_{0jh})_{p\times p} + \tau W'_{0jh} W_{0jh}]^{-1} \text{ and } \mu_{0jh} = \tau \Sigma_{0jh} W'_{0jh} Y_{0jh},$$

in which Y_{0jh} refers to the contribution for Θ_{0jh} from subjects with $z_{i0l} = 1-j$ and $\phi_{i0l} = h$. In addition, $W_{0jh} = \text{diag}(\text{w}_{01},\ldots,\text{w}_{0n}) Z_0 \Phi_0 R_0$, and $Y_{0jh} = \text{diag}(y_1,\ldots,y_n) Z_{0Y} \Phi_{0Y} \mathbf{1}_n$, where $\mathbf{1}_n$ is a $n \times 1$ vector of entries of 1.

6. Update λ_{0jh} from its conditional distribution $\text{Ga}(0.5+p/2, 0.5+\Theta'_{0jh}\Theta_{0jh}/2)$.

7. Update ν_{0l} from its conditional distribution $\text{Beta}(1+\sum_i z_{i0l}, \gamma_0+\sum_i(1-z_{i0l}))$.

8. Update the hyperparameter γ_0 from its conditional distribution $\text{Ga}(a_\gamma + p, b_\gamma - \sum_{l=1}^{p}\log(1-\nu_{0l}))$.

9. Update the hyperparameter α_0 from its conditional distribution $\text{Ga}(a_\alpha + \phi_0^*, b_\alpha - \sum_{h=1}^{\phi_0^*}\log(1-\pi_{0h}^*))$.

Step 2. Sample the coefficient regression functions $\{\beta_{ik}(t)\}_{k=1}^{K}$ part in a manner similar to *Step* 1.

1. Update the latent u_{ikl} from its conditional distribution $\text{Unif}(0, \pi_{\phi_{ikl}})$ for $k = 1, \ldots, K$, $l = 0, 1, \ldots, p$.

2. Update the latent z_{ikl} from its conditional distribution $\text{Ber}(p_{ikl})$, with

$$p_{ikl} = \frac{\nu_{kl} N(y_i; \theta_{i0}, \theta_{i(z_{ikl}=1)}, \tau)}{\nu_{kl} N(y_i; \theta_{i0}, \theta_{i(z_{ikl}=1)}, \tau) + (1-\nu_{kl}) N(y_i; \theta_{i0}, \theta_{i(z_{ikl}=0)}, \tau)},$$

where $\theta_{i(z_{ikl}=j)}$ refers to the current value of θ_i with inserting $\Theta_{0\phi_{ikl}l}$ to the lth component for $j = 1$, and $\Theta_{1\phi_{ikl}l}$ for $j = 0$.

3. Update the stick-breaking variable π_h^* with its conditional distribution

$$\text{Beta}(\sum_{l=0}^{p}\sum_{k=1}^{K}\sum_{i=1}^{n} 1(\phi_{ikl} = h) + 1, \sum_{l=0}^{p}\sum_{k=1}^{K}\sum_{i=1}^{n} 1(\phi_{ikl} > h) + \alpha)$$

for $h \leq \phi^*$, with $\phi^* = \max\{\phi_{ikl}, i = 1, \ldots, n, k = 1, \ldots, K, l = 0, 1, \ldots, p\}$; for $h > \phi^*$, sample it from $\text{Beta}(1, \alpha_1)$.

4. Update ϕ_{ikl} with its conditional probability $\Pr(\phi_{ikl} = h) \propto 1(h \in A_{ikl}) N(y_i; \theta_{i0}, \theta_{i(\phi_{ikl}=h)}, \tau)$, where $A_{ikl} = \{h : \pi_h > u_{ikl}\} \subset \{1, 2, \ldots, \infty\}$, which is obtained by sampling

π_h^* for $h = 1, \ldots, \tilde{\phi}$ with $\tilde{\phi}$ the smallest value satisfying

$$\sum_{h=1}^{\tilde{\phi}} \pi_h^* \prod_{l<h} (1 - \pi_l^*) \geq 1 - u^*,$$

where $u^* = \min\{u_{ikl}, i = 1, \ldots, n, k = 1, \ldots, K, l = 0, 1, \ldots, p\}$.

5. Update $\Theta_{jh} = (\Theta_{jh1}, \ldots, \Theta_{jhK})$ and λ_{jhk}, $j = 0, 1$:

a. Update Θ_{jh} from its conditional distribution $N_p(\mu_{jh}, \Sigma_{jh})$, where $\Sigma_{jh} = [\text{diag}(\lambda_{jh1}1_p, \ldots, \lambda_{jhK}1_p)_{pK\times pK} + \tau W_{jh}' W_{jh}]^{-1}$ and $\mu_{jh} = \tau \Sigma_{jh} W_{jh}' Y_{jh}$, with Y_{jh} referring to the contribution for Θ_{jh} from subjects with $z_{ikl} = 1 - j$ and $\phi_{ikl} = h$, so as to W_{jh}, for $l = 1, \ldots, p$. Specifically, $W_{jh} = \text{diag}(\text{w}_1, \ldots, \text{w}_n) Z \Phi R$, and $Y_{jh} = \text{diag}(y_1, \ldots, y_n) Z_Y \Phi_Y \mathbf{1}_n$, where $\mathbf{1}_n$ is a $n \times 1$ vector of entries of 1.

b. Update λ_{jhk} from its conditional distribution $\text{Ga}(0.5 + p/2, 0.5 + \Theta_{jhk}' \Theta_{jhk}/2)$.

6. Update ν_{kl} from its conditional distribution $\text{Beta}(1 + \sum_i z_{ikl}, \gamma_1 + \sum_i (1 - z_{ikl}))$.

7. Update the hyperparameter γ_1 from its conditional distribution $\text{Ga}(a_\gamma + pK, b_\gamma - \sum_{k=1}^K \sum_{l=1}^p \log(1 - \nu_{kl}))$.

8. Update the hyperparameter α_1 from its conditional distribution $\text{Ga}(a_\alpha + \phi^*, b_\alpha - \sum_{h=1}^{\phi^*} \log(1 - \pi_h^*))$.

Step 3. Update the precision parameter τ with its conditional distribution $\text{Ga}(a_\tau + \sum_i n_i/2, b_\tau + \sum_i \{y_i - \beta_{i0}(t) - \sum_{k=1}^K \beta_{ik}(t) x_{ik}\}' \{y_i - \beta_{i0}(t) - \sum_{k=1}^K \beta_{ik}(t) x_{ik}\}/2)$.

Appendix Figure 1

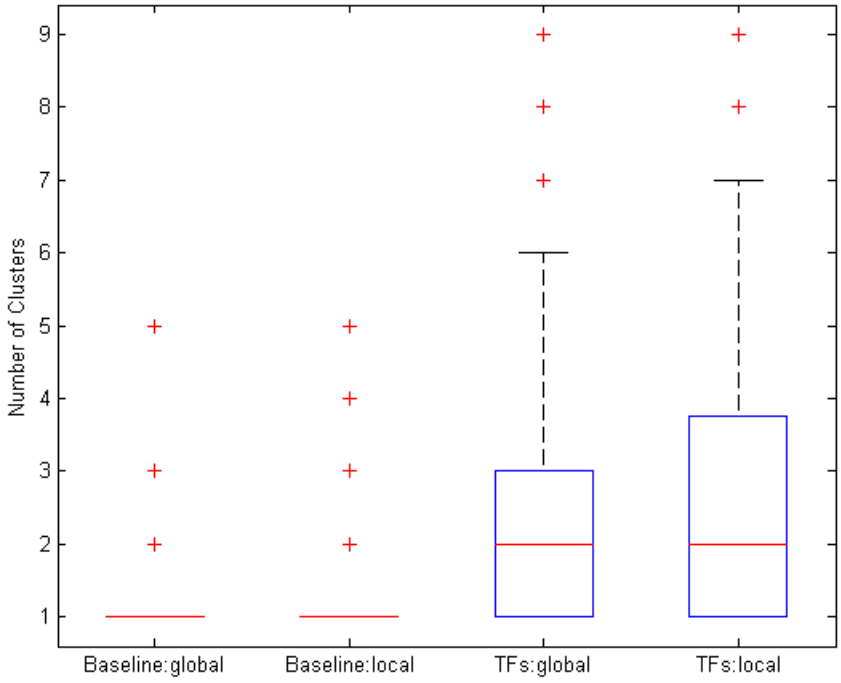

Fig. 2.3: Boxplots showing the number of global and local clusters for baseline effects and TFs effects

Appendix Figure 2

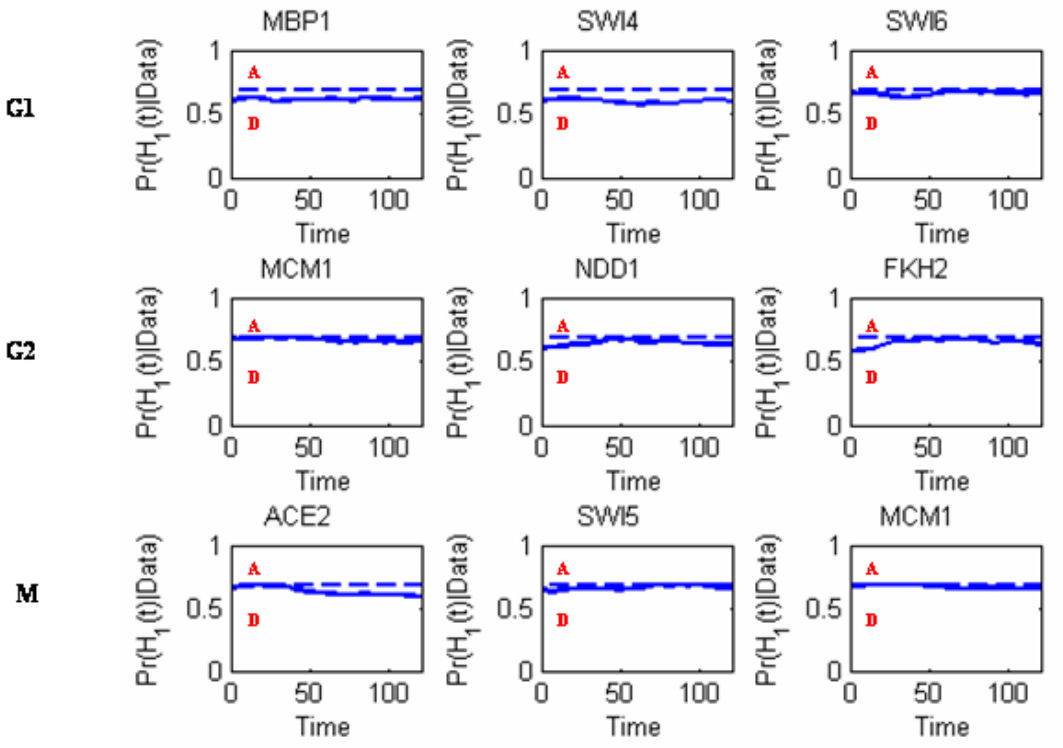

Fig. 2.4: Posterior probabilities of time-wise local alternative hypothesis for analyzing the yeast cell-cycle gene expression data using DP prior. The dashed line corresponds to the optimal rejecting threshold $r^* = 0.683$. G1: TFs reported to regulate gene expression at the G1 phase; G2: TFs reported to regulate gene expression at the G2 phase; M: TFs reported to regulate gene expression at the M phase. A: active phase above the threshold level; D: deactive phase below the threshold level.

Chapter 3

Semiparametric Bayesian Local Functional Models for Diffusion Tensor Tract Statistics

3.1 Introduction

Diffusion tensor imaging (DTI) is an magnetic resonance imaging (MRI) technique that measures the diffusion orientation of water molecules in tissue. The mobility of water molecules is affected by the tissue properties of white matter fiber tracts, such as the density of the fibers, the average fiber diameter, and the directionality of the fibers. In turn, the information collected on the diffusion properties of water molecules provides the structural organization of the white matter fiber tracts (Kubicki et al., 2007). The water diffusion directions and magnitudes at each voxel in the brain can be described by a 3×3 symmetric positive matrix, called a diffusion tensor (DT) (Basser et al., 1994a,b). The degree of diffusivity can be quantified by the three eigenvalue-eigenvector pairs of DT, and its related parameters, such as fractional anisotropy (FA)

(Pierpaoli and Basser, 1996; Hasan et al., 2001; Hasan and Narayana, 2003; Zhu et al., 2006). A rich literature in neuroimaging has been developed to analyze white matter fiber tract maturation and integrity via a set of water diffusion parameters such as FA, mean diffusivity (MD) and so on, used as markers (Moseley, 2002; Mukherjee and McKinstry, 2006; Cascio et al., 2007; Rollins, 2007).

Three major analytic approaches for grouping analysis of DTI data set have been explored including analyses based on region-of-interest (ROI), voxels, and fiber tracts (Smith et al., 2006; O'Donnell et al., 2009; Snook et al., 2007). The region-of-interest (ROI) method (Bonekam et al., 2008; Gilmore et al., 2008) suffers from difficulties in identifying meaningful ROIs, particularly the long curved structures common in fiber tracts, instability of statistical results, and the partial volume effect in relatively large ROIs (Snook et al., 2007). Voxel based analysis (Chen et al., 2009; Focke et al., 2008; Camara et al., 2007; Snook et al., 2005) suffers from issues of alignment quality and the arbitrary choice of smoothing extent (Hecke et al., 2009; Ashburner and Friston, 2000; Smith et al., 2006; Jones et al., 2005).

There is an extensive interest in the DTI literature in developing fiber tract based analysis of diffusion properties (Smith et al., 2006; O'Donnell et al., 2009; Yushkevich et al., 2008; Goodlett et al., 2009; Zhu et al., 2010, 2011; Goldsmith et al., 2011). Such analyses usually consist of DTI atlas building and a follow-up statistical analysis (Smith et al., 2006; Goodlett et al., 2009; Zhu et al., 2010). DTI atlas building is primarily to establish DTI correspondence across all DTI datasets from different subjects and to extract a set of individual DTI tracts (or skeleton) with the same corresponding geometry but varying DTs and diffusion properties. For instance, Smith et al. (2006) developed a *tract-based spatial statistics* framework to construct local diffusion properties along the white matter skeleton, fitted pointwise linear regression models, and performed pointwise hypothesis tests on the skeleton. This method essentially ignores the functional

nature of diffusion properties along the white matter skeleton, and thus suffers from low statistical power in detecting interesting features and exploring variability in functional data. Goodlett et al. (2009) used functional principal component analysis (fPCA) coupled with the Hotelling T^2 statistic to compare a univariate diffusion property, such as fractional anisotropy, across two (or more) populations for a single hypothesis test per tract. Zhu and his coauthors (Zhu et al., 2010, 2011) proposed a multivariate varying coefficient model based on fPCA for the analysis of fiber bundle multiple diffusion properties and their association with a set of covariates of interest such as age. Greven et al. (2010) developed a functional mixed effects model as a generalization of mixed effects models and fPCA for the analysis of longitudinal DTI fiber tract data. So far, frequentist inference is the primary approach for making statistical inferences in these statistical models for fiber tract based analysis of diffusion properties.

In this chapter, we propose a new semiparametric Bayes approach to model the association between multiple fiber bundle diffusion properties and covariates of interest. A multivariate random coefficient model is developed to characterize heterogeneity in the shape of the fiber bundle diffusion properties among subjects, while allowing the impact of covariates to vary across subjects. We assume a nonparametric Bayesian LPP2 prior (Dunson, 2009) on the distribution of the random coefficients to facilitate global and local borrowing of information among subjects. We consider a sparse latent factor model to more flexibly capture the within-subject correlation structure and assume a multiplicative gamma process shrinkage prior on the factor loadings which allows introduction of infinitely many factors (Bhattacharya and Dunson, 2011). We propose Bayesian local hypothesis testing to identify fiber segments, where multiple diffusion properties are significantly associated with covariates of interest, while controlling for multiple comparisons. We propose a Bayesian confidence band for the average effect of each covariate. Finally, we use the nonparametric LPP2 prior to randomly cluster

subjects via global and local clustering. Posterior computation proceeds via an efficient Markov chain Monte Carlo (MCMC) algorithm using the exact block Gibbs sampler.

3.2 Methodologies

This chapter focuses on developing a semiparametric Bayesian multivariate functional regression analysis pipeline, named as SBLFM, to assess the association between fiber bundle diffusion properties and a set of covariates of interest (e.g., age). Before SBLFM, we use DTI atlas building followed by atlas fiber tractography and fiber parametrization as described in Goodlett et al. (2009) to extract DTI fibers and establish DTI fiber correspondence across all DTI datasets from different subjects. We skip its description here for the sake of simplicity (Goodlett et al., 2009; Zhu et al., 2010). After performing the DTI atlas building step, we obtain a set of individual fiber tracts (or skeleton) with the same corresponding geometry but varying DTs and diffusion properties. Subsequently, we run the analysis pipeline SBLFM including a multivariate random varying coefficient model, a semiparametric Bayesian estimation method, a Bayesian approach to construct confidence bands, a Bayesian local hypothesis testing procedure, and a Bayesian posterior cluster analysis procedure (Fig. 3.1). The computational algorithm for SBLFM is developed by using Matlab.

3.2.1 Multivariate Random Coefficient Model

Assume n subjects are measured along a fiber bundle over a grid of T points for M diffusion properties (e.g. FA), denoted by $\{Y_i(d_t) : i = 1, \ldots, n, t = 1, \ldots, T\}$, where $Y_i(d_t) = (Y_i^{(1)}(d_t), \ldots, Y_i^{(M)}(d_t))'$ is an $M \times 1$ vector of diffusion properties for the i-th subject, $d_t \in [0, L]$ is the arc length of point t relative to a fixed end point of the fiber bundle, and L is arc length of this fiber bundle. For the i-th subject, we relate the m-th functional diffusion response $Y_i^{(m)}(d_t)$ to a set of K covariates via a multivariate

random coefficient model given by

$$Y_i^{(m)}(d_t) \;=\; \boldsymbol{\beta}_i^{(m)}(d_t)'X_i + \epsilon_i^{(m)}(d_t), \;\; \epsilon_i^{(m)}(d_t) \sim N(0, \sigma_{m,t}^2), \tag{3.1}$$

where $X_i = (1, \boldsymbol{x}_i')'$ is the design vector for the i-th subject with $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{iK})'$ being a $K \times 1$ vector of covariates of interest, and $\boldsymbol{\beta}_i^{(m)}(d_t) = (\beta_{i,0}^{(m)}(d_t), \ldots, \beta_{i,K}^{(m)}(d_t)'$ is a $(K+1) \times 1$ vector of random coefficients for the m-th diffusion property. The measurement error $\epsilon_i^{(m)}(d_t)$ is assumed to be drawn from a normal distribution with mean zero and standard deviation $\sigma_{m,t}$.

We model the random coefficient function $\beta_{i,k}^{(m)}(d_t)$ by using a linear combination of cubic B-spline basis functions $b_l(d_t)$s as follows:

$$\beta_{i,k}^{(m)}(d_t) = \sum_{l=1}^{p} \eta_{ikl}^{(m)} b_l(d_t) = \mathbf{b}(d_t)'\boldsymbol{\eta}_{ik}^{(m)}, \tag{3.2}$$

where $\mathbf{b}(d_t) = (b_1(d_t), \ldots, b_p(d_t))'$ is a $p \times 1$ vector for B-spline basis functions and $\boldsymbol{\eta}_{ik}^{(m)} = (\eta_{ik1}^{(m)}, \ldots, \eta_{ikp}^{(m)})'$ is a $p \times 1$ vector of random coefficients. Thus, the model (3.1) can be written as

$$Y_i^{(m)}(d_t) = B_i(d_t)'\boldsymbol{\eta}_i^{(m)} + \boldsymbol{\epsilon}_i^{(m)}(d_t) \text{ for } m = 1, \ldots, M, \tag{3.3}$$

where $\boldsymbol{\eta}_i^{(m)} = (\boldsymbol{\eta}_{i0}^{(m)'}, \ldots, \boldsymbol{\eta}_{iK}^{(m)'})'$ and $B_i(d_t) = X_i \otimes \mathbf{b}(d_t)$ are $(K+1)p \times 1$ vectors. Heterogeneity in the fiber bundle diffusion properties is then controlled by the variation of $\boldsymbol{\eta}_i^{(m)}$, which are usually treated as random effects and follow a specific parametric distribution. There is a serious concern about the sensitivity of associated inferences to the choice of the random effects distribution on the basis coefficients.

Let $\boldsymbol{\eta}_i = (\boldsymbol{\eta}_i^{(1)'}, \ldots, \boldsymbol{\eta}_i^{(M)'})'$ be an $M(K+1)p \times 1$ vector, $B_i = (B_i(d_1), \ldots, B_i(d_T))'$ be a $T \times (K+1)p$ matrix, and $Y_i = (Y_i^{(1)'}, \ldots, Y_i^{(M)'})'$ be an $MT \times 1$ vector, where

$Y_i^{(m)} = (Y_i^{(m)}(d_1), \ldots, Y_i^{(m)}(d_T))'$ is a $T \times 1$ vector for $m = 1, \ldots, M$. The concantenated representation of model (3.3) is given by

$$Y_i = (I_M \otimes B_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{MT}(0, \Sigma), \tag{3.4}$$

where $N_{MT}(\mathbf{0}, \Sigma)$ is an $MT \times 1$ Gaussian random vector with mean zero and covariance matrix $\Sigma = \text{diag}(\sigma^2_{1,1}, \ldots, \sigma^2_{M,T})$.

3.2.2 Infinite Latent Factor Model

Given the massive dimensionality of the random effect vector $\boldsymbol{\eta}_i$, it is important to favor lower dimensional representations of the dependence structure to address the curse of dimensionality. Instead of prespecifying a restrictive dependence structure, we follow the approach of using a Bayesian factor model, which relates the random effects $\boldsymbol{\eta}_i$ to latent factors $\boldsymbol{\theta}_i$ through the following characterization

$$\boldsymbol{\eta}_i = A\boldsymbol{\theta}_i, \tag{3.5}$$

where A is a $(K+1)pM \times \infty$ factor loadings matrix and $\boldsymbol{\theta}_i \sim N_\infty(0, I_\infty)$. For dimensionality reduction, one would typically restrict the dimension of the latent factor vector $\boldsymbol{\theta}_i$ to be orders of magnitude less than that of $\boldsymbol{\eta}_i$. However, following the motivation of Bhattacharya and Dunson (2011), we bypass the challenging issue of selecting the number of factors by incorporating infinitely many factors, while choosing a prior that favors the elements of A to be shrunk to zero increasingly as the column index increases.

We obtain the within-subject correlation structure through projecting Y_i as a linear combination of the underlying ∞-dimensional latent random vector θ_i after denoising,

via

$$Y_i = (I_M \otimes B_i)A\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{MT}(0, \Sigma), \tag{3.6}$$

where $B_i(d_t) = X_i \otimes \mathbf{b}(d_t)$ and $B_i = (B_i(d_1), \ldots, B_i(d_T))^T$. The common factors $\boldsymbol{\theta}_i$ explain the within-subject dependence structure among the MT variables for each Y_i.

Let Y_{if} and σ_f^2 be, respectively, the f-th component of Y_i and the f-th diagonal element of Σ. Conditional on $\boldsymbol{\theta}_i$, Y_{if_1} and Y_{if_2} are uncorrelated for all f_1 and f_2 $\in \{1, \ldots, MT\}$. Marginalizing over the distribution of $\boldsymbol{\theta}_i$, the covariance structure Ω of the data distribution $Y_i \sim N_{MT}(0, \Sigma)$ is induced as

$$\Omega = \mathrm{Var}(Y_i|A, \Sigma) = (I_M \otimes B_i)AA^T(I_M \otimes B_i)^T + \Sigma.$$

Specifically, conditional on A and Σ, we have:

$$\begin{aligned}
\mathrm{Var}(Y_{if_1}|\boldsymbol{\theta}_i, A, \Sigma) &= \sigma_{f_1}^2, \\
\mathrm{Cov}(Y_{if_1}, Y_{if_2}|\boldsymbol{\theta}_i, A, \Sigma) &= 0, \\
\mathrm{Var}(Y_{if_1}|A, \Sigma) &= (I_M \otimes B_iA)_{f_1}(I_M \otimes B_iA)_{f_1}^T + \sigma_{f_1}^2, \\
\mathrm{Cov}(Y_{if_1}, Y_{if_2}|A, \Sigma) &= (I_M \otimes B_iA)_{f_1}(I_M \otimes B_iA)_{f_2}^T,
\end{aligned}$$

where $(I_M \otimes B_iA)_{f_1}$ is the f_1th row of $(I_M \otimes B_iA)$.

A constraint is usually specified on A to define a unique model free from identification problems, since Ω is invariant under the transformations $A^* = AP$ for any semi-orthogonal matrix P ($PP^T = I$). The traditional full rank lower triangular constraint for identifiability implicity specifies order dependence among the responses (Geweke and Zhou, 1996). The choice of ordering of the response variables is a modeling decision (Carvalho et al., 2008). From the Baysian perspective, we don't require identifiability

of the loading elements in A for a wide class of applications such as covariance matrix estimation. In our case, we specify the multiplicative gamma process shrinkage prior, which will be given in (3.7), on a parameter expanded loadings matrix with redundant parameters. The induced prior on the covariance matrix is invariant to ordering of the data. This shrinkage prior adaptively selects a truncation of the infinite loadings to one having finite columns, which facilitates the posterior computation and provides an accurate approximation to the infinite factor model.

Although FRATS (Zhu et al., 2010) and FADTTS (Zhu et al., 2011) have been proposed to analyze the DTI fiber bundle data sets of multiple diffusion measures with a set of covariates such as gender and gestational age, they assume identical effects of the covariates on the diffusion properties for all subjects. Our SBLFM developed here relaxes the assumption to allow the subject specific functional covariate effects on the multiple diffusion outcome functions. In addition, FRATS and FADTTS assume a covariate-free correlation structure among the multiple diffusion properties. We introduce the random factors underlying the subject specific covariate effects to allow the correlation structure among the multiple diffusion measures to vary with the levels of the covariates along the fiber tracts. Moreover, we specify a LPP2 prior on the distributions of the subject specific random factors to produce a global and local clustering structure of the multiple diffusion trajectories across the subjects.

3.2.3 Priors

We choose independent priors for A, $\boldsymbol{\theta}_i$, and Σ and develop an efficient Markov chain Monte Carlo (MCMC) algorithm given in the Appendix for posterior computation. To remove the redundant factors, we place the multiplicative gamma process shrinkage prior (Bhattacharya and Dunson, 2011) on the loadings matrix A to increasingly shrink the loadings toward zero with the column index. This specification avoids the

traditional drawback of order dependence from the lower triangular constraint for identifiability. We use the inverse gamma priors on the diagonal elements of Σ. To allow the latent clustering structure among subjects, we specify a LPP2 prior (Dunson, 2009) on the prior distribution P of $\boldsymbol{\theta}_i$. Let $\mathrm{Ga}(a, b)$ be a gamma distribution with scale a and shape b. Specifically, these priors in details are given as follows:

$$\begin{aligned}
A &= \{A_{gh}\}, \;\; g = 1, \ldots, (K+1)pM \; ; h = 1, \ldots, \infty, \\
A_{gh}|\phi_{gh}, \tau_h &\sim N(0, \phi_{gh}^{-1}\tau_h^{-1}), \;\; \phi_{gh} \sim \mathrm{Ga}(v/2, v/2), \;\; \tau_h = \prod_{l=1}^{h} \delta_l, \\
\delta_1 &\sim \mathrm{Ga}(a_1, 1), \;\; \delta_l \sim \mathrm{Ga}(a_2, 1), \;\; l \geq 2, \\
\sigma_f^{-2} &\sim \mathrm{Ga}(a_\sigma, b_\sigma), f = 1, \ldots, MT, \\
\boldsymbol{\theta}_i &\sim P, \;\; P \sim \mathrm{LPP2}(\alpha, \gamma, P_0), \;\; P_0 : \boldsymbol{\theta}_i \sim N_\infty(0, \zeta I_\infty), \;\; \zeta \sim \mathrm{Ga}(0.5, 0.5),
\end{aligned} \tag{3.7}$$

where δ_l, $l = 1, \ldots, \infty$, are independent, τ_h is a global shrinkage parameter for the hth column, and the ϕ_{gh}s are local shrinkage parameters for the elements in the hth column. When $a_2 > 1$, the τ_hs increase stochastically as the column index h increases, which means more shrinkage favored over the columns of higher indexes. The loading component specific prior precision $\phi_{gh}^{-1}\tau_h^{-1}$ allows shrinking the coefficients of the B-spline basis functions. The LPP2 prior assumption specifies an ∞-dimensional central probability measure P_0 as a prior guess, a precision parameter α expressing confidence in the prior guess, and a hyperparameter γ determining the overall allocation weight on the local family.

3.2.4 LPP2 prior

In modelling the multiple trajectories of diffusion measures over subjects, we are interested in identifying the latent cluster structure of Y_is among subjects, which can be induced by specifying a nonparametric prior on the unknown distribution of the

random factors $\boldsymbol{\theta}_i$. Dirichlet process prior (Ferguson, 1973, 1974) is the routine approach to induce a sparse representation of subjects. However, it has the drawback of global clustering, in which two subjects are either allocated to an identical cluster or two different clusters. In DTI imaging data analysis, it is common that two subjects have similar diffusion trajectories over the major region of fiber bundles while having local deviations. Under the Dirichlet process, either such subjects are inappropriately clustered together, obscuring local differences, or allocated to separate clusters causing unnecessary computational burden. We used the LPP2 prior (Dunson, 2009) to model the unknown random effects distributions to facilitate both global and local clustering of random effects. It relaxes the global clustering assumption of the widely-used Dirichlet process prior, while accomplishing sparseness for substantial gains in computational efficiency.

After truncating the loadings matrix A to $h^* \ll (K+1)pM$ columns, we assume the resulting h^*-dimensional random effects $\theta_i = (\theta_{i1}, \ldots, \theta_{ih^*})^T \sim P$ with P unknown, $i = 1, \ldots, n$. The LPP2 prior models P as a hybrid mixture distribution:

$$P = \sum_{z_1=0}^{1} \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_h^*=0}^{1} \sum_{\psi_{h^*}=(1-z_h^*,1)}^{(1-z_{h^*},\infty)} \pi_{\psi_1,\ldots,\psi_{h^*}} \delta_{\Theta_\psi}, \tag{3.8}$$

where δ_x denotes a degenerate distribution with all its mass at x, $\pi_{\psi_1,\ldots,\psi_{h^*}}$ is the probability of $\theta_i = \Theta_\psi$ having the property that $\pi_{\psi_1,\ldots,\psi_{h^*}} \geq 0$ and

$$\sum_{z_1=0}^{1} \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_{h^*}=0}^{1} \sum_{\psi_{h^*}=(1-z_{h^*},1)}^{(1-z_{h^*},\infty)} \pi_{\psi_1,\ldots,\psi_{h^*}} = 1.$$

The indicator $z_j \sim \text{Ber}(\nu_j)$, $j = 1, \ldots, h^*$, denotes the allocation to global clustering or local clustering, taking the value of 1 for global clustering and 0 for local clustering. The hybrid atom $\Theta_\psi = (\Theta_{\psi_1,1}, \ldots, \Theta_{\psi_{h^*},h^*})^T$ is obtained by setting the jth element of Θ_ψ

equal to $\Theta_{\psi_j,j}$ with $\psi_j = (1 - z_j, \phi_j)$ and $\Theta_{\psi_j} \sim P_0$ for $\phi_j \in \{1, \ldots, \infty\}$, $j = 1, \ldots, h^*$, where P_0 is a base distribution.

Let π_h denote $\Pr(\theta_{ij} = \Theta_{ghj})$, $g = 0, 1$, $h = 1, \ldots, \infty$, $j = 1, \ldots, h^*$. Those random effects having $z_j = 1$ are assigned together to a component in the global family $\Xi_0 = \{\Theta_{0h}\}_{h=1}^{\infty}$, while others having $z_j = 0$ are assigned to their own component in the local family $\Xi_1 = \{\Theta_{1h}\}_{h=1}^{\infty}$. Let $J_0 = \{j : z_j = 1\}$ and $J_1 = \{j : z_j = 0\}$. Conditional on the values of z, we have

$$\begin{aligned}
&\Pr(\psi_j = (0, h), j \in J_0 | z_1, \ldots, z_{h^*}) = \pi_h, \ h = 1, \ldots, \infty, \\
&\Pr(\psi_j = (1, h_j), j \in J_1 | z_1, \ldots, z_{h^*}) = \prod_{j \in J_1} \pi_{h_j}, \ h_j = 1, \ldots, \infty.
\end{aligned}$$

The allocation probability of $\theta_i = \Theta_\psi$ in (3.8) is then simply

$$\pi_{\psi_1,\ldots,\psi_{h^*}} = \Pr(\psi_1 = (1 - z_1, h_1), \ldots, \psi_{h^*} = (1 - z_{h^*}, h_{h^*})) = \pi_h \Big\{ \prod_{j \in J_1} \pi_{h_j} \Big\} \Big\{ \prod_{j=1}^{h^*} \nu_j^{z_j} (1 - \nu_j)^{1 - z_j} \Big\}.$$

The specification is completed by choosing the hyperpriors

$$\begin{aligned}
&\nu_j \sim \text{Beta}(1, \gamma), \quad j = 1, \ldots, h^*, \\
&\pi_h = \pi_h^* \prod_{l<h} (1 - \pi_l^*), \quad \pi_h^* \sim \text{Beta}(1, \alpha), \quad h = 1, \ldots, \infty,
\end{aligned}$$

where γ controls the overall weight on the local family, and α controls the overall number of clusters. As short hand, the LPP2 prior on the random effects distribution P is denoted by $P \sim \text{LPP2}(\alpha, \gamma, P_0)$.

The LPP2 prior specification (3.8) can also be viewed as a hybrid mixture model of infinitely many components drawn from P_0 via 2-stage clustering. Stage 1 determines the membership of global or local clustering for each of the h^* components. Those components allocated to the global clustering membership will be clustered together

to an atom in the global family at stage 2, while those allocated to the local family will be allocated individually to their own clusters. The joint cluster membership probability at stage 1 corresponds to $\Pr(z_1, \ldots, z_{h^*}) = \prod_{j=1}^{h^*} {\nu_j}^{z_j}(1-\nu_j)^{1-z_j}$. The joint cluster allocation weight at stage 2 conditional on stage 1 corresponds to $\Pr(\psi_1 = (1-z_1, h_1), \ldots, \psi_{h^*} = (1-z_{h^*}, h_{h^*})|z_1, \ldots, z_{h^*})$. The overall joint cluster allocation weight corresponds to $\pi_{\psi_1,\ldots,\psi_{h^*}}$.

3.2.5 Hypothesis Formulation & Testing

We are interested in making pointwise inference for the covariate effects along a fiber bundle. The local null and alternative hypotheses for the kth covariate effect on the mth diffusion property specific to a location $d_t \in [0, L]$ are formulated as:

$$H_{0k}^{(m)}(d_t) : \left|\overline{\beta}_k^{(m)}(d_t)\right| \leq \epsilon \text{ versus } H_{1k}^{(m)}(d_t) : \left|\overline{\beta}_k^{(m)}(d_t)\right| > \epsilon,$$

where $\overline{\beta}_k^{(m)}(d_t)$ represents the mean of the subject-specific random coefficients for the kth predictor on the mth diffusion property at location d_t. The zero-neighborhood size ϵ is chosen as being proportional to the maximum posterior standard deviation of the posterior samples of $\beta_{i,k}^{(m)}(d_t)$ multiplied by ϵ^*. It is important to be aware that the data have been normalized prior to analysis.

To conduct local hypotheses testing, we use the Bayesian decision rule for multiple testing proposed by Müller et al. (2004). Following the implementation in Wang and Dunson (2010), our strategy is to reject $H_{0k}^{(m)}(t)$ if the posterior alternative hypothesis probability $\upsilon_{kt}^{(m)} = \Pr(H_{1k}^{(m)}(t)|Data) \geq r$ for any $t \in [0, T]$, with r as a common threshold for all the local hypotheses. r is chosen to minimize the posterior expected false negative rate (FNR) under the constraint of the posterior expected false discovery rate (FDR) being no greater than α_T, where α_T is pre-specified (we focus on a value

of 0.05). Denote $d_{kt}^{(m)} = 1(\upsilon_{kt}^{(m)} \geq r)$, an indicator of rejecting $H_{0k}^{(m)}(t)$. The posterior expected FNR and FDR are calculated as:

$$\overline{FNR} = \frac{\sum_{m=1}^{M}\int_0^T (1-d_{kt}^{(m)})\upsilon_{kt}^{(m)}\mathrm{d}t}{M\int_0^T \mathrm{d}t - \sum_{m=1}^{M}\int_0^T d_{kt}^{(m)}\mathrm{d}t + \kappa_0} \approx \frac{\sum_{m=1}^{M}\sum_{t=t_1}^{t_W}(1-d_{kt}^{(m)})\upsilon_{kt}^{(m)}T/W}{KT - \sum_{m=1}^{M}\sum_{t=t_1}^{t_W} d_{kt}^{(m)}T/W + \kappa_0},$$

$$\overline{FDR} = \frac{\sum_{m=1}^{M}\int_0^T d_{kt}^{(m)}(1-\upsilon_{kt}^{(m)})\mathrm{d}t}{\sum_{m=1}^{M}\int_0^T d_{kt}^{(m)}\mathrm{d}t + \kappa_0} \approx \frac{\sum_{m=1}^{M}\sum_{t=t_1}^{t_W} d_{kt}^{(m)}(1-\upsilon_{kt}^{(m)})T/W}{\sum_{m=1}^{M}\sum_{t=t_1}^{t_W} d_{kt}^{(m)}T/W + \kappa_0},$$

where $t_1, \ldots, t_W$ is a fine grid of points equally spaced along $[0, T]$ and κ_0 is a small positive constant to avoid a zero denominator. In summary, our decision rule is to determine the optimal threshold r^* by $r^* = \mathrm{argmin}\, \overline{FNR}\{r \in [0,1], \overline{FDR} \leq \alpha_T\}$. For each combination of m and $t \in \{t_1, \ldots, t_W\}$, we reject $H_{0k}^{(m)}(t)$ if $\upsilon_{kt}^{(m)} \geq r^*$. For a sufficiently fine grid, the results are robust to w, with the optimal threshold appropriately adapting to the chosen w.

3.2.6 Bayesian Confidence Bands

We construct a Bayesian simultaneous confidence (credible) band for the mean coefficient curve $\overline{\beta}_k^{(m)}(t)$, $k = 1, \ldots, K$, $m = 1, \ldots, M$, from its posterior MCMC samples. Assuming there is a collection of posterior sampled curves $\beta_k^{(m),s} = (\beta_k^{(m),s}(t_1), \ldots, \beta_k^{(m),s}(t_L))^T$, $s = 1, \ldots, S$ indexing posterior iterations after burn-in, our goal is to compute a simultaneous confidence band for $\overline{\beta}_k^{(m)}(t)$. The principle in constructing a Bayesian confidence band is to search for a region $R_\alpha = \{R_\alpha(t), t \in [0,T]\}$ such that

$$\Pr\{\overline{\beta}_k^{(m)}(t) \in R_\alpha(t), t \in [0,T]\} = 1 - \alpha,$$

where α is a pre-specified significance level.

The strategy is based on pointwise measures of uncertainty. We used the method

proposed by Crainiceanu et al. (2007) which assumes approximate posterior normality at each grid point and derives the $1-\alpha$ sample percentile $c_{1-\alpha}$ of

$$\max_{l=1,\ldots,L} \left| \frac{\beta_k^{(m),s}(t_l) - \hat{\beta}_k^{(m)}(t_l)}{\sqrt{\widehat{\text{var}}(\hat{\beta}_k^{(m)}(t_l))}} \right|, \quad s = 1, \ldots, S,$$

where $\hat{\beta}_k^{(m)}(t_l)$ is the posterior mean at time t_l and $\sqrt{\widehat{\text{var}}(\hat{\beta}_k^{(m)}(t_l))}$ is its posterior standard deviation. A simultaneous confidence region is given by the hyperrectangular

$$[\hat{\beta}_k^{(m)}(t_l) - c_{1-\alpha}\sqrt{\widehat{\text{var}}(\hat{\beta}_k^{(m)}(t_l))}, \hat{\beta}_k^{(m)}(t_l) + c_{1-\alpha}\sqrt{\widehat{\text{var}}(\hat{\beta}_k^{(m)}(t_l))}], \quad l = 1, \ldots, L.$$

In our implementation, we replaced $c_{1-\alpha}$ by c_b which is calculated by $c_b = \max(|c_{\alpha/2}|, |c_{1-\alpha/2}|)$ to account for skewness.

3.2.7 Bayesian Cluster Analysis

We propose a Bayesian clustering approach to identify the number of clusters and to probabilistically assign each individual to the identified clusters. The nonparametric LPP2(α, γ, P_0) prior has the feature of randomly clustering subjects via global and local clustering. Given a truncated h^*-dimensional random effects vector θ_i for the i-th subject, the LPP2 prior assigns those elements having $z_{ij} = 1$ for $j = 1, \ldots, h^*$ together to a global cluster, while assigning the remaining elements having $z_{ij} = 0$ to their own clusters. If subjects i and i' are in the same cluster for the j-th random effect, such that $\theta_{ij} = \theta_{i'j}$, the LPP2 prior will propagate such clustering information to other random effects, increasing the probability of $\theta_{ij'} = \theta_{i'j'}$, $j' \neq j$.

The set of probability weights assignment for both global clustering and local clustering are formulated by a sticking-breaking process:

$$\phi_{ij} = \sum_{h=1}^{\infty} \pi_h \delta_h, \;\; \pi_h = \pi_h^* \prod_{l<h} (1 - \pi_l^*), \;\; \pi_h^* \sim \text{Beta}(1, \alpha), \;\; j = 0, 1, \ldots, h^*,$$

where ϕ_{ij} indexes the assigned cluster for θ_{ij}, with $j = 0$ for global clustering and $j = 1, \ldots, h^*$ for each element's own clustering, and π_h is the weight for a set of the elements (global clustering) or an element (local clustering) to be assigned to the h-th cluster. Small α favors the assigned clusters to be concentrated on the first few ones to achieve sparseness for posterior computation efficiency.

In practical posterior computation, we developed a posterior cluster analysis procedure on the clustering of the subject specific latent random factors $\{\boldsymbol{\theta}_i : i = 1, \ldots, n\}$. We specified a LPP2 prior on the unknown distributions of the subject specific latent factors $\boldsymbol{\theta}_i$. Denote $o_{g,\theta}^b$ for the number of global clusters among $\{\boldsymbol{\theta}_i : i = 1, \ldots, n\}$ at the b-th iteration in the Gibbs sampler, and similarly denote $o_{l,\theta}^b$ for the number of local clusters. We obtained the posterior samples of $\{o_{g,\theta}^b : b = 1, \ldots, B\}$ and $\{o_{l,\theta}^b : b = 1, \ldots, B\}$. The overall weight allocated to the local cluster family equals *a priori* to $1/(1+\gamma)$ according to the specification of LPP2(α, γ, P_0) in Dunson (2009). Given the posterior sample average $\hat{\gamma}$, the proportion of local clustering *a posteriori* for $\boldsymbol{\theta}_i$ is estimated by $1/(1+\hat{\gamma})$.

3.3 Monte Carlo Simulations

We conducted a set of Monte Carlo simulations to evaluate the false discovery rate (FDR) and the power of the local hypothesis testing approach. We simulated the two diffusion measures FA and MD along the right internal capsule tract obtained in the

clinical study from the following model:

$$
\begin{aligned}
&(\mathrm{FA}_i(d_t), \mathrm{MD}_i(d_t))^T = (\mathbf{x}_i^T B_i^{(1)}(d_t), \mathbf{x}_i^T B_i^{(2)}(d_t))^T + \boldsymbol{\epsilon}_i(d_t), \\
&\mathbf{x}_i^T B_i^{(m)}(d_t) = (\beta_{0i}^{(m)}(d_t), \beta_{1i}^{(m)}(d_t), \beta_{2i}^{(m)}(d_t))\mathbf{x}_i, \quad m = 1, 2,
\end{aligned} \tag{3.9}
$$

where $d_t \in [0, L]$, $\mathbf{x}_i = (1, G_i, A_i)$ with G_i and A_i referring to the gender and gestational age of the i-th subject, $B_i^{(m)}(d_t) = (\beta_{0i}^{(m)}(d_t), \beta_{1i}^{(m)}(d_t), \beta_{2i}^{(m)}(d_t))$ is a 3×1 coefficient vector with its component $\beta_{ki}^{(m)}(d_t)$, $k = 0, 1, 2$, to be drawn from a Gaussian process $\mathrm{GP}(\beta_k^{(m)}(d_t), \sigma_0^2 \rho^{d_t - d_t'})$, and the measurement error $\boldsymbol{\epsilon}_i(d_t)$ is a 2×1 vector of Gaussian random variables with mean zero and covariance matrix $\Sigma(d_t)$. To mimic the realistic clinic imaging data, we obtain the estimates $\hat{\beta}_k^{(m)}(d_t)$ of $\beta_k^{(m)}(d_t)$ and $\hat{\Sigma}(d_t)$ of $\Sigma(d_t)$ by analyzing the FA and MD measures along the right internal capsule using the methodology proposed by Zhu et al. (2011). We set $\sigma_0^2 = 0.004$ and $\rho = 0.5$. Further, fixing $(\beta_{0i}^{(m)}(d_t), \beta_{1i}^{(m)}(d_t))$ for all m and d_t at their obtained estimates from the clinic data, we let $(\beta_2^{(1)}(d_t), \beta_2^{(2)}(d_t)) = c(d_t)(\hat{\beta}_2^{(1)}(d_t), \hat{\beta}_2^{(2)}(d_t))$ and consider five scenarios for $(\beta_2^{(1)}(d_t), \beta_2^{(2)}(d_t))$: null case of $c(d_t) = 0$ and local alternative cases for $c(d_t) = 0.2, 0.4, 0.6$, and 0.8.

We analyzed each simulated scenario using 100 repeated simulations. We set $n = 32, 64$. We randomly chose 16 males and 16 females for $n = 32$ (32 males and 32 females for $n = 64$) from our clinical data and used their values of gender and gestational age to simulate the values of FA and MD along the right internal capsule tract. We assigned $\mathrm{Ga}(1, 1)$ hyper priors for the hyperparameters α and γ, and $\mathrm{Ga}(2, 1)$ priors on a_1 and a_2. We specified $\nu = 3$ and chose $5 \log((K + 1)pM)$ as the starting number of factors. The exact block Gibbs sampler ran 10,000 iterations, with the first 5000 samples discarded as burn-in. Every 5th sample was collected to thin the chain. For each case, our Gibbs sampler converges rapidly and exhibits efficient mixing.

We conducted the local hypothesis testing of $H_0^{(m)}(d_t) : \left|\beta_2^{(m)}(d_t)\right| \leq \epsilon$ for a point d_t along the right internal capsule tract and the m-th diffusion property against $H_1^{(m)}(d_t) : \left|\beta_2^{(m)}(d_t)\right| > \epsilon$. We specified $\epsilon = 0.1, 0.15$, and 0.2. We evaluated FDR for the local hypothesis testing procedure at the null case $c(d_t) = 0$ and the local alternative cases $c(d_t) = 0.2, 0.4, 0.6$, and 0.8, and its powers at the local alternative case $c(d_t) = 0.2, 0.4, 0.6$, and 0.8. For each simulation, the significant levels for the local hypothesis testing procedure were set at $\alpha = 0.01$, and 100 replications were used to estimate the FDR and powers. For a fixed α at the alternative cases, if FDR is smaller than α, the testing procedure is conservative, whereas if it is observed to be greater than α, the testing procedure is liberal. At the null case, FDR equals to 1 based on its definition. Table 4.1 displays that the FDR is accurate for the sample sizes of 32 and 64. Consistent with our expectations, the statistical power for rejecting the local null hypotheses increases with the sample size and decreases with the zero neighborhood size ϵ.

3.4 Clinical DTI Fiber Tract Data

3.4.1 Background and Analysis

The DTI fiber tract data comes from a clinical study approved by the Institute Review Board of the University of North Carolina at Chapel Hill. This larger study was designed to investigate early brain development. Healthy infants less than 1-year old were recruited with written informed consents obtained from their patients before imaging acquisition. In our study, a total of 128 healthy full-term infants (75 males and 53 females) are included, whose mean gestational age at MR scanning is 298$\pm$17.6 days (range: 262 to 433 days). All infants were placed with efforts such that they slept comfortably inside the MR scanner and none of them was sedated during the imaging

procedure. They were fed and calmed to sleep on a warm blanket with suitable ear protection.

The device used to acquire all images is a 3T Allegra head only MR system (Siemens Medical Inc., Erlangen, Germany), featuring in 40 mT/m for the maximal gradient strength and 400 mT/(m.msec) for the maximal slew rate. Every subject was scanned for contiguous slices with slice thickness of 2mm to cover the whole brain. Each slice was repeated 5 times such that an average was obtained to improve the signal-to-noise ratio. These acquired DTI images are obtained via a single shot EPI DTI sequence (TR/TE=5400/73 msec) with eddy current compensation. Diffusion gradients were collected at six non-collinear directions:(1,0,1), (-1,0,1), (0,1,1), (0,1,-1), (1,1,0), and (-1,1,0), at the b-value level of 1000 s/mm^2. The reference scan at the b-value of 0 was implemented to construct diffusion tensor matrices. The voxel resolution was set at isotropic 2mm, and the in-place field of view was set at 256mm in both directions.

The diffusion tensors were constructed using a weighted least square estimation method (Zhu et al., 2007; Basser et al., 1994b). We then processed all 128 DTI data sets using the image processing steps in the DTI atlas building and compute diffusion properties along all fiber tracts of interest. We focused on analyzing two tracts of interest including the splenium of the corpus callosum tract and the right internal capsule tract (Fig. 3.2(a) and Fig. 3.7(a)). We computed fractional anistropy (FA), mean diffusivity (MD), and the three eigenvalues of the diffusion tensors, denoted by, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, at each grid point on both tracts for each of the 128 subjects. FA denotes the inhomogeneous extent of local barriers to water diffusion, and MD measures the averaged magnitude of local water diffusion. The three eigenvalues of diffusion tensor reflect the magnitude of water diffusivity along and perpendicular to the long axis of white matter fibers (Song et al., 2003).

We focused on addressing four goals in the analysis of the multiple diffusion properties along the right internal capsule tract and the splenium tract. Our first goal was to determine if there is a difference in the multiple diffusion properties along the fiber bundles between male and female infants. Our secondary goal was to investigate the development of the diffusion properties along the fiber bundles vs gestational age. Our third goal was to accomplish a parsimonious representation of the multiple diffusion trajectories across subjects. Our fourth goal was to examine the covariate-dependent correlation pattern among the multiple diffusion properties along the fiber bundles.

We applied SBLFM (3.9) to model the smoothed FA and MD functions (M=2) along the right internal capsule and the splenium capsule with $x_i = (G_i, A_i)^T$, where G_i and A_i refer to the gender (1 for female and 0 for male) and the gestational age of the ith infant. We estimated the ith subject specific functional regression coefficients $\beta_i^m(d_t)$ for the mth functional diffusion measure. We then estimated the mean coefficient curve $\overline{\beta}_k^{(m)}(d_t)$ *a posteriori* and constructed their 95% Bayesian confidence bands from the posterior MCMC samples. We also performed local hypothesis testing procedures to determine if and where gender and gestational age are significant in the development of the diffusion trajectories along both of these two fiber tracts. The posterior cluster analysis is conducted to describe how many clusters exist among the 128 infants in terms of the water diffusion properties FA and MD. The correlation analysis is then implemented to illustrate the development of the correlation pattern between FA and MD along both fiber tracts for male infants of average gestational age of 298 days, as well as the pattern between the gestational age effects on FA and MD. Finally, based on the significant results for MD, we analyzed the three eigenvalues of diffusion tensors with FA along the two fiber tracts.

We assigned Ga$(1, 1)$ hyper priors for the hyperparameters α and γ, a Ga$(0.1, 0.1)$ prior for the precision parameters σ_f^{-2}, and Ga$(2, 1)$ priors on a_1 and a_2. We specified

$\nu = 3$ and a default choice of $5\log((K+1)pM)$ as the initialized number of factors. We recommend these settings as default values in other analyses of DTI fiber tract data. We summarized the 1000 posterior samples from the MCMC output with thinning of 5 iterations after the burn-in of 5000 iterations. Using multiple chains with widely-distributed points, the proposed exact block Gibbs sampler exhibits good rates of convergence and mixing. This formulation took 519 seconds per 100 iterations in Matlab 2010b on a Lenovo X61 laptop. We repeated the analysis for a variety of hyperparameter values. We multiplied the mean and variance of α and γ by 2 and 0.5, and multiplied the variance of P_0 by 2 and 0.5. We used $\nu = 3.5, 4, 5$ and varied b_σ between 0.1 and 0.5. We also used different starting number of factors between 3 and 10. The results are robust with no conclusions changed.

3.4.2 Results: Right Internal Capsule Tract

We initially focused on analyzing the smoothed FA and MD curves (Fig. 3.2(b) and (c)) of the 128 infants, which were obtained from the adaptive local polynomial kernel smoothing technique. Our primary interest lies in investigating if and where gender and gestational age are significant in the development of FA and MD along the right internal capsule tract. We conducted Bayesian local hypothesis testing procedures at each grid point along the tract to examine the effects of gender and gestation age on FA and MD values. Fig. 3.3(c) showed that the degrees of both fractional anisotropy and mean diffusivity vary significantly between males and females along partial regions of the right internal capsule tract. The posterior mean effect curves of gender on FA and MD and their 95% confidence bands (Fig. 3.3(a) and (b)) provided more evidence that females have less fractional anisotropy than males in a major region of the right internal capsule tract, and females have stronger mean diffusivity than males in half of the tract with weaker differences in the other half. The difference in mean diffusivity between

males and females are also observed in the sample average MD values (Fig. 3.3(d)). Such sophisticated differences between males and females were not discovered by the previous FRADS and FADTTS methods. Similar to FRADS and FADTTS, Fig. 3.4(b) found uniformly significant effects of gestational age on the FA and MD trajectories. Additional supportive evidence were observed that FA increases with gestational age (Fig. 3.4(a) and (c)), while MD decreases with gestational age (Fig. 3.4(d)).

We performed posterior cluster analysis to examine latent clusters in the FA and MD trajectories along the right internal capsule tract across the 128 infants. The estimated value of the hyperparameter γ, controlling allocation weight on the local family, equals 0.51 with 95% credible interval of $[0.12, 1.20]$. This implies that the posterior proportion of local clustering is 1/3, indicating that the data favors global clustering of the FA and MD trajectories across infants. The estimated value of the hyperparameter $\alpha = 5.34$, with 95% credible interval of $[2.59, 9.65]$. These values suggest that few clusters are present in the data and a sparse representation of the data is obtained. This conclusion is further supported by the estimated values of 4.45, 3.94, and 4.42 for the number of overall clusters, global clusters, and local clusters respectively (Fig. 3.2(d)).

Along the right internal capsule tract, we further computed the posterior mean curves of FA and MD (Fig. 3.5(a)) for males of average gestational age, and estimated the posterior covariance matrix (Fig. 3.5(b) and (c)). Then we obtained correlation coefficient between FA and MD at each grid point for males of average gestation age. We observed negative correlations at around 60% of grid points and positive correlation at the other around 40% of grid points (Fig. 3.5(d)). Moreover, we computed the posterior covariance matrix for the gestational age effects on FA and MD values (Fig. 3.5(f) and (g)) and derived the grid pointwise correlation between the gestational age effects on FA and MD. Fig. 3.5(h) displayed negative correlations at most grid points.

In the analysis of the three eigenvalues with FA, we focused on the association between gender and gestation age with the three eigenvalues and the correlation among the three eigenvalues along the right internal capsule tract. We examined the association between both covariates of interest and all eigenvalues along the tract by performing local hypothesis testing at each grid point. A similar locally significant pattern of gender effect was found for all three eigenvalues (Fig. 3.6(b)). Gestational age effect was found uniformly significant for λ_2 and λ_3, while locally head-and-tail significant for λ_1 (Fig. 3.6(c)). We observed positive correlation between λ_2 and λ_3 for both males and females of mean gestational age (Fig. 3.6(d) and (e)), and positive correlation between the gestational age effects on λ_2 and λ_3 (Fig. 3.6(f)). Overall, these correlations are stronger than those between λ_1 and λ_3 or λ_1 and λ_2. This agrees with the small differences between λ_2 and λ_3, while the relatively large differences between λ_1 and λ_3 or λ_1 and λ_2 in the middle of the right internal capsule tract (Fig. 3.2(e), (f) and (g)). Finally, we performed a cluster analysis among the three eigenvalues and obtained a sparse representation of the three eigenvalue functions along the right internal capsule tract (Fig. 3.2(h)).

3.4.3 Results: Splenium Tract

We applied SBLFM (3.9) to analyze the FA and MD trajectories along the splenium tract from 128 infants. We started with the Bayesian local hypothesis testing procedure to identify the region in which gender or gestational age is significantly associated with FA and MD. Fig. 3.8 (c) shows that gender is only slightly significant in the head region for MD development in the splenium tract. The posterior estimation for the mean effects of gender and the corresponding 95% confidence bands agrees with the testing results. Fig. 3.8(d)-(e) display the uniformly increasing effects of gestational age on FA development and decreasing effects on MD development at a major region. The

local hypothesis testing results (Fig. 3.8(f)) support that gestational age is uniformly significant along the splenium tract for the developmental change of the degree of fractional anisotropy, while locally significant for that of the degree of mean diffusivity.

The posterior sampling distribution of the number of clusters among the 128 infants is given in Fig. 3.7 (d). The posterior mean estimates of the number of overall clusters, global clusters, and local clusters are 3.42, 3.07, and 3.40 respectively, indicating that the 128 infants concentrate on a few clusters. The weight parameter γ for allocation on the local family was estimated to be 0.38 with 95% credible interval of $[0.08, 0.98]$. The estimated posterior mean of the proportion of local clustering is 0.28, which means a favoring of global clustering on infants. The estimated posterior mean of α equals to 6.46, with 95% credible interval of $[2.94, 11.80]$. This result agrees with the sparse clusters identified in the data.

The posterior estimation of the mean curves of FA and MD for males of average gestational age and the posterior estimates of their covariance matrices is given in Fig. 3.9 (a)-(c). Fig. 3.9 (d) displays that males of average gestational age have positive correlations between FA and MD at a major region of the splenium tract. Furthermore, Fig. 3.9 (e)-(g) displays the estimated gestational age functional effect curves and their corresponding covariance matrices. We observed that the gestational age effects on FA and MD are negatively correlated almost over the whole splenium tract (Fig. 3.9 (h)), which is consistent with the opposite trend of gestational age on FA and MD (Fig. 3.9 (e)).

We additionally found that both gender and gestational age are locally significantly associated with the developmental pattern of the three eigenvalues along the splenium tract (Fig. 3.10 (b) and (c)). We further observed large correlation between λ_2 and λ_3 and relatively weak correlation between λ_1 and λ_2 or λ_1 and λ_3 for both male and female of average gestational age. This agrees with the small difference between λ_2 and λ_3

while relatively large difference between λ_1 and λ_3 in the first half of the splenium tract (Fig. 3.7(e), (f) and (g)). We didn't observe noticeable difference among the pairwise correlations of the gestational age effects on the three eigenvalues (Fig. 3.10(h)).

3.5 Discussion

The contributions of our work are twofold. From the statistical perspective, we have developed a new Bayesian functional analysis pipeline SBLFM for delineating the structure of the variability of multiple diffusion properties along major white matter fiber bundles and their association with a set of covariates of interest. The SBLFM pipeline integrates five advanced Bayesian statistical tools from the statistical literature. Compared with the existing literation (Zhu et al., 2010, 2011), our SBLFM explicitly model the subject-specific functional covariate effects on the multiple diffusion outcome functions and the covariate-specific correlation structure among the multiple diffusion measures. From the application perspective, we have demonstrated SBLFM in a clinical study of neurodevelopment. Our results have revealed the complex inhomogeneous spatiotemporal maturation patterns of fiber bundle diffusion properties.

Several limitations need to be addressed in future research. First, since the existing fiber tract based methods including SBLFM requires DTI fiber correspondence across subjects, SBLFM cannot be used to investigate some scenarios in practice. For instance, it is possible that the centroid of the localization of white matter lesion may vary across time and subjects. In this case, one cannot use both ROI-based methods and tract based methods. Secondly, although we have used the B-spline basis functions in SBLFM, it is interesting to consider other basis functions, such as wavelet. Thirdly, SBLFM is limited to diffusion properties along fiber bundle, but it is interesting to consider full DTs and other representations based on high angular resolution diffusion image (HARDI) (Lenglet et al., 2009; Tuch et al., 2002; Schwartzman, 2006; Lepore

et al., 2008; Schwartzman et al., 2005; Zhu et al., 2009; Whitcher et al., 2007). Fourthly, SBLFM can be readily extended to more complex brain structures, such as the medial manifolds of fiber tracts (Yushkevich et al., 2008), functional neuroimaging data (Bowman et al., 2008; Woolrich et al., 2004; Lei et al., 2009; Gössl et al., 2001), and group analysis of neuroimaging data (Rosa et al., 2010; Penny et al., 2007).

3.6 Appendix: Posterior Computation

We develop an efficient Markov chain Monte Carlo (MCMC) algorithm for posterior computation. After truncating the loadings matrix A to $h^* \ll (K+1)pM$ columns, we adapt an efficient exact block Gibbs sampler (Papaspiliopoulos, 2008) for Dirichlet mixture models. This exact block Gibbs sampler combines the advantages of the retrospective MCMC method (Papaspiliopoulos and Roberts, 2008) and the slice sampling method (Walker, 2007). It introduces auxiliary variables to avoid truncated approximations. This algorithm is straightforward to implement and exhibits good performance in convergence and mixing. We will describe the adaptation strategy on the truncated level h^* after introducing the Gibbs sampler.

Allowing data to inform information about the hyperparameters involved in the model through $(3.1)-(3.7)$, we specify $\alpha \sim Ga(a_\alpha, b_\alpha)$ and $\gamma \sim Ga(a_\gamma, b_\gamma)$. Introducing auxiliary variables $u = \{u_{ih} : i = 1, \ldots, n; \quad h = 0, 1, \ldots, h^*\}$ to avoid truncated approximations, the complete data joint likelihood of y, u and z is:

$$\prod_{i=1}^{n} g(Y_i; A, \theta_{\psi_i}, \Sigma)\mathbf{1}(u_{i0} < \pi_{\psi_{i0}}) \prod_{h=1}^{h^*} \mathbf{1}(u_{ih} < \pi_{\psi_{ih}})\nu_h^{z_{ih}}(1-\nu_h)^{1-z_{ih}},$$

where Y_i is a $\sum_{m=1}^{M} n_{m_i} \times 1$ outcome vector for the M measures of the i-th subject, $g(\cdot)$ is the density function of the outcome vector Y_i, $i = 1, \ldots, n$, and the elements in u are constrained to the interval $(0, 1)$.

Starting from the initiation step, the Gibbs sampler at the truncated level h^* proceeds as follows:

1. Update the hth column of the factor loadings matrix A_{h^*}, denoted by A_h, from its conditional distribution

$$p(A_h|-) \sim N((C_h^T\Sigma_N^{-1}C_h + \Sigma_{A_h}^{-1})^{-1}C_h^T\Sigma_N^{-1}Y_h, (C_h^T\Sigma_N^{-1}C_h + \Sigma_{A_h}^{-1})^{-1}),$$

where $Y_h = (Y_{1h}^T, \ldots, Y_{nh}^T)^T$ with $Y_{ih} = Y_i - \sum_{j\neq h} I_M \otimes B_i A_j \theta_{\psi_{ij}}$, $C_h = \begin{pmatrix} \theta_{\psi_{1h}}(I_M \otimes B_1) \\ \vdots \\ \theta_{\psi_{nh}}(I_M \otimes B_n) \end{pmatrix}$, $\Sigma_N = I_n \otimes \Sigma$, $\Sigma_{A_h} = \mathrm{diag}(\phi_{1,h}^{-1}\tau_h^{-1}, \ldots, \phi_{(K+1)pM,h}^{-1}\tau_h^{-1})$, for $h = 1, \ldots, h^*$.

2. Update ϕ_{gh} from its conditional distribution

$$p(\phi_{gh}|-) \sim \mathrm{Ga}(\frac{v+1}{2}, \frac{v + A_{gh}^2\tau_h}{2}).$$

3. Update δ_1 from its conditional distribution

$$p(\delta_1|-) \sim \mathrm{Ga}(a_1 + \frac{1}{2}(K+1)pMh^*, 1 + \frac{1}{2}\sum_{h=1}^{h^*}\tau_h^{(1)}\sum_{g=1}^{(K+1)pM}\phi_{gh}A_{gh}^2),$$

and update δ_l, $l \geq 2$ from its conditional distribution

$$p(\delta_l|-) \sim \mathrm{Ga}(a_2 + \frac{1}{2}(K+1)pM(h^* - l + 1), 1 + \frac{1}{2}\sum_{h=l}^{h^*}\tau_h^{(l)}\sum_{g=1}^{(K+1)pM}\phi_{gh}A_{gh}^2),$$

where $\tau_h^{(l)} = \prod_{t=1,h\neq l}^{h}\delta_t$, $l = 1, \ldots, h^*$.

4. Update σ_f^{-2}, $f = 1, \ldots, MT$, from its conditional distribution

$$p(\sigma_f^{-2}|-) \sim \text{Ga}(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\sum_{i=1}^{n}(y_{if} - (I_M \otimes B_i)_f A\theta_{\psi_i})^2),$$

where $(I_M \otimes B_i)_f$ denotes the fth row of $(I_M \otimes B_i)$.

5. Update u_{ih} from its conditional distribution

$$p(u_{ih}|-) \sim \text{Unif}(0, \pi_{\psi_{ih}}).$$

6. Update the latent z_{ih} from its conditional distribution

$$p(z_{ih}|-) \sim \text{Ber}(\frac{\nu_h N(y_i; A, \theta_{i(z_{ih}=1)}, \Sigma)}{\nu_h N(y_i; A, \theta_{i(z_{ih}=1)}, \Sigma) + (1-\nu_h)N(y_i; A, \theta_{i(z_{ih}=0)}, \Sigma)}),$$

where $\theta_{i(z_{ih}=j)}$ refers to the current value of θ_i with inserting $\theta_{0\phi_{i0}h}$ to the hth component for $j = 1$, and $\theta_{1\phi_{ih}h}$ for $j = 0$.

7. Update the stick-breaking weights π_q^* from its conditional distribution

$$p(\pi_q^*|-) \sim \text{Beta}(\sum_{h=0}^{h^*}\sum_{i=1}^{n}\mathbf{1}(\psi_{ih} = q) + 1, \sum_{h=0}^{h^*}\sum_{i=1}^{n}\mathbf{1}(\psi_{ih} > q) + \alpha),$$

for $q \leq \psi^*$ with $\psi^* = \max\{\psi_{ih} : i = 1, \ldots, n, h = 0, 1, \ldots, h^*\}$, for $q > \psi^*$, sample π_q^* from $p(\pi_q^*|-) \sim \text{Beta}(1, \alpha)$.

8. Update the allocation index ψ_{ih} from its conditional distribution

$$P(\psi_{ih} = q|-) \propto \mathbf{1}(q \in C_\pi(u_{ih}))N(y_i; A, \theta_{\psi_{ih}=q}, \Sigma),$$

where $C_\pi(u_{ih}) = \{q : \pi_q > u_{ih}\} \subset \{1, 2, \ldots, \infty\}$, and π_q^* is sampled for $q =$

$1, \ldots, \tilde{\psi}$ such that

$$\sum_{q=1}^{\tilde{\psi}} \pi_q^* \prod_{l \leq q} (1 - \pi_l^*) \geq 1 - u^*,$$

where $u^* = \min\{u_{ih} : i = 1, \ldots, n, h = 0, 1, \ldots, h^*\}$.

9. Update θ_{jq} from its conditional distribution

$$p(\theta_{jq}|-) \sim N_{h^*}([\text{diag}(\mathbf{1}_{h^*}\zeta_{jq}) + W'_{jq}\Sigma_N^{-1}W_{jq}]^{-1}W'_{jq}\Sigma_N^{-1}Y_{jq}, [\text{diag}(\mathbf{1}_{h^*}\zeta_{jq}) + W'_{jq}\Sigma_N^{-1}W_{jq}]^{-1}),$$

in which Y_{jq} refers to the contribution for θ_{jq} from subjects with $z_{ih} = 1 - j$ and $\phi_{ih} = q$. In addition, $W_{jq} = \text{diag}(\text{w}_1, \ldots, \text{w}_n)Z\Phi R$, and $Y_{jq} = \text{diag}(y_1, \ldots, y_n)Z_Y\Phi_Y\mathbf{1}_n$, where

- $\text{w}_i = I_M \otimes B_i A$,
- $Z = \text{Blkdiag}(Z_1, \ldots, Z_n)$ with $Z_i = \text{diag}(1(z_{i1} = 1 - j), \ldots, 1(z_{ih^*} = 1 - j))$,
- $\Phi = \text{Blkdiag}(\Phi_1, \ldots, \Phi_n)$ with $\Phi_i = \text{diag}(1(\psi_{i1} = q), \ldots, 1(\psi_{ih^*} = q))$,
- $R = \mathbf{1}_n \otimes I_{h^*}$,
- $Z_Y = \text{Blkdiag}(Z_{Y_1}, \ldots, Z_{Y_n})$ with $Z_{Y_i} = 1(\sum_{h=1}^{h^*} 1(z_{ih} = 1 - j) \geq 1)$,
- $\Phi_Y = \text{Blkdiag}(\Phi_{Y_1}, \ldots, \Phi_{Y_n})$ with $\Phi_{Y_i} = 1(\sum_{h=1}^{h^*} 1(\psi_{ih} = q) \geq 1)$,
- and $\mathbf{1}_n$ is a $n \times 1$ vector of entries of 1.

10. Update ζ_{jq} from its conditional distribution $\text{Ga}(0.5 + h^*/2, 0.5 + \theta'_{jq}\theta_{jq}/2)$.

11. Update the hyperparameter γ from its conditional distribution $\text{Ga}(a_\gamma + h^*, b_\gamma - \sum_{h=1}^{h^*} \log(1 - \nu_h))$.

12. Update the hyperparameter α from its conditional distribution $\text{Ga}(a_\alpha + \psi^*, b_\alpha - \sum_{q=1}^{\psi^*} \log(1 - \pi_q^*))$.

To choose the effective number of factors, we use an adaptation approach (Bhattacharya and Dunson, 2011) to update the truncated number of factors h^* across iterations. Starting with a conservative guess $\tilde{h^*}$ of h^*, we adapt with probability $p(t) = \exp(\alpha_0 + \alpha_1 t)$ at the tth iteration. We specify α_0 and α_1 such that the chain comes with adaptation around every 10 iterations at the beginning and exponentially fast decreasing in adaptation frequency. In detail, at the tth iteration, we monitor the columns in the factor loadings by comparing u_t with $p(t)$, where u_t is a sequence randomly generated from Unif$(0, 1)$. If $u_t \leq p(t)$, we remove the columns having all elements within some pre-specified small zero neighborhood, called redundant columns, and retain non-redundant columns for other parameters. If no such redundant column is detected, we add a new column to the loadings, and sample other parameters from their prior distributions to fill in the additional columns.

Table 3.1: Simulation study: FDR and power of the local hypothesis tests for the effects of gestational age on FA and MD along the right internal capsule tract, evaluated at five different values of c and three different values of ϵ for sample sizes of 32 and 64 subjects.

		$\epsilon = 0.1$		$\epsilon = 0.15$		$\epsilon = 0.2$	
n	c	FDR	Power	FDR	Power	FDR	Power
32	0	1.00	-	1.00	-	1.00	-
	0.2	0.01	0.52	0.01	0.45	0.01	0.40
	0.4	0.01	0.88	0.01	0.84	0.01	0.80
	0.6	0.01	0.98	0.01	0.96	0.01	0.95
	0.8	0.01	0.99	0.01	0.99	0.01	0.99
64	0	1.00	-	1.00	-	1.00	-
	0.2	0.01	0.74	0.01	0.68	0.01	0.64
	0.4	0.01	0.98	0.01	0.96	0.01	0.95
	0.6	0.01	0.99	0.01	0.99	0.01	0.99
	0.8	0.01	1.00	0.01	1.00	0.01	1.00

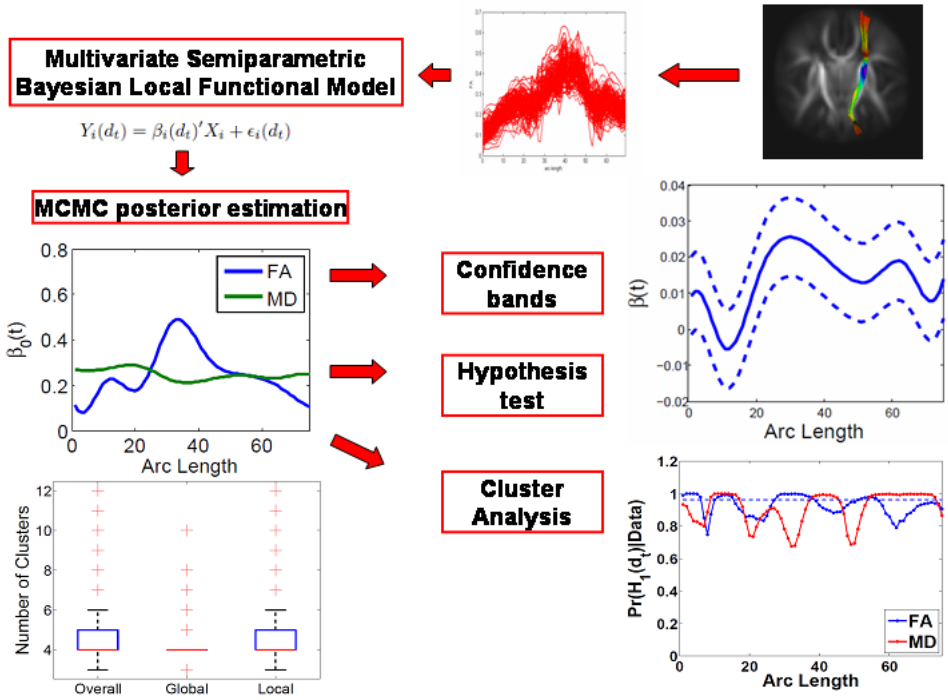

Fig. 3.1: A schematic overview of SBLFM: a multivariate semiparametric varying-coefficient model for multiple diffusion properties, a MCMC posterior estimation method for estimating the coefficient functions, a construction of Bayesian confidence bands of the mean covariate effect curves, a Bayesian local hypothesis testing procedure, and a posterior cluster analysis procedure.

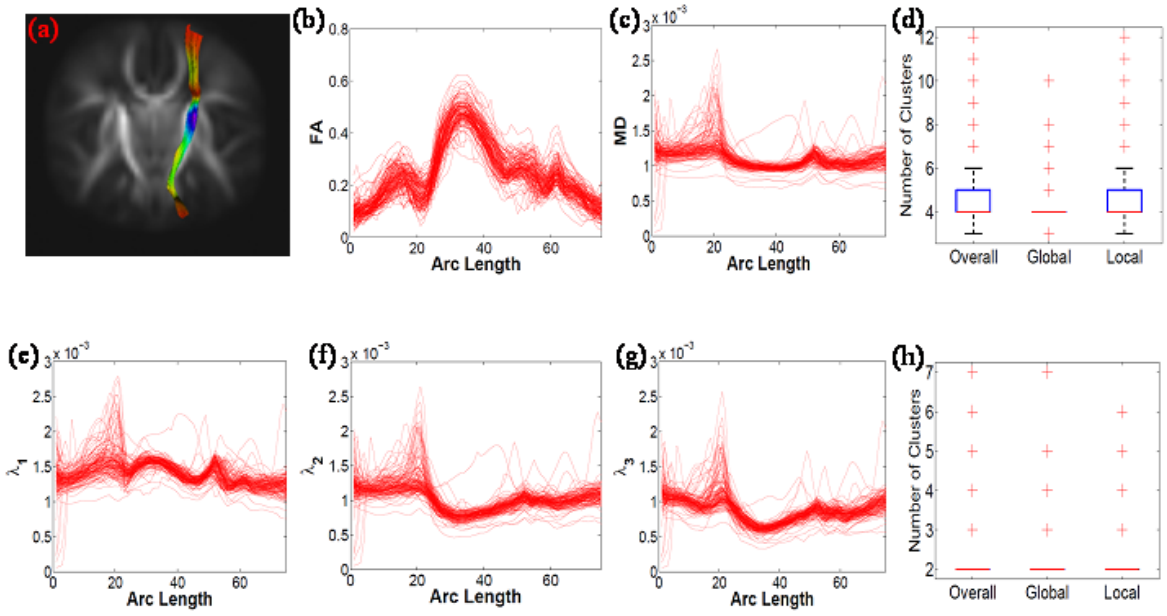

Fig. 3.2: Right internal capsule tract description: (a) the right internal capsule tract extracted from the tensor atlas with color presenting mean FA value; diffusion properties of FA in panel (b), MD in panel (c), λ_1 in panel (e), λ_2 in panel (f), λ_3 in panel (g); the number of clusters posteriori in FA and MD trajectories in panel (d) and the number of clusters posteriori in λ_1, λ_2 and λ_3 trajectories in panel (h).

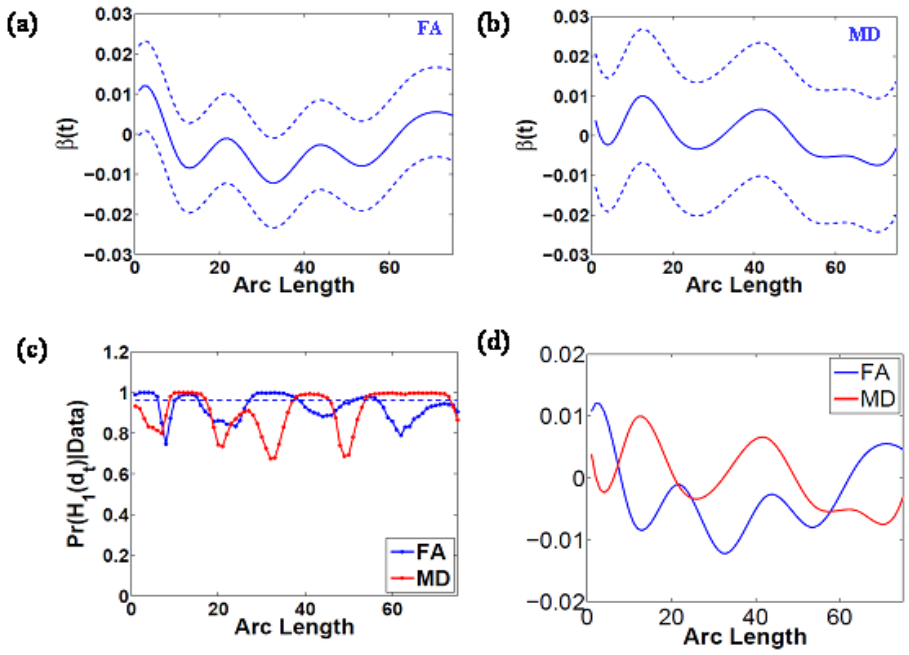

Fig. 3.3: Results of the gender effects on FA and MD along the right internal capsule tract: the estimated coefficient curves (–) and the corresponding 95% confidence bands (- -) for FA in panel (a) and MD in panel (b); (c) the posterior probability curves in favor of the grid point-wise alternative hypotheses; (d) the estimated mean FA and MD.

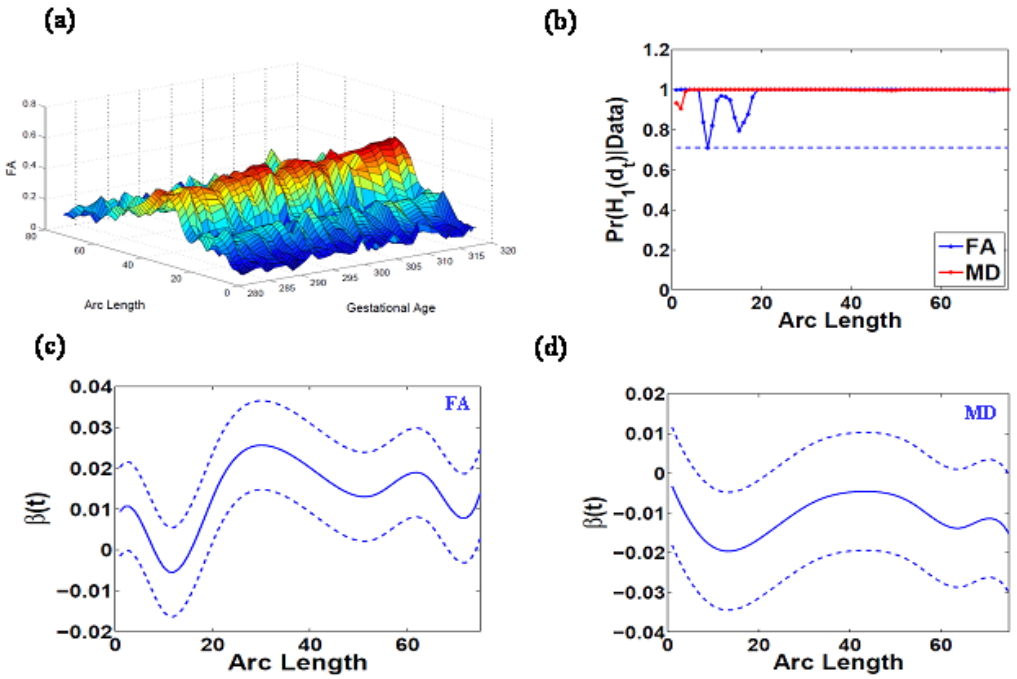

Fig. 3.4: Results of the gestational age effects on FA and MD along the right internal capsule tract: (a) 3D surf plot of FA along the right internal capsule tract for visualizing the gestational age effects ; (b) the posterior probability curves in favor of the grid point-wise alternative hypotheses; the estimated coefficient curves (–) and the corresponding 95% confidence bands (- -) for FA in panel (c) and MD in panel (d).

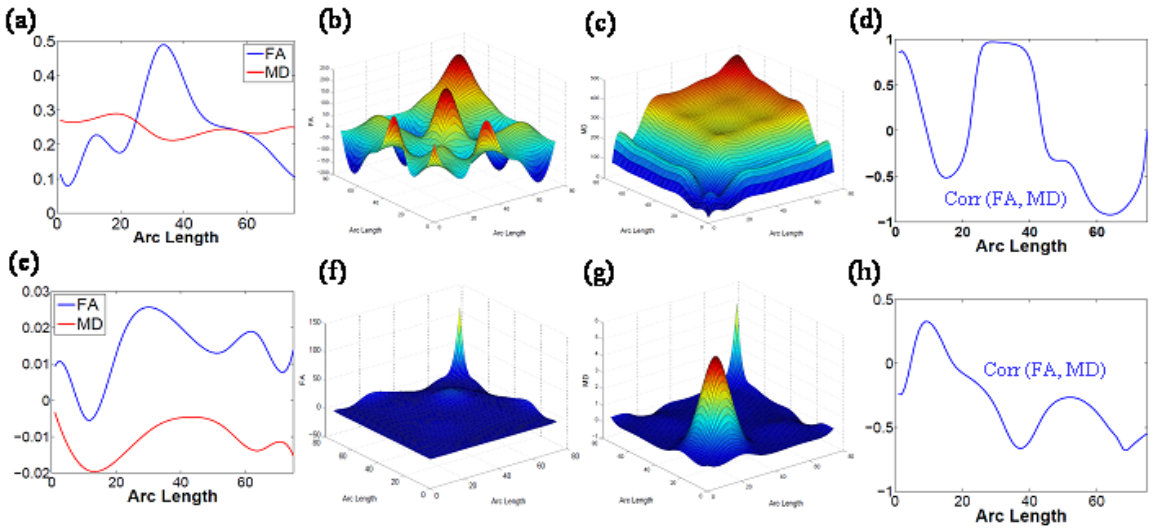

Fig. 3.5: Results from the analysis of FA and MD along the right internal capsule tract: (a) the posterior estimated curves of FA and MD for males of average gestational age; the estimated covariance matrices for FA in panel (b) and MD in panel (c) for males of average gestational age; (d) the estimated correlations between FA and MD for males of average gestational age; (e) the posterior estimated gestational age effects; the estimated covariance matrices for the gestational age effects on FA in panel (f) and MD in panel (g); (h) the estimated correlations between the gestational age effects on FA and MD.

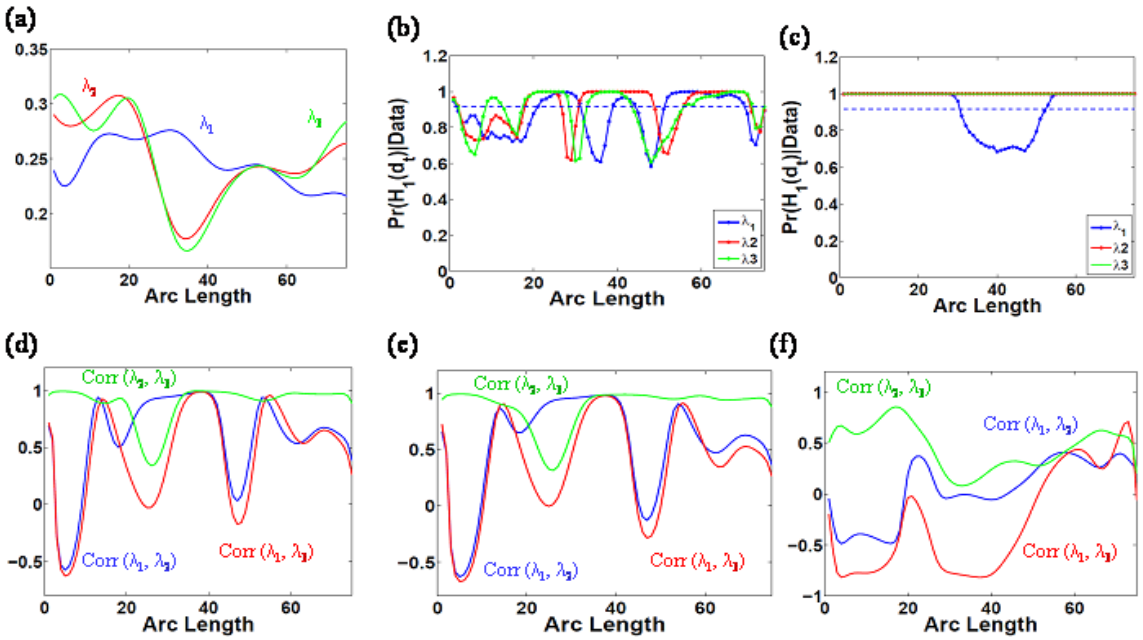

Fig. 3.6: Results from the analysis of the three eigenvalues of diffusion tensor on the right internal capsule tract: (a) the estimated mean functions for λ_1 (blue), λ_2 (red) and λ_3 (green); the posterior probability curves in favor of the grid point-wise alternative hypotheses for the gender effects in panel (b) and the gestational age effects in panel (c); the estimated correlations among λ_1, λ_2 and λ_3 for males of mean gestational age in panel (d) and females of mean gestational age in panel (e), and among the gestational age effects on λ_1, λ_2 and λ_3 in panel (f).

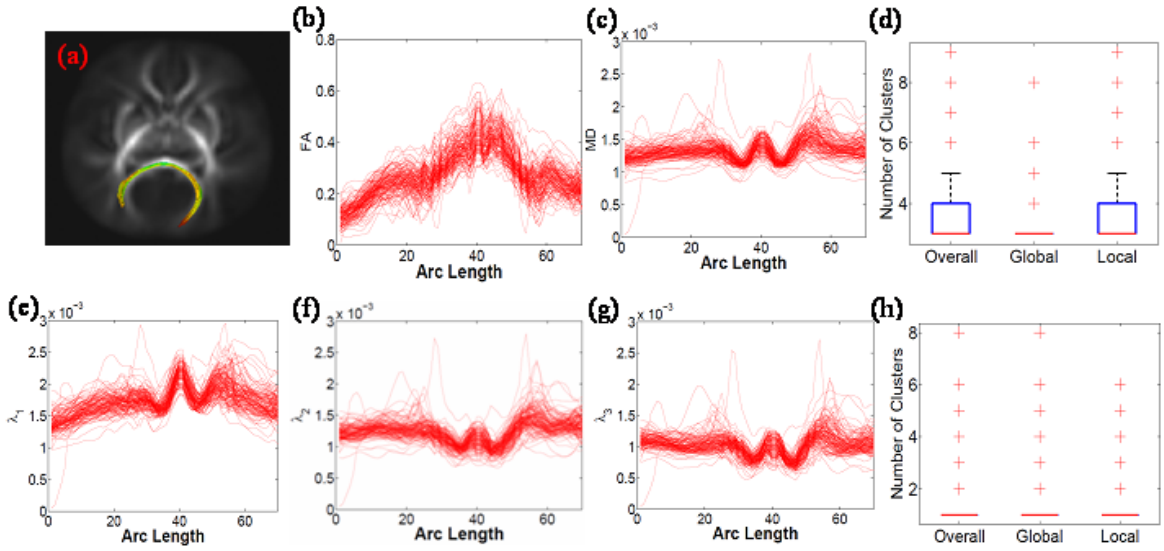

Fig. 3.7: Splenium tract description: (a) the splenium capsule tract extracted from the tensor atlas with color presenting mean FA value; diffusion properties of FA in panel (b), MD in panel (c), λ_1 in panel (e), λ_2 in panel (f), λ_3 in panel (g); the number of clusters *posteriori* in FA and MD trajectories in panel (d) and the number of clusters *posteriori* in λ_1, λ_2 and λ_3 trajectories in panel (h).

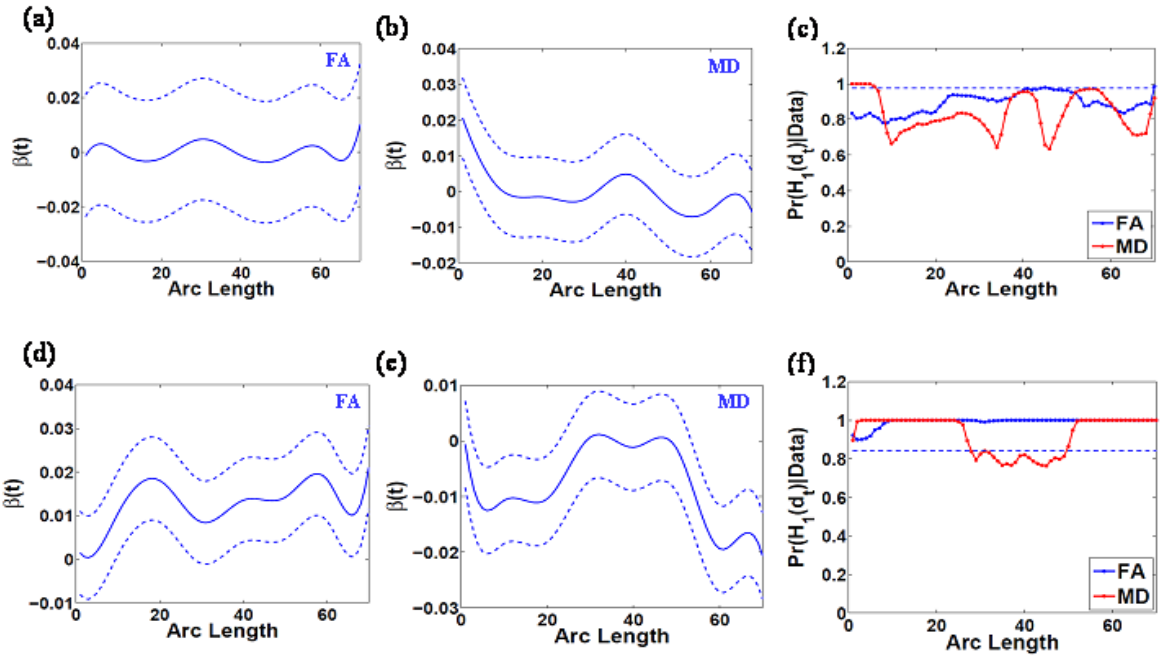

Fig. 3.8: Results from the analysis of FA and MD along the splenium tract: the estimated coefficient curves (–) and the corresponding 95% confidence bands (- -) for the gender effects (panel (a) for FA and panel (b) for MD) and the gestational age effects (panel (d) for FA and panel (e) for MD); the posterior probability curves in favor of the grid point-wise alternative hypotheses for the gender effects (panel (c)) and the gestational age effects (panel (f)).

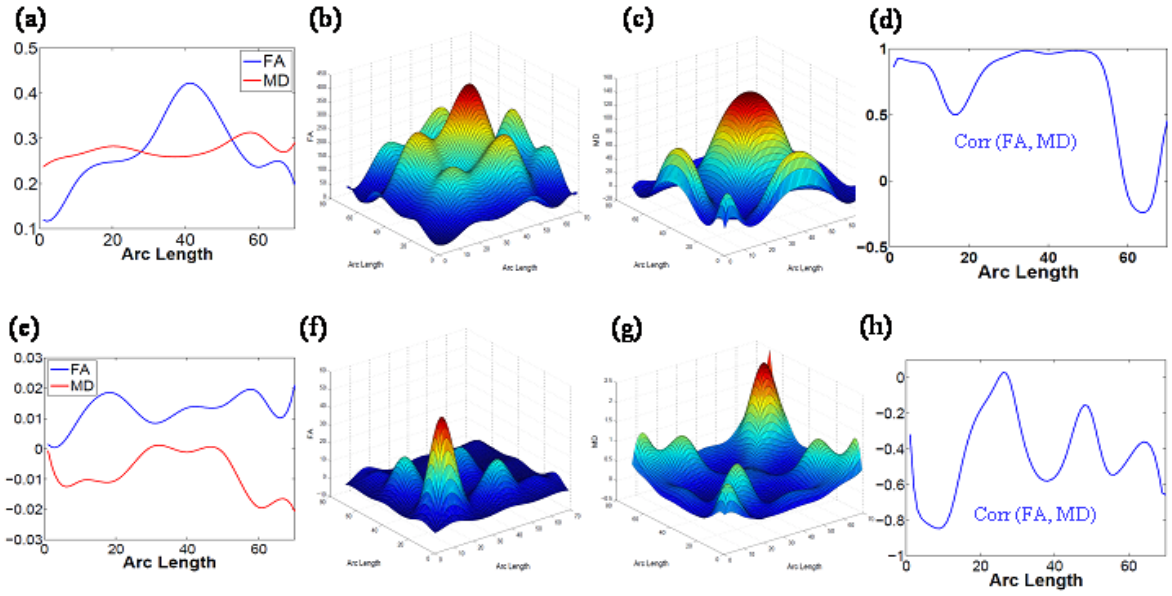

Fig. 3.9: Results from the analysis of FA and MD along the splenium tract: (a) the posterior estimated curves of FA and MD for males of average gestational age; the estimated covariance matrices for FA in panel (b) and MD in panel (c) for males of average gestational age; (d) the estimated correlations between FA and MD for males of average gestational age; (e) the posterior estimated gestational age effects; the estimated covariance matrices for the gestational age effects on FA in panel (f) and MD in panel (g); (h) the estimated correlations between the gestational age effects on FA and MD.

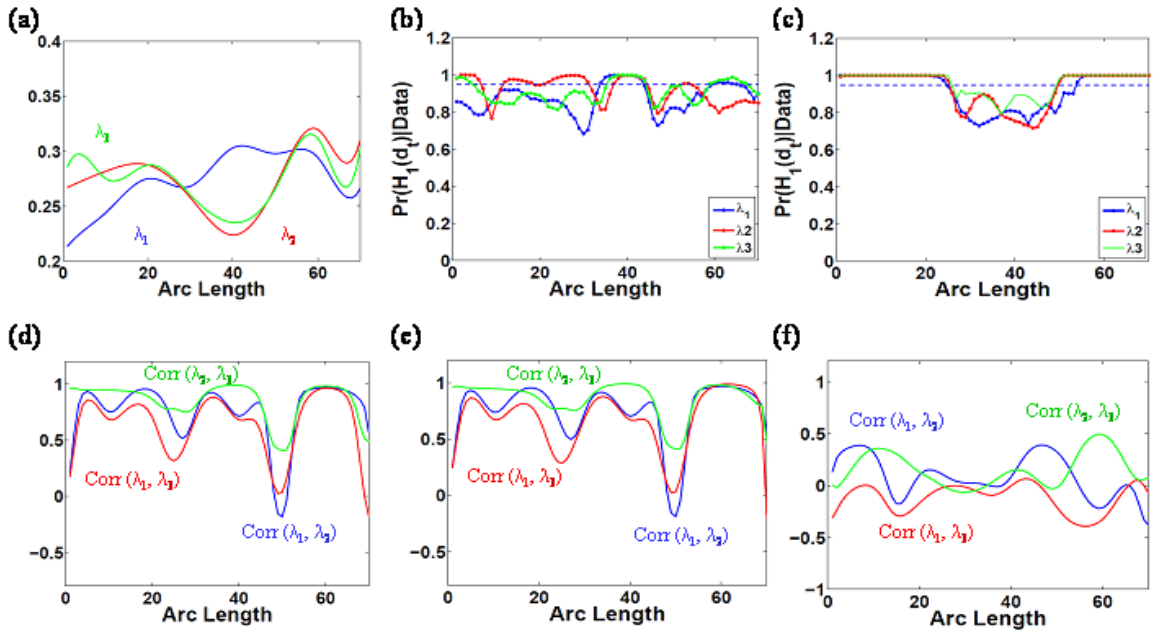

Fig. 3.10: Results from the analysis of the three eigenvalues of diffusion tensor on the splenium tract: (a) the estimated mean functions for λ_1 (blue), λ_2 (red) and λ_3 (green); the posterior probability curves in favor of the grid point-wise alternative hypotheses for the gender effects in panel (b) and the gestational age effects in panel (c); the estimated correlations among λ_1, λ_2 and λ_3 for males of mean gestational age in panel (d) and females of mean gestational age in panel (e), and among the gestational age effects on λ_1, λ_2 and λ_3 in panel (f).

Chapter 4

Bayesian Sticky HDP Infinite Hidden Markov Circumplex Model

4.1 Introduction

The circumplex model is useful in the study of affects, which refer to the experience of feeling or emotion (Huitt, 2003). Affective states are depicted by a circular structure based on the dimensions of valence and arousal (Russell and Carroll, 1999), as shown in Fig (4.1), where valence refers to the intrinsic attractiveness (positive valence) or aversiveness (negative valence) (Frijda, 1986), and arousal refers to being awake or reactive to stimuli (Csikszentmihalyi, 1998; Mirr, 2001). Each emotion arises as a linear combination of these two dimensions, or as varying degrees of both valence and arousal. For example, Joy is conceptualized as an affect produced by strong activation in the neural systems associated with positive valence or pleasure and moderate activation in the neural systems associated with arousal. Similarly, other affects are the products of the activities of these two neurophysiological systems in which the degrees of activation differ. Hence, specific emotions arise from cognitive interpretations and labeling of these activation patterns of these two neurophysiological systems.

Within this circumplex model, affects having the same valence tend to have substantial positive correlation, while affects of the opposite valence tend to have weak negative correlation. This phenomenon is referred to as "two fundamental psychometric principles" in Watson and Clark (1997). These two principles underlie their conceptual theory of the independence between "Positive Affect" and "Negative Affect" (Watson and Tellegen, 1985) and have been tested in the scales of Positive and Negative Affect Schedule (PANAS) for the assessment of affects (Watson et al., 1988). The correlation between two affects is modeled as a function of the angle between them in the circular ordering.

Circumplex model has been studied in the literature to model the correlation structure among affects. Correlations between adjacent scales with the circular ordering are the highest. As circular separation increases, correlations decrease. Hence, the corresponding correlation matrix has a circumplex structure, such that correlations decrease first and then increase when one moves away from the main diagonal. Under the constraint of the circumplex structure, Guttman (1954) and Anderson (1960) proposed stochastic processes on the circumference of the circle which produce positive correlations with moving average and Markov properties respectively. Methods allowing negative correlations are developed in Cudeck (1986) and Wiggins et al. (1981). Browne (1992) extended the model of Anderson (1960) to allow for negative correlations. Lenk et al. (2006) extended the work of Browne (1992) from Bayesian perspective and accommodated subject-specific random effects to account for personal response scale usage and subject-specific factors.

Circumplex model is a type of multidimensional scaling (MDS) methods, which is a family of models that map a group of objects into a low dimensional space where points representing the objects and the distances between the points corresponding to the similarity/dissimlarity measure among the objects. It aims to seek a configuration

of points such that the distances "match" the similarities as well as possible. MDS methods have been applied into a wide variety of disciplines, including behavior science, social science, political science, marketing research, biomedical science, and so on. It sprung from psychometric analysis to evaluate the similarity of people's judgements over a set of objects with the first appearance (Torgerson, 1952). During the recent decade of the bursting of high-dimensional large data sets, MDS was applied to gene analysis like revealing the relational pattern of gene-expressions (Taguchi and Oono, 2005). In the frontier of the biotechnological field, MDS was applied for protein structure prediction (Mooney et al., 2006) and the quality assessment of mass spectrometry data (Harezlak et al., 2007).

This chapter is motivated by a data set from a clinical study which is designed to understand the mechanism of emotion regulation based on empirical measurements. Participants were measured on a 17-scale PANAS measures in a daily diary study for a total of 7 days. Each subject reports their affect scales 5 times each day, with 4 times at day and 1 time at night. The resultant data structure corresponds to the traditional cross-sectional circumplex data while measured repeatedly over time. Hence, we call such kind of data as longitudinal circumplex data. A major interest to analyze this data set is to discover the latent emotion states hidden in the emotion items of the PANAS ratings. An appropriate circumplex model for the analysis of such longitudinal circumplex data has not been developed in the literature. We are motivated to bridge the gap between the statistical methods of emotion regulation and the existing longitudinal circumplex data.

The analysis of the longitudinal circumplex data should consider the temporal structure of emotions. In the past few decades, researchers tended to conceptualize emotion as a process. Emotion is regarded as versatile, "dynamic", and often shows persistent patterns of self-regularity (Carver and Scheier, 1982; Chow et al., 2004; Larsen, 2000).

In this chapter, we target to extend the work of Lenk et al. (2006) to a semiparametric Bayesian state-space circumplex model for the analysis of longitudinal circumplex data. Assuming an infinite many collection of emotion states hidden in the population, a hierarchical Dirichlet proces (Teh et al., 2006) is used to facilitate the discovery of the time related state-switching structure. In addition, a sticky component (Fox et al., 2008) is added to address the phenomenon of emotion persistence.

Section 4.2 introduces the model and associated inferences motivated by the above considerations. Section 4.3 introduces an MCMC algorithm for posterior computation. Section 4.4 is for an illustrative simulation. Section 4.5 applies the model to an emotion data set. Section 4.6 makes some further discussion.

4.2 Bayesian Sticky HDP-iHMM Circumplex Model

4.2.1 Circumplex Model

Circumplex model is a useful tool to quantify the correlation structure among affective states where the underlying structure of affective states can be characterized on the circumference of a circle. Consider a latent response score $\mathrm{Y}_{i,j}$ of the ith subject to the jth item representing an affective state, $i = 1, \ldots, n$, $j = 1, \ldots, J$. Lenk et al. (2006) proposed a random effects circumplex model for the latent response variable $\mathrm{Y}_{i,j}$:

$$\mathrm{Y}_{i,j} = \mathrm{U}_j + \mathrm{H}_i + \mathrm{A}_i \sin\theta_j + \mathrm{B}_i \cos\theta_j + \epsilon_{i,j}, \;\; \theta_1 = 0, \; \theta_j \in [0, 2\pi], \tag{4.1}$$

where U_j is a scalar for the mean latent response for the jth item, H_i is a scalar for the ith subject specific random effect to capture scale-usage effects, A_i and B_i are scalars specifying the random factor scores for the ith subject, and $\sin\theta_j$ and $\cos\theta_j$ are the jth item specific loadings, which are constrained to the unit circle. The error term $\epsilon_{i,j}$, $i = 1, \ldots, n$, $j = 1, \ldots, J$ are independently and normally distributed with mean 0 and

variance σ_j^2 specific to the jth item.

The subject by item interaction term $\mathrm{A}_i \sin\theta_j + \mathrm{B}_i \cos\theta_j$ characterizes the circumplex correlation. Specifying $\mathrm{H}_i \sim N(0, \lambda^2)$, $\mathrm{A}_i \sim N(0, \tau_a)$, $\mathrm{B}_i \sim N(0, \tau_b)$, and letting $\tau_a = \tau_b = \tau$ makes the inter-item correlations to have a circumplex structure, after adjusting for the scale-usage effects. Conditional on the angles, the variance of $\mathrm{Y}_{i,j}$ and the covariance between $\mathrm{Y}_{i,j}$ and $\mathrm{Y}_{i,k}$ are given by

$$\begin{aligned} \mathrm{Var}(\mathrm{Y}_{i,j}|\theta_j) &= \lambda^2 + \tau^2 + \sigma^2 \\ \mathrm{Cov}(\mathrm{Y}_{i,j}, \mathrm{Y}_{i,k}|\theta_j, \theta_k) &= \lambda^2 + \tau^2 \cos(\theta_j - \theta_k). \end{aligned}$$

4.2.2 Sticky HDP-iHMM State-space Circumplex Model

Considering our motivated data, suppose $i = 1, \ldots, n$ subjects respond to $j = 1, \ldots, J$ items on a rating scale with G ordinal categories over $t = 1, \ldots, T$ time points. Let $\mathrm{W}_{i,j,t}$ denote the rating of the ith subject toward the jth item at time t. We introduce a set of the ith subject specific ordered latent cut points $\{c_{i,g}\}$: $c_{i,0} = -\infty, c_{i,1} = -1 < c_{i,2} < \ldots < c_{i,G-2} < c_{i,G-1} = 1, c_{i,G} = \infty$. We assume that the observed ordinal response $\mathrm{W}_{i,j,t}$ is associated with a latent response variable $\mathrm{Y}_{i,j,t}$ falling between two latent cutpoints:

$$\mathrm{W}_{i,j,t} = g, \quad \text{iff} \quad c_{i,g-1} \leq \mathrm{Y}_{i,j,t} < c_{i,g}, \quad \text{for} \quad g = 1, \ldots, G.$$

We extend model (4.1) to accommodate the dynamics of the latent response over time by modeling the latent responses $\mathrm{Y}_{i,j,t}$:

$$\mathrm{Y}_{i,j,t} = \mathrm{U}_j + \mathrm{H}_{i,t} + \mathrm{A}_{i,t} \sin\theta_j + \mathrm{B}_{i,t} \cos\theta_j + \epsilon_{i,j,t}, \;\; \theta_1 = 0, \; \theta_j \in [0, 2\pi]. \tag{4.2}$$

where U_j is the mean latent response of the jth item, $\mathrm{H}_{i,t}$, $\mathrm{A}_{i,t}$ and $\mathrm{B}_{i,t}$ are the ith subject specific latent random effects vector at time t, and $\sin\theta_j$ and $\cos\theta_j$ are the jth item specific loadings. The error terms $\epsilon_{i,j,t}$ are assumed to be independently drawn from $N(0, \sigma_j^2)$.

We introduce the sticky infinite state hidden Markov model via hierarchical Dirichlet proces (HDP-iHMM) (Fox et al., 2008) for the subject specific latent random factor scores $\phi_{i,t} = (\mathrm{H}_{i,t}, \mathrm{A}_{i,t}, \mathrm{B}_{i,t})^T$:

$$
\begin{aligned}
&\phi_{i,t}|s_{i,t}, \{\phi_k^*\}_{k=1}^{\infty} = \phi_{s_{i,t}}^* \\
&\{\phi_k^*\}_{k=1}^{\infty}|P_0 \sim P_0, \quad P_0 : \phi_k^* \sim N(0, \mathrm{diag}(\lambda^2, \tau^2 1_2)) \\
&s_{i,t}|s_{i,t-1}, \{\boldsymbol{\omega}_e\}_{e=1}^{\infty} \sim \mathrm{Mult}(\boldsymbol{\omega}_{s_{i,t-1}}) \\
&\{\boldsymbol{\omega}_e\}_{e=1}^{\infty}|\alpha, \kappa, \boldsymbol{\beta} \sim \mathrm{DP}(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\delta_e}{\alpha + \kappa}) \\
&\boldsymbol{\beta}|\gamma \sim \mathrm{Stick}(\gamma),
\end{aligned} \tag{4.3}
$$

where $s_{i,t}$ denotes the state of the ith subject at time t, the infinite collection of state specific parameters $\{\phi_k^*\}_{k=1}^{\infty}$ are drawn from a 3-dimensional normal base measure P_0, $\boldsymbol{\omega}_e$ corresponds to the state e specific transition distribution, $\kappa > 0$ is the self-transition parameter for the current state e, $\boldsymbol{\beta}$ is the average transition distribution. The average transition distribution $\boldsymbol{\beta}$ is formulated by a stick-breaking process of parameter γ. The state specific transition distribution $\boldsymbol{\omega}_e$ is generated from the Dirichlet process with central transition distribution $\boldsymbol{\beta}$ as the base measure and a precision parameter α controlling the degree of deviation. The current state $s_{i,t}$ depends on the previous state $s_{i,t-1}$ via the multinomial distribution of transition parameters $\boldsymbol{\omega}_{s_{i,t-1}}$.

The Bayesian specification is completed by assigning priors on other parameters in the model (4.2). We place the multivariate normal prior on the item mean responses vector $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_J)^T$, the conditionally uniform prior on the latent cutpoints c_i,

the gamma priors on σ_j^{-2}, λ^{-2}, and τ^{-2}, and the extended Von Mises priors on the angles θ_j:

$$\begin{aligned}
\mathbf{U} &\sim N(\mu_0, \Sigma_0), \\
c_i &\sim \mathcal{X}(-1 < c_{i,2} < \ldots < c_{i,G-2} < 1), \\
c_{i,g}|c_{i,g-1}, c_{i,g+1} &\sim \text{Unif}(c_{i,g-1}, c_{i,g+1}), \quad g = 2, \ldots, G-2, \\
\sigma_j^{-2} &\sim \text{Ga}(r_0, s_0), \quad \lambda^{-2} \sim \text{Ga}(u_{0,1}, v_{0,1}), \quad \tau^{-2} \sim \text{Ga}(u_{0,2}, v_{0,2}), \\
\theta_j &\sim \begin{cases} \text{VM}(d_0, Q_0, [0, \pi]), & \text{for } j = 2, \\ \text{VM}(d_0, Q_0, [0, 2\pi]), & \text{for } j > 2. \end{cases}
\end{aligned} \tag{4.4}$$

where $d_0 = (0, 1)^T$, $Q_0 = 0.2I_2$. This Von Mises distribution is fairly flat on $[0, 2\pi]$. The probability density function of the extended Von Mises distribution $\text{VM}(d, Q, \mathcal{C})$ is given by $p(\theta|d, Q, \mathcal{C}) \propto \exp\{-0.5(\xi(\theta) - d))^T Q(\xi(\theta) - d)\}\mathcal{X}(\theta \in \mathcal{C})$, where $\xi(\theta) = (\sin(\theta), \cos(\theta))^T$, d is a two-dimensional vector, Q is a 2×2 matrix, $\mathcal{X}(\cdot)$ is the indicator function, and $\mathcal{C}$ is a subset of $[0, 2\pi]$.

4.2.3 Background: Dirichlet Processes

Dirichlet Process

A Dirichlet process (DP) (Ferguson, 1973, 1974) defines a distribution for a random density function $G(\phi)$ on a parameter space Φ, denoted by $G(\phi) \sim \text{DP}(\gamma, P_0)$. The stick-breaking representation (Sethuraman, 1994) expresses $G(\phi)$ as

$$\begin{aligned}
G(\phi) &= \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k^*}(\cdot), \quad \phi_k^* \overset{i.i.d}{\sim} P_0 \\
\beta_k &= \beta_k^* \prod_{l<k} (1 - \beta_l^*), \quad \beta_k^* \overset{i.i.d}{\sim} \text{Beta}(1, \gamma), \quad k = 1, \ldots, \infty,
\end{aligned} \tag{4.5}$$

where β_k is a probability weight formulated from a stick-breaking procedure, $\delta_{\phi^*}(\cdot)$ denotes a point mass at ϕ^* which is sampled from P_0, P_0 is a central probability measure as a prior guess, and γ is a precision parameter expressing confidence in the prior guess. Under this formulation, the probability of ϕ allocated to $\delta_{\phi_k^*}(\cdot)$ is equivalent to β_k. We denote the distribution on the set of weights $\boldsymbol{\beta} = \{\beta_k\}_{k=1}^{\infty}$ by $\boldsymbol{\beta} \sim \text{Stick}(\gamma)$.

The discreteness of G implied from the definition of the DP process generates a clustering structure in the samples drawn from G. Suppose $\phi_i \sim G$, $i = 1, \ldots, n$. Integrating over G, the predictive distribution of ϕ_{n+1} is obtained from the Pólya urn representation (Blackwel and Macqueen, 1973)

$$\phi_{n+1}|\boldsymbol{\phi}^n, \gamma, P_0 \sim \sum_{k=1}^{K} \frac{n_k}{\gamma + n}\delta_{\phi_k^*} + \frac{\gamma}{\gamma + n}P_0, \tag{4.6}$$

where $\{\phi_k^*\}_{k=1}^K$ denote the distinct values (clusters) taken by the set $\{\phi_i\}_{i=1}^n$, n_k is the number of ϕ_i's taking the value of ϕ_k^*, and the hyperparameter γ determines the weight for a new cluster to be created.

The above Pólya urn sampling scheme is closely related to the Chinese restaurant process. Let ϕ_i as a customer entering a restaurant with infinitely many tables, and each table serving a unique dish ϕ_k^*. Every arriving customer chooses a already occupied table with a probability in proportion to the number of customers currenly sitting at this table, and a new table with a probability in proportion to γ. It is implied that small γ favors fewer tables to be selected.

The DP is commonly used as a nonparametric prior for the distribution of the parameters in a mixture model of unknown complexity. Assume we have an observation $y_i \sim F(\phi_i)$ with $\phi_i \sim G(\phi)$. The above sample scheme generates an indication variable z_i to denote the mixture component for the observation $y_i \sim F(\phi_{z_i}^*)$. Unlike standard mixture models having fixed number of mixture components, the DP prior allows an

infinite number of mixture components and lets the data inform the actual number of components.

Hierarchical Dirichlet Process

Teh et al. (2006) extends the DP to the hierarchical Dirichlet Process (HDP) for cases where groups of mixture model generative data are related but with an unknown sharing structure. Assume J groups of data $y_{ji} \sim F(\phi_{ji})$. The HDP generates a global probability measure $G_0 \sim \text{DP}(\gamma, P_0)$ on the parameter space Φ and then draws the group specific distribution $G_j \sim \text{DP}(\alpha, G_0)$. The global measure G_0 acts as a central distribution for G_j, with α controlling the degree of variability. The stick-breaking representation expresses G_0 and G_j as:

$$\begin{aligned} G_0 &= \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k^*}(\cdot), & \boldsymbol{\beta} = \{\beta_k\}_{k=1}^{\infty} \sim \text{Stick}(\gamma) \\ G_j &= \sum_{k=1}^{\infty} \omega_{jk} \delta_{\phi_k^*}(\cdot), & \boldsymbol{\omega}_j = \{\omega_{jk}\}_{k=1}^{\infty} \sim \text{DP}(\alpha, \boldsymbol{\beta}), \end{aligned} \tag{4.7}$$

where $\phi_k^* \overset{i.i.d}{\sim} P_0$. Since G_0 is discrete, each G_j shares the same set of global parameters but with different sets of weights. An indicator variable $z_{ji} \sim \text{Mult}(\boldsymbol{\omega}_j)$ (multinomial distribution) is generated to link an observation y_{ji} to the unique global parameter $\phi_{z_{ji}}^*$. Multiple ϕ_{ji}'s might take identical values of some global parameter ϕ_k^*.

An equivalent representation of the HDP mixture model is given via the following conditional distributions:

$$\begin{aligned} y_{ji}|z_{ji}, \{\phi_k^*\}_{k=1}^{\infty} &\sim F(\phi_{z_{ji}}^*), & \phi_k^* &\sim P_0, \\ z_{ji} &\sim \text{Mult}(\boldsymbol{\omega}_j), & \boldsymbol{\omega}_j|\alpha, \boldsymbol{\beta} &\sim \text{DP}(\alpha, \boldsymbol{\beta}), \\ \boldsymbol{\beta} &\sim \text{Stick}(\gamma). \end{aligned} \tag{4.8}$$

The underlying process generating the HDP mixture model can be described using the metaphor of a Chinese restaurant franchise (CRF) (Teh et al., 2006). There are J restaurants (groups) sharing a global menu G_0. Each jth restaurant has infinitely many tables (clusters) serving dishes (parameters) for customers (observations), and its own menu G_j defining the customer sitting structure. Integrating out the jth restaurant specific menu G_j, customer y_{ji} sits at already occupied table t_{ji} with probability proportional to the number of currently seated customers or selects a new table with probability proportional to α. The conditional distribution of dish ϕ_{ji} for customer y_{ji} is given by

$$\phi_{ji}|\boldsymbol{\phi}_j^{-i}, \alpha, G_0 \sim \sum_{t=1}^{T_j} \frac{n_{jt}}{\alpha + N_j - 1}\delta_{\phi^*_{jt}} + \frac{\alpha}{\alpha + N_j - 1}G_0, \tag{4.9}$$

where n_{jt} is the number of customers in the jth restaurant sitting at table t and sharing dish (parameter) ϕ^*_{jt}, and $\boldsymbol{\phi}_j^{-i} = \{\phi_{j1}, \ldots, \phi_{ji-1}, \phi_{ji+1}, \ldots, \phi_{jN_j}\}$. Integrating out G_0, each table serves a dish (parameter) $\phi^*_{jt} = \phi^*_{k_{jt}}$ with probability proportional to the number of other tables serving that dish in the franchise, or chooses a new dish with probability proportional to γ. The conditional distribution of the dish ϕ^*_{jt} is written as

$$\phi^*_{jt}|\boldsymbol{\Phi}^*_1, \ldots, \boldsymbol{\Phi}^*_{j-1}, \boldsymbol{\Phi}^{*(-t)}_j, \gamma, P_0 \sim \sum_{k=1}^{M} \frac{m_k}{\gamma + \sum_k m_k}\delta_{\phi^*_k} + \frac{\gamma}{\gamma + \sum_k m_k}P_0, \tag{4.10}$$

where $\boldsymbol{\Phi}^*_r = \{\phi^*_{r1}, \ldots, \phi^*_{rT_r}\}$ includes all the dishes in the rth restaurant, $\boldsymbol{\Phi}^{*(-t)}_j$ includes all the dishes in the jth restaurant except the dish ϕ^*_{jt}, m_k is the number of tables in all restaurants serving the dish ϕ^*_k, and M denotes the number of global dishes already served in the franchise. Observation y_{ji} is then drawn from the generating distribution $F(\phi^*_{z_{ji}})$ with the parameter $\phi^*_{z_{ji}} = \phi^*_{jt_{ji}} = \phi^*_{k_{jt_{ji}}}$.

The HDP mixture model can be derived as the limit of a sequence of finite hierarchical mixture model (Ishwaran and Zarepour, 2002; Teh et al., 2006; Fox et al.,

2008):

$$\begin{aligned} \boldsymbol{\beta} &\sim \text{Dir}(\gamma/L, \ldots, \gamma/L) \\ \boldsymbol{\omega}_j &\sim \text{Dir}(\alpha\beta_1, \ldots, \alpha\beta_L), \end{aligned} \tag{4.11}$$

where the number of mixture components $L \to \infty$. This weak limit approximation is useful to develop efficient blocked posterior sampling algorithm.

4.2.4 The Sticky HDP-iHMM Model

To model the situations where a system changes over time in a discontinuous fashion, regime-switching state-space models is a good choice. Regimes refer to discrete stages between which certain parameters or characteristics of the system switch. The dynamics of emotions experienced by each subject can be regarded as two discrete "regimes" along the valence dimension, positive and negative emotion phases, and unknown, potentially infinitely many "regimes" along the arousal dimension. The emotion switching system has the feature of state persistence, where a currently experienced emotion state is reluctant to move to another state. Hence, we use the sticky HDP-iHMM model (Teh et al., 2006; Ni et al., 2007; Fox et al., 2008) to address the time related state-switching structure and the self-transition feature in the time-dependent latent emotion factor $\phi_{i,t}$ in model (4.3).

Assume the subject specific time-dependent parameters $\phi_{i,t} = (\text{H}_{i,t}, \text{A}_{i,t}, \text{B}_{i,t})^T$ switching among an infinite collection of emotion and arousal regimes. Let each HDP group-specific set of weights, $\boldsymbol{\omega}_j$ act as a state-specific transition distribution in the infinite state space (infinitely many groups). Denote z_{it} for the current emotion state of the Markov chain for the ith subject at time t. The current state z_{it} is generated from the

multinomial distribution of transition parameters $\boldsymbol{\omega}_{z_{it-1}}$,

$$z_{it} \sim \text{Mult}(\boldsymbol{\omega}_{z_{it-1}}), \tag{4.12}$$

where z_{it-1} indexes the group to which y_{it} is assigned. The current state z_{it} links the parameter $\phi^*_{z_{it}}$ to $\phi_{i,t}$. To incorporate the self-transition feature, the state-specific transition distribution $\boldsymbol{\omega}_j$ is specified as,

$$\boldsymbol{\omega}_j \sim \text{DP}(\alpha + \kappa, \frac{\alpha\boldsymbol{\beta} + \kappa\delta_j}{\alpha + \kappa}), \tag{4.13}$$

where $\kappa > 0$ is the self-transition sticky parameter for the current state j, and δ_j denotes a point mass at the current state j. The original HDP-iHMM model is retrieved when $\kappa = 0$.

The Chinese restaurant metaphor of the sticky HDP-iHMM model now has a special family of loyal customers to the franchise. Denote z_{t-1} as the grandparent,z_t as the parent, and z_{t+1} as the child. The parent enters a restaurant j which its parent chose $z_{t-1} = j$, and chooses a table $t_{ji} \sim \boldsymbol{\omega}_j$ which serves dish $k_{jt_{ji}}$. It is implied that $z_t = z_{ji} = k_{jt_{ji}}$. The increased family favor will likely to drive the child to choose the same dish as its parent, and in turn, develops a loyal family to the franchise over generations. On the other side, it is also possible that the child turns away from its parent's occasional un-popular choice to the most popular dish served in this franchise. Hence, such sticky (loyalty) feature can be observed in the increased popularity of a table dish randomly chosen by

$$k_{jt} \sim \text{Mult}(\frac{\alpha\boldsymbol{\beta} + \kappa\delta_j}{\alpha + \kappa}). \tag{4.14}$$

A set of auxiliary random variables $\bar{k}_{jt}$ and q_{jt} is introduced to facilitate the inference

algorithm:

$$\begin{aligned} \bar{k}_{jt} &\sim \boldsymbol{\beta}, \\ q_{jt} &\sim \text{Ber}(\frac{\kappa}{\alpha+\kappa}), \\ k_{jt} &= \begin{cases} \bar{k}_{jt}, & q_{jt}=0; \\ j, & q_{jt}=1, \end{cases} \end{aligned} \tag{4.15}$$

where q_{jt} is an indicator variable with the value of 1 to override the current state as the previous state, and 0 to move to any state randomly silencing the sticky tendency.

4.3 Posterior Computation

In this section we develop a block z Gibbs sampler to implement the sticky HDP-iHMM circumplex model. We incorporate the block z sampler of Fox et al. (2008) to sample the parameters $\phi_{i,t} = (\mathrm{H}_{i,t}, \mathrm{A}_{i,t}, \mathrm{B}_{i,t})^T$, where a variant of the HMM forward-backward procedure is used to harness the Markov structure and jointly sample the state sequence $z_{1:T}$. A set of auxiliary variables m_{jk}, $\bar{m}_{jk}$, and q_{jt} is added to facilitate the block z sampling. We place the block constraints (Lenk et al., 2006) to sample the item angles θ_j.

We define the following notations for convenient use in the description of the sampler:

- $\mathrm{Y} = (\mathrm{Y}_{\cdot,1}, \ldots, \mathrm{Y}_{\cdot,J})$, where $\mathrm{Y}_{\cdot,j} = (\mathrm{Y}_{1,j}^T, \ldots, \mathrm{Y}_{n,j}^T)^T$, and $\mathrm{Y}_{i,j} = (\mathrm{Y}_{i,j,1}, \ldots, \mathrm{Y}_{i,j,T})^T$, for $i = 1, \ldots, n$, $j = 1, \ldots, J$,
- $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_J^2)$,
- $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)^T$,
- $X = (X_s X_c)$, where $X_s = (\sin(\theta_1), \ldots, \sin(\theta_J))^T$ and $X_c = (\cos(\theta_1), \ldots, \cos(\theta_J))^T$,

- $\mathbf{H} = (\mathrm{H}_1^T, \ldots, \mathrm{H}_n^T)^T$, where $\mathrm{H}_i = (\mathrm{H}_{i,1}, \ldots, \mathrm{H}_{i,T})^T$.
- $\mathbf{A} = (\mathrm{A}_1^T, \ldots, \mathrm{A}_n^T)^T$, where $\mathrm{A}_i = (\mathrm{A}_{i,1}, \ldots, \mathrm{A}_{i,T})^T$.
- $\mathbf{B} = (\mathrm{B}_1^T, \ldots, \mathrm{B}_n^T)^T$, where $\mathrm{B}_i = (\mathrm{B}_{i,1}, \ldots, \mathrm{B}_{i,T})^T$.

Starting from the initiation step, the posterior sampler proceeds as follows:

1. Conditional on the previous set of state-specific transition densities $\boldsymbol{\omega}^{(n-1)}$, the global transition density $\boldsymbol{\beta}^{(n-1)}$, and the model parameters $\boldsymbol{\phi}^{(n-1)}$, $\mathbf{U}^{(n-1)}$, $\boldsymbol{\theta}^{(n-1)}$, block sample the state sequences z and the parameters $\phi_{i,t}$:

   (a) Compute backward messages $m_{t,t-1}(k)$ sequentially in time. Set $\boldsymbol{\omega} = \boldsymbol{\omega}^{(n-1)}$, $\mathbf{U} = \mathbf{U}^{(n-1)}$, $\boldsymbol{\theta} = \boldsymbol{\theta}^{(n-1)}$. For $t \in \{T, \ldots, 1\}$, $k \in \{1, \ldots, L\}$, initialize messages $m_{T+1,T}(k) = 1$ and then compute $m_{t,t-1}(k) = \sum_{l=1}^{L} \omega_k(l) N(y_t; \mathbf{U}_l, \boldsymbol{\phi}_l, \boldsymbol{\theta}_l, \Sigma_l)$.

   (b) Initialize $n_{kl} = 0$ and $\mathcal{Y}_k = \emptyset$, $(k, l) \in \{1, \ldots, L\}^2$. Sequentially sample the state assignments $z_{1:T}$ forward in time

   $$z_t \sim \sum_{l=1}^{L} f_l(y_t)\delta_l,$$

   where $f_l(y_t) = \omega_{z_{t-1}}(l) N(y_t; \mathbf{U}_l, \boldsymbol{\phi}_l, \boldsymbol{\theta}_l, \Sigma_l) m_{t+1,t}(l)$. Increment $n_{z_{t-1} z_t}$ and update y_t to the cached statistics $\mathcal{Y}_{z_t}$ for sampling $\boldsymbol{\phi}_{z_t}$.

   (c) Sample the auxiliary variables m, $\bar{m}$, and q:

      i. Denote $\mathcal{J}_{kl} = \{t | z_{t-1} = k, z_t = l\}$, for each $(k, l) \in \{1, \ldots, L\}^2$. Start with $m_{kl} = 0$, $n = 0$, for each $t \in \mathcal{J}_{kl}$, sample

      $$x \sim \mathrm{Ber}\left(\frac{\alpha\beta_l + \kappa\delta(k, l)}{n + \alpha\beta_l + \kappa\delta(k, l)}\right),$$

      increment n, and if $x = 1$ increment m_{kl}.

ii. Sample the number of override variables in restaurant k:

$$q_{k\cdot} \sim \text{Binomial}(m_{kk}, \frac{\rho}{\rho + \beta_k(1-\rho)}),$$

for $k \in \{1, \ldots, L\}$, where $\rho = \frac{\kappa}{\alpha+\kappa}$.

iii. Set the number of informative tables in restaurant k serving dish l as:

$$\bar{m}_{kl} = \begin{cases} m_{kl}, & k \neq l; \\ m_{kk} - q_{k\cdot}, & k = l, \end{cases}$$

(d) Sample the global transition density via

$$\boldsymbol{\beta} \sim \text{Dir}(\gamma/L + \bar{m}_{\cdot 1}, \ldots, \gamma/L + \bar{m}_{\cdot L}).$$

(e) Sample the new transition densities $\boldsymbol{\omega}_k$ and the model parameters $\boldsymbol{\phi}_k$, $k = 1, \ldots, L$ via

$$\begin{aligned} \boldsymbol{\omega}_k &\sim \text{Dir}(\alpha\beta_1 + n_{k1}, \ldots, \alpha\beta_k + \kappa + n_{kk}, \ldots, \alpha\beta_L + n_{kL}), \\ \boldsymbol{\phi}_k &\sim p(\boldsymbol{\phi}|\eta, \mathcal{Y}_k). \end{aligned}$$

2. Sample $\mathbf{U}$ from its conditional distribution $N(\mu_u, \Sigma_u)$, where $\mu_u = \Sigma_u(\Sigma^{-1}(\mathrm{Y} - \mathbf{H}1_J^T - \mathbf{A}X_s^T - \mathbf{B}X_c^T)^T 1_{nT} + \Sigma_0^{-1}\mu_0)$ and $\Sigma_u = (n\Sigma^{-1} + \Sigma_0^{-1})$.

3. Sample σ_j^{-2} from its conditional distribution $\text{Ga}(r_0 + nT/2, s_0 + 0.5\sum_{i,T}(\mathrm{Y}_{i,j,t} - \mathrm{U}_j - \mathrm{H}_{i,t} - \mathrm{A}_{i,t}\sin(\theta_j) - \mathrm{B}_{i,t}\cos(\theta_j))^2)$.

4. Sample λ^{-2} from its conditional distribution $\text{Ga}(u_{0,1} + nT/2, v_{0,1} + \mathbf{H}^T\mathbf{H}/2)$.

5. Sample τ^{-2} from its conditional distribution $\text{Ga}(u_{0,2} + nT/2, v_{0,2} + \mathbf{A}^T\mathbf{A}/2 + \mathbf{B}^T\mathbf{B}/2)$.

6. Sample the latent cutpoints $c_{i,g}$ from the conditional distributions $\text{Unif}(\max(\bar{y}_g, c_{i,g-1}), \min(\underline{y}_{g+1}, c_{i,g+1}))$, where $\bar{y}_g = \max\{\text{Y}_{i,j,t} : \text{W}_{i,j,t} = g\}$ and $\underline{y}_{g+1} = \min\{\text{Y}_{i,j,t} : \text{W}_{i,j,t} = g+1\}$, $g = 2, \ldots, G-2$.

7. Sample the latent responses $\text{Y}_{i,j,t}$ from the conditional distributions:

   (a) for observed $\text{W}_{i,j,t}$,

$$p(\text{Y}_{i,j,t}|-) \sim N(\text{U}_j + \text{H}_{i,t} + \text{A}_{i,t}\sin\theta_j + \text{B}_{i,t}\cos\theta_j, \sigma_j^2)\mathcal{X}(c_{i,\text{w}_{i,j,t}-1} < \text{Y}_{i,j,t} < c_{i,\text{w}_{i,j,t}}).$$

   (b) for missing $\text{W}_{i,j,t}$,

$$p(\text{Y}_{i,j,t}|-) \sim N(\text{U}_j + \text{H}_{i,t} + \text{A}_{i,t}\sin\theta_j + \text{B}_{i,t}\cos\theta_j, \sigma_j^2).$$

8. Sample the item angles $\boldsymbol{\theta}$ using a hybrid sampling strategy. These angles are placed in C blocks of constraints $\mathcal{B}_1, \ldots, \mathcal{B}_C$. Denote $\underline{\mathcal{B}}_c = \min\{\theta_j : j \in \mathcal{B}_c\}$ and $\bar{\mathcal{B}}_c = \max\{\theta_j : j \in \mathcal{B}_c\}$ for the minimum and maximum angles in the cth block, $c = 2, \ldots, C$. For the first block, the "minimum" and "maximum" angles are defined as

$$\begin{aligned}
\underline{\mathcal{B}}_1 &= \begin{cases} 2\pi, & \text{if } \underline{\mathcal{B}}_C > \max\{\theta_j : j \in \mathcal{B}_1\}; \\ \min\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j > \bar{\mathcal{B}}_C\}, & \text{if } \underline{\mathcal{B}}_C < \max\{\theta_j : j \in \mathcal{B}_1\}, \end{cases} \\
\bar{\mathcal{B}}_1 &= \max\{\theta_j : j \in \mathcal{B}_1 \text{ and } \theta_j < \underline{\mathcal{B}}_2\}.
\end{aligned}$$

   These blocks satisfy the ordering

$$0 \leq \bar{\mathcal{B}}_1 < \underline{\mathcal{B}}_2 < \bar{\mathcal{B}}_2 < \ldots < \underline{\mathcal{B}}_C < \bar{\mathcal{B}}_C < \underline{\mathcal{B}}_1 \leq 2\pi.$$

(a) Gibbs step: sample $\boldsymbol{\theta}$ from the conditional distribution

$$\begin{aligned} p(\boldsymbol{\theta}|-) &\propto \prod_{j=1}^{J} VM(d_j, Q_j, \mathcal{C}_j), \\ Q_j &= \sigma_j^{-2} \begin{pmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{A}^T\mathbf{B} \\ \mathbf{B}^T\mathbf{A} & \mathbf{B}^T\mathbf{B} \end{pmatrix} + Q_0, \\ d_j &= Q_j^{-1}\{(\mathbf{A} \ \ \mathbf{B})^T(\mathrm{Y}_{\cdot,j} - \mathrm{U}_j 1_{nT} + \mathbf{A}) + Q_0 d_0\} \end{aligned}$$

where $\mathcal{C}_j$ is the constraint set for θ_j.

(b) Metropolis step: generate θ_j from a mixture of R uniform distributions. Denote $b_1, \ldots, b_C$ for the indices of the last angle in each block such that

$$1, \ldots, b_1 \in \mathcal{B}_1 \ \text{ and } \ b_{c-1}+1, \ldots, b_c \in \mathcal{B}_c, \ \ c = 2, \ldots, C.$$

i. For $\theta_j \in \mathcal{B}_1$, generate the candidate angle from the jump distributions

$$\begin{aligned} g_1(\psi_m|\psi_2, \ldots, \psi_{m-1}, \theta_m, \ldots, \theta_J) &\sim \sum_{r=1}^{R} p_r \frac{\mathcal{X}(a_{r,m,1} < \psi_m < b_{r,m,1})}{b_{r,m,1} - a_{r,m,1}}, \\ a_{r,m,1} &= \max\{\bar{\mathcal{B}}_C - 2\pi, \theta_m - u_r\}, \\ b_{r,m,1} &= \min\{\underline{\mathcal{B}}_2, \theta_m + u_r\}, \end{aligned}$$

for $m = 2, \ldots, b_1$. If $\psi_m < 0$, set $\psi_m = 2\pi + \psi_m$.

ii. For $\theta_j \in \bigcup_{c=2}^{C} \mathcal{B}_c$, generate the candidate angle from the uniform random walk

$$\begin{aligned} g_c(\psi_m|\psi_2, \ldots, \psi_{m-1}, \theta_m, \ldots, \theta_J) &\sim \sum_{r=1}^{R} p_r \frac{\mathcal{X}(a_{r,m,c} < \psi_m < b_{r,m,c})}{b_{r,m,c} - a_{r,m,c}}, \\ a_{r,m,1} &= \max\{\bar{\mathcal{B}}_{c-1}, \theta_m - u_r\}, \\ b_{r,m,1} &= \min\{\underline{\mathcal{B}}_{c+1}, \theta_m + u_r\}, \end{aligned}$$

for $m = b_{c-1}+1, \ldots, b_c$. Note $\underline{\mathcal{B}}_{C+1} = \underline{\mathcal{B}}_1$.

The acceptance probability for the candidates are

$$\min\{1, \frac{[\boldsymbol{\psi}|Y]\prod_{c=1}^{C}\prod_{m=b_{c-1}+1}^{b_c} g_c(\theta_m|\theta_2,\ldots,\theta_{m-1},\psi_m,\ldots,\psi_J)}{[\boldsymbol{\theta}|Y]\prod_{c=1}^{C}\prod_{m=b_{c-1}+1}^{b_c} g_c(\psi_m|\psi_2,\ldots,\psi_{m-1},\theta_m,\ldots,\theta_J)}\},$$

where $b_0 + 1 = 2$.

We use a modified Brier score (Lenk et al., 2006) as the fit measure. After the burn in, we compute a Brier score at each iteration h:

$$BS^{(h)} = \frac{1}{NG}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{g=1}^{G}\delta_{i,j,t}(z_{i,j,t,g} - P(\mathrm{W}_{i,j,t} = g|\Omega^{(h)}))^2,$$

where N is the total number of observations, the indicator variable $z_{i,j,t,g} = 1$ if the ith subject responded g to the jth item at time t and 0 otherwise, $\delta_{i,j,t}$ takes the value of 1 if this record is not missing, and $P(\mathrm{W}_{i,j,t} = g|\Omega)$ is the predictive probability for $\mathrm{W}_{i,j,t} = g$ conditional on the parameters Ω and the data.

4.4 Simulation Study

We conducted a Monte Carlo simulation study to access the performance of the sticky HDP-iHMM circumplex model. We simulated a three-state sequence data containing $T = 9$ time points. At each state s, we generated 16 item responses on a 7 point ordinal scale for 50 subjects from the following model:

$$\begin{aligned} \mathrm{Y}_{i,j,s} &= \mathrm{U}_j + \mathrm{H}_{i,s} + \mathrm{A}_{i,s}\sin\theta_j + \mathrm{B}_{i,s}\cos\theta_j + \epsilon_{i,j,s}, \\ \mathrm{W}_{i,j,s} &= g, \quad \text{iff} \quad c_{g-1} \le \mathrm{Y}_{i,j,s} < c_g, \quad \text{for} \quad g = 1,\ldots,7. \end{aligned} \tag{4.16}$$

where the item specific response means U_j were drawn from $N(0,1)$, the item specific angles θ_j were randomly generated under the constraints of four blocks along the circle with four angles in each block, and the state specific random factors $(\mathrm{H}_{i,s}, \mathrm{A}_{i,s}, \mathrm{B}_{i,s})^T$ were generated from a 3-dimensional normal distribution with the Gaussian mean of $(-10, 0, 10)^T$ and the diagonal covariance matrix with 0.01 variances. The measurement error $\epsilon_{i,j,s}$ was assumed to be drawn from a normal distribution with mean 0 and variance 0.01. The cutoff points were set as $(-1, -0.6, -0.2, 0.2, 0.6, 1)$.

We analyzed the simulated data using our sticky HDP-iHMM circumplex model and applied the block-z Gibbs sampler. The block-z Gibbs sampler ran 7,000 iterations, with the first 2000 samples discarded as burn-in. Every 5th sample was collected to thin the chain. Our Gibbs sampler converges rapidly and exhibits efficient mixing. We set 0.98 for the self-transition probability in the sticky HDP-iHMM model. We assigned $\mathrm{Ga}(1,1)$ hyperpriors for the hyperparameters α and γ. We used a truncation level of $L = 5$ in the block sampler, while letting the sampler learn a strict subset of the available states.

Our algorithm was able to recover the true states accurately. Fig. 4.2 plots the true state sequence in the upper panel and the labeled state sequence learned in the sampler in the lower panel. Three colors stand for three states, navy for state 1, dodger blue for state 2, and green yellow for state 3. The true state sequence shares the same color map as the labeled state sequence. The posterior estimates of the angles are accurate for the angles in the 2nd to 4th block, while have relatively large bias for the angles in the first block (Table 4.1). The brier score comparing the posteriori predictive probabilities and the ordinal outcomes was 0.056.

4.5 PANAS Data Analysis

The PANAS data comes from a large emotion study. This study was designed to investigate the emotion dynamics. Participants were asked to complete a "daily" survey four times a day for a total of thirty days in addition to one "nightly" survey each day. We included 55 individuals who did the daily self-reports of their feelings over the first 7 consecutive days. Eight ordinal positive emotion (PE) items and nine ordinal negative emotion (NE) items measured on a scale from 1 (= none) to 4 (= always) were used for model fitting purpose in the chapter. The PE items include enthusiastic, interested, determined, active, strong, proud, inspired, and attentive. The NE items included scared, afraid, upset, nervous, ashamed, guilty, distressed, jittery, and irritable.

We analyzed the PANAS data set using the sticky HDP-iHMM circumplex model. We collected 1000 samples for every 5th samples from the block-z Gibbs sampler, after the initial burn in of 2000 iterations. We initiated a truncation level of $L = 5$, and placed $\text{Ga}(1, 1)$ hyperpriors for the hyperparameters α and γ. We chose 0.98 for the self-transition probability, favoring state persistence. The chain converges rapidly and exhibits efficient mixing. We estimated the item specific direction posteriori and labeled the emotion states over the seven days.

The analysis results for the PANAS data are displayed in Fig. 4.3 and 4.4. The emotions were labeled as two states over the seven days (Fig. 4.3), where the first six days fell in one state, and the seventh day fell in the other state. This result is supported by the fact that day 7 has the smallest averages of both the difference and the ratio between the PE scores and the NE scores (Table 4.2). This phenomenon concludes that the prolonged effects on emotions from the emotion induction exposure in the laboratory experiment extended for six days and elapsed in the seventh day. The estimates of the emotion directions along the unit circle were graphed in Fig. 4.4, where the PE items are in the right panel and the NE items are in the left panel. The

brier score evaluating the distance between the posteriori predictive probabilities and the ordinal outcomes was 0.063.

4.6 Discussion

We demonstrated the considerable benefits of the extended circumplex model in which the time related state switching structure is constructed by the sticky HDP-iHMM model. The circumplex model has been appealing to social science researchers because of its circular correlation structure of measure items. Our approach further yields another appealing feature allowing the capture of the state persistence/switching dynamics. We have also presented an efficient sampling technique. The synthetic data analysis and the empirical application clearly demonstrate the practical importance of our state-space extensions. We believe that our contributions will facilitate rigorous tests and increase the popularity of circumplex models in the analysis of psychological measures.

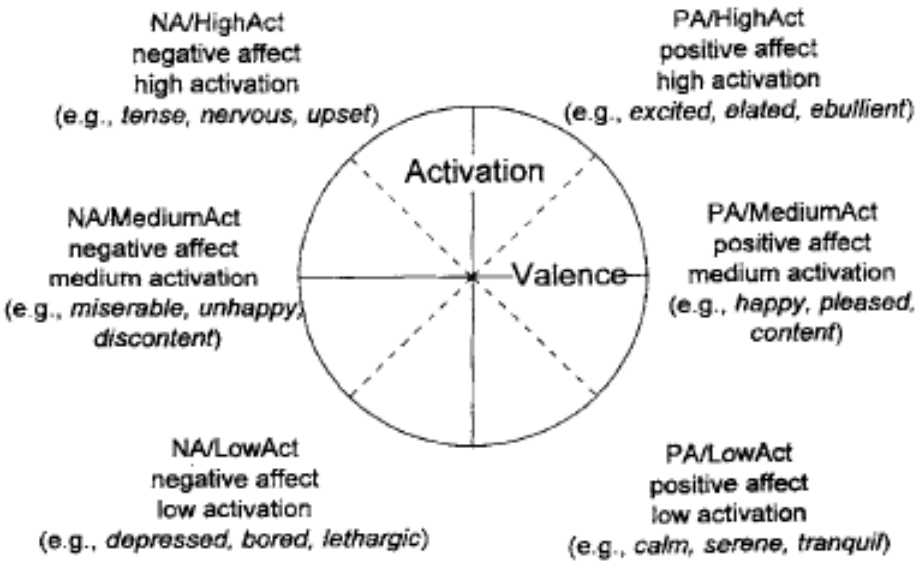

Fig. 4.1: A circular structure of affects classified into six clusters. PA = positive affect; NA = negative affect; HighAct = high activation; MediumAct = medium activation; LowAct = low activation.

Table 4.1: Simulation study: estimated directions for the synthetic data.

Item	Block	True	Estimate	Bias	Standardized Bias
1	1	0.000	0.000	0.000	0.000
2	1	1.425	0.697	-0.723	-0.511
3	1	0.865	0.701	-0.165	-0.190
4	1	1.210	0.650	-0.559	-0.462
5	2	2.828	3.054	0.226	0.080
6	2	2.731	2.853	0.121	0.045
7	2	1.793	1.540	-0.254	-0.141
8	2	2.259	2.840	0.581	0.257
9	3	3.692	3.571	-0.121	-0.033
10	3	3.893	3.594	-0.299	-0.077
11	3	4.064	3.597	-0.467	-0.115
12	3	3.371	3.323	-0.048	-0.014
13	4	6.135	6.148	0.014	0.002
14	4	5.718	5.875	0.157	0.028
15	4	4.968	4.996	0.028	0.006
16	4	5.601	5.884	0.283	0.051

Table 4.2: PANAS data: the averages of the difference and the ratio between positive emotion (PE) scores and negative emotion (NE) scores.

		day						
Measure	Average	1	2	3	4	5	6	7
PE-NE	Average	1.44	1.41	1.48	1.37	1.31	1.37	1.29
PE/NE	Average	2.23	2.22	2.34	2.29	2.18	2.22	2.17

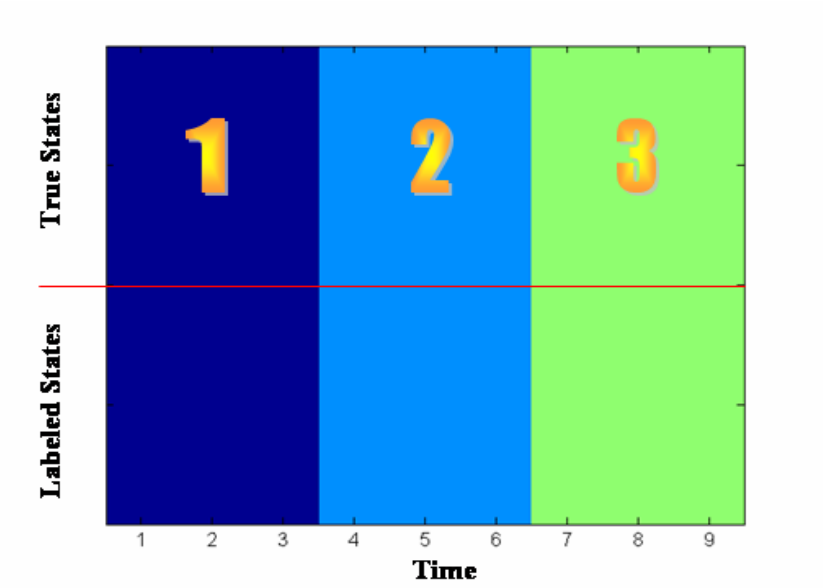

Fig. 4.2: Performance of state labeling in the simulated data. Upper panel: the true state sequence; lower panel: the labeled state sequence.

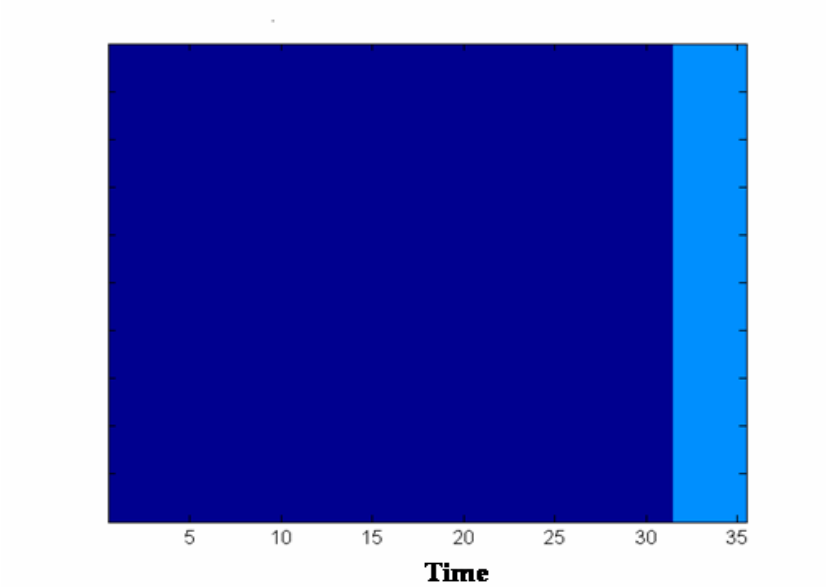

Fig. 4.3: The classified states for the PANAS data.

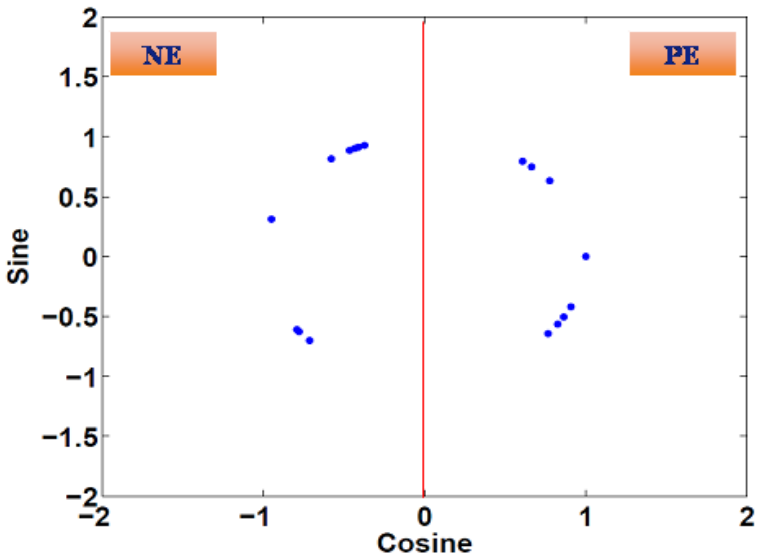

Fig. 4.4: Plot of the sine versus cosine of the posteriori estimated angles for the PANAS data. Right panel: positive emotion (PE) items; left panel: positive emotion (NE) items;

Bibliography

Ahmad, I., Leelahanon, S., and Li, Q. (2005), "Efficient estimation of a semiparametric partially linear varying coefficient model," *The Annals of Statistics*, 33, 258283.

Anderson, T. (1960), "Some stochastic process models for intelligence test scores," in *In K.J. Arrow, S. Karlin, and P. Suppes (Eds.), Mathematical methods in the social sciences*, Stanford, University Press, pp. 205–220.

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

Baladandayuthapani, V., Mallick, B. K., Young Hong, M., Lupton, J. R., Turner, N. D., and Carroll, R. J. (2008), "Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis," *Biometrics*, 64, 64–73.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994a), "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance Ser. B*, 103, 247–254.

— (1994b), "MR diffusion tensor spectroscopy and imaging," *Biophysical Journal*, 66, 259–267.

Behseta, S., Kass, R. E., and Wallstrom, G. L. (2005), "Hierarchical models for assessing variability among functions," *Biometrika*, 92, pp. 419–434.

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010), "A spatio-temporal downscaler for output from numerical models," *Journal of Agricultural Biological and Environmental Statistics*, 15, 176–197.

Bhattacharya, A. and Dunson, D. B. (2011), "Sparse Bayesian infinite factor models," *Biometrika*, 98, 291–306.

Biller, C. and Fahrmeir, L. (1997), "Bayesian spline-type smoothing in generalized regression models," .

Blackwel, D. and Macqueen, J. (1973), "Ferguson distributions via pólya urn schemes," *Annals Of Statistics*, 1, 353–355.

Bonekam, D., Nagae, L. M., Degaonkar, M., Matson, M., Abdalla, W. M., Barker, P. B., Mori, S., and Horská, A. (2008), "Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences," *NeuroImage*, 34, 733–742.

Botts, C. H. and Daniels, M. J. (2008), "A flexible approach to Bayesian multiple curve fitting," *Computational Statistics & Data Analysis*, 52, 5100–5120.

Bowman, F. D., Caffo, B., Bassett, S. S., and Kilts, C. (2008), "A Bayesian hierarchical framework for spatial modeling of fMRI data," *NeuroImage*, 39, 146 – 156.

Brezger, A., Fahrmeir, L., and Hennerfeind, A. (2007a), "Adaptive Gaussian Markov random fields with applications in human brain mapping," *Journal of Royal Statistical Society C*, 56, 327–345.

— (2007b), "Adaptive Gaussian Markov random fields with applications in human brain mapping," *J. Roy. Statist. Soc. C*, 56, 327–345.

Brezger, A. and Lang, S. (2006), "Generalized structured additive regression based on Bayesian P-splines," *Comput. Statist. Data Anal*, 50, 967–991.

Browne, M. (1992), "Circumplex models for correlation matrices," *Psychometrika*, 57, 469–497, 10.1007/BF02294416.

Brumback, B. and Rice, J. A. (1998), "Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion)," *J. Amer. Statist. Assoc.*, 93, 961994.

Bush, C. and MacEachern, S. (1996), "A semiparametric Bayesian model for randomised block designs," *Biometrika*, 83, 275–285.

Cai, J., Fan, J., Zhou, H., and Zhou, Y. (2007), "Hazard models with varying-coefficients for multivariate failure time data," *Ann Statist*, 35, 324354.

Cai, Z., Fan, J., and Li, R. (2000a), "Efficient estimation and inferences for varying-coefficient models," *J. Amer. Statist. Assoc.*, 95, 888902.

Cai, Z., Fan, J., and Yao, Q. (2000b), "Functional-coefficient regression models for nonlinear time series," *J. Amer. Statist. Assoc.*, 95, 941956.

Cai, Z. and Sun, Y. (2003), "Local linear estimation for time-dependent coefficients in Coxs regression model," *Scand J of Statist*, 30, 93111.

Camara, E., Bodammer, N., Rodriguez-Fornells, A., and Tempelmann, C. (2007), "Age-related water diffusion changes in human brain: A voxel-based approach," *NeuroImage*, 34, 1588–1599.

Carroll, R. J., Ruppert, D., and Welsh, A. H. (1998), "Local estimating equations," *J. Amer. Statist. Assoc.*, 93, 214227.

Carvalho, C. M., Chang, J., Lucas, J. E., Nevins, J. R., Wang, Q., and West, M. (2008), "High-dimensional sparse factor modeling: Applications in gene expression genomics," *J. Amer. Statist. Assoc.* , 103, 1438–1456.

Carver, C. S. and Scheier, M. F. (1982), "Control-theory - a useful conceptual-framework for personality-social, clinical, and health psychology," *Psychological Bulletin*, 92, 111–135.

Cascio, C. J., Gerig, G., and Piven, J. (2007), "Diffusion tensor imaging: Application to the study of the developing brain," *J Am Acad Child Adolesc Psychiatry*, 46, 213–223.

Chen, R. and Tsay, R. J. (1993), "Functional-coefficient autoregressive models," *J. Amer. Statist. Assoc.*, 88, 298308.

Chen, Y. S., An, H. Y., Zhu, H. T., Stone, T., Smith, J. K., Hall, C., Bullitt, E., Shen, D. G., and Lin, W. L. (2009), "White matter abnormalities revealed by diffusion tensor imaging in non-demented and demented HIV+ patients," *NeuroImage*, 47, 1154–1162.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001), "Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables," *J. Amer. Statist. Assoc.*, 96, 605619.

Chow, S.-M., Nesselroade, J. R., Shifren, K., and McArdle, J. J. (2004), "Dynamic structure of emotions among individuals with Parkinsons disease," *Structural Equation Modeling*, 11, 560 – 582.

Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991), *Local regression models. In statistical models in S (Chambers, J. M. and Hastie, T. J., eds),*, Wadsworth & Brooks, Pacific Grove.

Cox, D. R. (1972), "Regression models and life tables (with discussion)," *J. Roy. Statist. Soc. B*, 34, 187220.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007), "Spatially adaptive Bayesian penalized splines with heteroscedastic errors," *Journal of Computational and Graphical Statistics*, 16, 265–288.

Cruz-Mesía, R. D. L., Fernando, A. Q., and Müeller, P. (2007), "Semiparametric Bayesian classification with longitudinal markers," *Journal of Royal Statistical Society C*, 56, 119–137.

Csikszentmihalyi, M. (1998), *Finding Flow: The Psychology of Engagement With Everyday Life*, Basic Books.

Cudeck, R. (1986), "A note on structural models on the circumplex," *Psychometrika*, 51, 143147.

Dahl, D. B. (2007), "Comment on article by jain and neal," *Bayesian Analysis*, 2, 473–477.

De Iorio, M., Müller, P., Rosner, G., and MacEachern, S. (2004), "An ANOVA model for dependent random measures," *J. Amer. Statist. Assoc.*, 99, 205–215.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998a), "A Bayesian CART algorithm," *Biometrika*, 85, 363–377.

— (1998b), "Automatic bayesian curve fitting," *Journal of Royal Statistical Society B*, 60, pp. 333–350.

Dimatteo, I., Genovese, C. R., and Kass, R. E. (2001), "Bayesian curve fitting with free-knot splines," *Biometrika*, 88, 1055–1071.

Dunson, D. B. (2009), "Nonparametric Bayes local partition models for random effects," *Biometrika*, 96, 249–262.

Eilers, P. H. C. and Marx, B. D. (1996), "Flexible smoothing with B-splines and penalties," *Statistical Science*, 11, 89–102.

Escobar, M. and West, M. (1995), "Bayesian density estimation and inference using mixtures," *J. Amer. Statist. Assoc.*, 90, 577–588.

Eubank, R. L., Huang, C. F., Maldonado, Y. M., Wang, N., Wang, S. J., and Buchanan, R. J. (2004), "Smoothing spline estimation in varying-coefficient models," *J. Roy. Statist. Soc. B*, 66, 653–667.

Fahrmeir, L., Kneib, T., and Lang, S. (2004), "Penalized structured additive regression for space-time data: A Bayesian perspective ," *Statistica Sinica*, 14, 715–745.

Fahrmeir, L. and Lang, S. (2001), "Bayesian inference for generalized additive mixed models based on Markov random field priors," *Journal of Royal Statistical Society C*, 50, pp. 201–220.

Fan, J. (1993), "Local linear regression smoothers and their minimax efficiencies," *The Annals of Statistics*, 21, 196–216.

Fan, J. and Gijbels, I. (1992), "Variable bandwidth and local linear regression smoothers," *Annals of statistics*, 21, 433–446.

— (1995), "Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation," *J. Roy. Statist. Soc. B*, 57, 371394.

— (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Fan, J., Heckman, N. E., and Wand, M. P. (1995), "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," *Journal of the American Statistical Association*, 90, pp. 141–150.

Fan, J., Huang, T., and Li, R. Z. (2007), "Analysis of longitudinal data with semiparametric estimation of covariance function," *J. Amer. Statist. Assoc.*, 35, 632641.

Fan, J., Lin, H., and Zhou, Y. (2006), "Local partial-likelihood estimation for lifetime data," *Ann Statist*, 34, 290325.

Fan, J. and Zhang, J. T. (2000a), "Functional linear models for longitudinal data," *J. Roy. Statist. Soc. B*, 62, 303322.

Fan, J. and Zhang, W. (1999), "Statistical estimation in varying coefficient models," *Ann. Statist.*, 27, 14911518.

— (2000b), "Simultaneous confidence bands and hypothesis testing in varying-coefficient models," *Scand. J. Statist.*, 27, 715731.

— (2008), "Statistical methods with varying coefficient models," *Statistics and Its Interface*, 1, 179195.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, 1, 209–230.

— (1974), "Prior distributions on spaces of probability measures," *The Annals of Statistics*, 2, pp. 615–629.

Focke, N. K., Yogarajah, M., Bonelli, S. B., Bartlett, P. A., Symms, M. R., and Duncan, J. S. (2008), "Voxel-based diffusion tensor imaging in patients with mesial temporal lobe epilepsy and hippocampal sclerosis," *NeuroImage*, 40, 728–737.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2008), "An HDP-HMM for systems with state persistence," in *Proceedings of the 25th international conference on Machine learning*, New York, NY, USA: ACM, ICML '08, pp. 312–319.

Frijda, N. H. (1986), "The Emotions," in *The Emotions*, Cambridge(UK): Cambridge University Press, p. 207.

Gelfand, A. E. and Vounatsou, P. (2003), "Proper multivariate conditional autoregressive models for spatial data analysis," *Biostatistics*, 4, 11–15.

Geweke, J. and Zhou, G. (1996), "Measuring the pricing error of the arbitrage pricing theory," *Review of Financial Studies*, 9, 557–587.

Gilmore, J. H., Smith, L. C., Wolfe, H. M., Hertzberg, B. S., Smith, J. K., Chescheir, N. C., Evans, D. D., Kang, C., Hamer, R. M., Lin, W., and Gerig, G. (2008), "Prenatal mild ventriculomegaly predicts abnormal development of the neonatal brain," *Biol Psychiatry*, 64, 1069–1076.

Goldsmith, A., Feder, J., Crainiceanu, C., Caffo, B., and Reich, D. (2011), "Penalized functional regression," *Journal of Computational and Graphical Statistics*.

Goodlett, C. B., Fletcher, P. T., Gilmore, J. H., and Gerig, G. (2009), "Group analysis of DTI fiber tract statistics with application to neurodevelopment," *NeuroImage*, 45, S133–S142.

Gössl, C., Auer, D. P., and Fahrmeir, L. (2001), "Bayesian Spatiotemporal Inference in Functional Magnetic Resonance Imaging," *Biometrics*, 57, 554–562.

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010), "Longitudinal functional principal component analysis," *Electron. J. Stat.*, 4, 1022–1054.

Guttman, L. (1954), "A new approach to factor analysis: The radex." in *In P.F. Lazarsfeld (Ed.), Mathematical thinking in the social sciences*, New York: Columbia University Press, pp. 258–348.

Harezlak, J., Wang, M., Christiani, D., and Lin, X. (2007), "Quantitative quality-assessment techniques to compare fractionation and depletion methods in SELDI-TOF mass spectrometry experiments," *Bioinformatics*, 23, 2441–2448.

Hasan, K. M., Basser, P. J., Parker, D. L., and Alexander, A. L. (2001), "Analytical computation of the eigenvalues and eigenvectors in DT-MRI," *J Magn Reson*, 152, 41–47.

Hasan, K. M. and Narayana, P. A. (2003), "Computation of the fractional anisotropy and mean diffusivity maps without tensor decoding and diagonalization: theoretical analysis and validation," *Magn Reson Med*, 50, 589–598.

Hastie, T. and Tibshirani, R. (2000), "Bayesian backfitting," *Statistical Science*, 15, pp. 196–213.

Hastie, T. J. and Tibshirani, R. J. (1993), "Varying-coefficient models," *J. Roy. Statist. Soc. B.*, 55, 757796.

He, J. H., Mcgee, D. L., and Niu, X. F. (2010), "Application of the Bayesian dynamic survival model in medicine," *Statistics in Medicine*, 29, 347–360.

Hecke, W. V., Sijbers, J., Backer, S. D., Poot, D., Parizel, P. M., and Leemans, A. (2009), "On the construction of a ground truth framework for evaluating voxel-based diffusion tensor MRI analysis methods," *NeuroImage*, 46, 692–707.

Hogan, J. W., Lin, X., and Herman, B. (2004), "Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout," *Biometrics*, 60, 854864.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998), "Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data," *Biometrika*, 85, 809822.

Huang, J. Z. and Shen, H. (2004), "Functional coefficient regression models for nonlinear time series: A polynomial spline approach," *Scandinavian Journal of Statistics*, 31, 515534.

Huang, J. Z., Wu, C. O., and Zhou, L. (2002), "Varying-coefficient models and basis function approximations for the analysis of repeated measurements," *Biometrika*, 89, 111128.

— (2004), "Polynomial spline estimation and inference for varying coefficient models with longitudinal data," *Statistica Sinica*, 14, 763788.

Huang, Y., Liang, H., and Wu, H. (2008), "Identifying significant covariates for anti-HIV treatment response: Mechanism-based differential equation models and empirical semiparametric regression models," *Statistics in Medicine*, 27, 4722–4739.

Huitt, W. (2003), "The affective system," Tech. rep., Valdosta, GA: Valdosta State University.

Ishwaran, H. and James, L. F. (2001), "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association*, 96, pp. 161–173.

Ishwaran, H. and Zarepour, M. (2002), "Exact and approximate representations for the sum Dirichlet process," *Canadian Journal Of Statistics-Revue Canadienne De Statistique*, 30, 269–283.

Jeffreys, W. and Berger, J. (1992), "Ockham's razor and Bayesian analysis," *American Statistician*, 80, 64–72.

Jones, D. K., Symms, M. R., Cercignani, M., and Howard, R. J. (2005), "The effect of filter size on VBM analyses of DT-MRI data," *NeuroImage*, 26, 546–554.

Kauermann, G. (2005), "A note on smoothing parameter selection for penalized spline smoothing," *Journal of Statistical Planning and Inference*, 127, 53–69.

Kauermann, G. and Tutz, G. (1999), "On model diagnostics using varying coefficient models," *Biometrika*, 86, 119128.

Kleinman, K. and Ibrahim, J. (1998), "A semiparametric Bayesian approach to the random effects model," *Biometrics*, 54, 921–938.

Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F. A., and Shenton, M. E. (2007), "A review of diffusion tensor imaging studies in schizophrenia," *Journal of Psychiatric Research*, 41, 15–30.

Larsen, R. (2000), "Toward a science of mood regulation," *Psychological Inquiry*, 11, 129–141.

Lee, D. and Shaddick, G. (2007), "Time-varying coefficient models for the analysis of air pollution and health outcome data," *Biometrics*, 63, 1253–1261.

Lei, X., Timothy, D. J., Thomas, E. N., and Derek, N. E. (2009), "Modeling Inter-Subject Variability in fMRI Activation Location: A Bayesian Hierarchical Spatial Model," *Biometrics*, 65, 1041–1051.

Lenglet, C., Campbell, J. S. W., Descoteaux, M., Haro, G., Savadjiev, P., Wassermann, D., Anwander, A., Deriche, R., Pike, G. B., Sapiro, G., Siddiqi, K., and Thompson, P. M. (2009), "Mathematical methods for diffusion MRI processing," *NeuroImage*, 45, S111–S122.

Lenk, P., Wedel, M., and B¨ckenholt, U. (2006), "Bayesian estimation of circumplex models subject to prior theory constraints and scale-usage bias," *Psychometrika*, 71, 33–55, 10.1007/s11336-001-0958-4.

Lepore, N., Brun, C. A., Chou, Y., Chiang, M., Dutton, R. A., Hayashi, K. M., Luders, E., Lopez, O. L., Aizenstein, H. J., Toga, A. W., Becker, J. T., and Thompson, P. M. (2008), "Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors," *IEEE Transactions in Medical Imaging*, 27, 129–141.

Liang, H., Wu, H. L., and Carroll, R. J. (2003), "The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error," *Biostatistics*, 4, 297–312.

Lin, D. Y. and Ying, Z. (2001), "Semiparametric and nonparametric regression analysis of longitudinal data (with discussions)," *J. Amer. Statist. Assoc.*, 96, 103126.

Lin, X. and Carroll, R. J. (2001), "Semiparametric regression for clustered data using generalized estimating equations," *J. Amer. Statist. Assoc.*, 96, 10451056.

LO, A. (1984), "On a class of Bayesian nonparametric estimates. 1. Density estimates." *The Annals of Statistics*, 12, 351–357.

Lu, Z., Steinskog, D. J., Tjstheim, D., and Yao, Q. (2009), "Adaptively varying-coefficient spatiotemporal models," *Journal of Royal Statistical Society B*, 71, 859–880.

Maceachern, S. (1994), "Estimating normal means with a conjugate style Dirichlet process prior," *Communications in Statistics-Simulation and Computation*, 23, 727–741.

Malloy, E. J., Morris, J. S., Adar, S. D., Suh, H., Gold, D. R., and Coull, B. A. (2010), "Wavelet-based functional linear mixed models: An application to measurement error-corrected distributed lag models," *Biostatistics*, 11, 432–452.

Marina, S. P., Dani, G., Flávia, M. L., and Esther, S. (2008), "Spatially varying dynamic coefficient models," *Journal of Statistical Planning and Inference*, 138, 1038–1058.

Mirr, M. P. (2001), ""Abnormally Increased Behavioral Arousal" Cris Stewart- Amidei and Joyce A. Kunkel." in *Neuroscience Nursing: Human Response to Neurologic Dysfunction*, W. B. Sunders Philadelphia: PA.

Mooney, C., Vullo, A., and Pollastri, G. (2006), "Protein structural motif prediction in multidimensional phi-psi space leads to improved secondary structure prediction," *Journal of Computational Biology*, 13, 1489–1502.

Morris, J. and Carroll, R. (2006), "Wavelet-based functional mixed models," *Journal of Royal Statistical Society B*, 68, 179–199.

Moseley, M. (2002), "Diffusion tensor imaging and aging-a review," *NMR Biomed.*, 15, 553–560.

Mukherjee, P. and McKinstry, R. C. (2006), "Diffusion tensor imaging and tractography of human brain development," *Neuroimaging Clin N Am*, 16, 19–43.

Müller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004), "Optimal sample size for multiple testing: The case of gene expression microarrays," *Journal of the American Statistical Association*, 99, 990–1001.

Müller, P. and Rosner, G. (1997), "A Bayesian population model with hierarchical mixture priors applied to blood count data," *J. Amer. Statist. Assoc.*, 92, 1279–1292.

Ni, K., Qi, Y., and Carin, L. (2007), "Multiaspect target detection via the infinite hidden Markov model," *Journal Of The Acoustical Society Of America*, 121, 2731–2742.

Nicholls, D. F. and Quinn, B. G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, New York.

O'Donnell, L., Westin, C.-F., and Golby, A. (2009), "Tract-based morphometry for white matter group analysis," *Neuroimage*, 45, 832–844.

Ohlssen, D. I., Sharples, L. D., and Spiegelhalter, D. J. (2007), "Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons," *Statistics in Medicine*, 26, 2088–2112.

Papaspiliopoulos, O. (2008), "A note on posterior sampling from Dirichlet mixture models." Tech. rep., Universitat Pompeu Fabra.

Papaspiliopoulos, O. and Roberts, G. (2008), "Retrospective Markov chain Monte Carlo Methods for Dirichlet process hierarchical models," *Biometrika*, 95, 169–186.

Penny, W., Kilner, J., and Blankenburg, F. (2007), "Robust Bayesian general linear models," *NeuroImage*, 36, 661 – 671.

Petrone, S., Guindani, M., and Gelfand, A. E. (2009), "Hybrid Dirichlet mixture models for functional data," *Journal of Royal Statistical Society B*, 71, 755–782.

Pierpaoli, C. and Basser, P. J. (1996), "Toward a quantitative assessment of diffusion anisotropy," *Magn Reson Med*, 36, 893–906.

Rajan, J. J. and Rayner, P. J. W. (1996), "Generalized feature extraction for time-varying autoregressive models," *IEEE Transactions on Signal Processing*, 44, 2498–2507.

Ramsay, J. O. and Silverman, B. W. (1997), *Functional data analysis*, Springer, New York.

Reich, B. J., Fuentes, M., Herring, A. H., and Evenson, K. R. (2010), "Bayesian variable selection for multivariate spatially varying coefficient regression ," *Biometrics*, 66, 772–782.

Rice, J. A. and Silverman, B. W. (1991), "Estimating the mean and covariance structure nonparametrically when the data are curves," *J. Roy. Statist. Soc. B*, 53, 233243.

Rice, J. A. and Wu, C. O. (2001), "Nonparametric mixed effects models for unequally sampled noisy curves," *Biometrics*, 57, 253259.

Rollins, N. K. (2007), "Clinical applications of diffusion tensor imaging and tractography in children," *Pediatr Radiol*, 37, 769–780.

Rosa, M. J., Bestmann, S., Harrison, L., and Penny, W. (2010), "Bayesian model selection maps for group studies," *NeuroImage*, 49, 217 – 224.

Ruppert, D. (1997), "Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation," *Journal of the American Statistical Association*, 92, pp. 1049–1062.

Ruppert, D. and Wand, M. P. (1994), "Multivariate locally weighted least squares regression," *The Annals of Statistics*, 22, pp. 1346–1370.

Russell, J. A. and Carroll, J. M. (1999), "On the bipolarity of positive and negative affect," *Psychological Bulletin*, 125, 3–30.

Schwartzman, A. (2006), "Random Ellipsoids and False Discovery Rates: Statistics for Diffusion Tensor Imaging Data," *Ph.D. thesis, Stanford University*.

Schwartzman, A., Dougherty, R. F., and Taylor, J. E. (2005), "Cross-subject comparison of principal diffusion direction maps," *Magn Reson Med*, 53, 1423–1431.

Sethuraman, J. (1994), "A constructive definition of dirichlet priors," *Statistica Sinica*, 4, 639–650.

Shi, M. G., Weiss, R. E., and Taylor, J. M. G. (1996), "An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves," *J. Roy. Statist. Soc. C*, 45, 151–163.

Smith, M. and Kohn, R. (1996), "Nonparametric regression using Bayesian variable selection," *Journal of Econometrics*, 75, 317–343.

— (1998), "Nonparametric estimation of irregular functions with independent or autocorrelated errors," *In Practical Nonparametric and Semiparametric Bayesian Statistics. Lecture Notes in Statist.*, 133, 157–179.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M., Matthews, P., and Behrens, T. E. (2006), "Tractbased spatial statistics: voxelwise analysis of multi-subject diffusion data," *NeuroImage*, 31, 1487–1505.

Snook, L., Paulson, L. A., Roy, D., Phillips, L., and Beaulieu, C. (2005), "Diffusion tensor imaging of neurodevelopment in children and young adults," *NeuroImage*, 26, 1164–1173.

Snook, L., Plewes, C., and Beaulieu, C. (2007), "Voxel based versus region of interest analysis in diffusion tensor imaging of neurodevelopment," *NeuroImage*, 34, 243–252.

Song, S. K., Sun, S. W., Ju, W. K., Lin, S. J., Cross, A. H., and Neufeld, A. H. (2003), "Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia," *Neuroimage*, 20, 1714–1722.

Spellman, P., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M., Brown, P., Botstein, D., and Futcher, B. (1998), "Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," *Molecular Biology of Cell*, 9, 3273–3297.

Sun, Y., Zhang, W., and Tong, H. (2007), "Estimation of the covariance matrix of random effects in longitudinal studies," *Ann. Statist.*, 35, 27952814.

Taguchi, Y. and Oono, Y. (2005), "Relational patterns of gene expression via non-metric multidimensional scaling analysis," *Bioinformatics*, 21, 730–740.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006), "Hierarchical Dirichlet processes," *Journal Of The American Statistical Association*, 101, 1566–1581.

Thompson, W. K. and Rosen, O. (2008), "A Bayesian model for sparse functional data," *Biometrics*, 64, 54–63.

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox model with time-varying coefficients," *J Am Statist Assoc*, 100, 172183.

Torgerson, W. (1952), "Multidimensional scaling: I. Theory and method," *Psychometrika*, 17, 401–419, 10.1007/BF02288916.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002), "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magnetic Resonance in Medicine*, 48, 577–582.

Walker, S. G. (2007), "Sampling the Dirichlet mixture model with slices," *Communications in Statistics - Simulation and Computation*, 36, 45–54.

Wang, L., Chen, G., and Li, H. (2007), "Group SCAD regression analysis for microarray time course gene expression data," *Bioinformatics*, 23, 1486–1494.

Wang, L. and Dunson, D. (2010), "Semiparametric Bayes multiple testing: Applications to tumor data," *Biometrics*, 66, 493–501.

Wang, L., Li, H., and Huang, J. Z. (2008), "Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements," *J. Amer. Statist. Assoc.*, 103, 1556–1569.

Wang, Y. X. and Taylor, J. M. G. (1995), "Inference for smooth curves in longitudinal data with application to an aids clinical trial," *Statistics in Medicine*, 14, 1205–1218.

Watson, D. and Clark, L. A. (1997), "The measurement and mismeasurement of mood: Recurrent and emergent issues," *Journal of Personality Assessment*, 86, 267–296.

Watson, D., Clark, L. A., and Tellegen, A. (1988), "Development and validation of brief measures of positive and negative affect: The PANAS scales," *Journal of Personality and Social Psychology*, 54, 1063–1070.

Watson, D. and Tellegen, A. (1985), "Toward a consensual structure of mood," *Psychological Bulletin*, 98, 219–235.

Whitcher, B., Wisco, J. J., Hadjikhani, N., and Tuch, D. S. (2007), "Statistical group comparison of diffusion tensors via multivariate hypothesis testing," *Magnetic Resonance in Medicine*, 57, 1065–1074.

Wiggins, J. S., Steiger, J. H., and Gaelick, L. (1981), "Evaluating circumplexity models in personality data," *Multivariate Behavioral Research*, 16, 263289.

Winnett, A. and Sasieni, P. (2003), "Iterated residuals and time-varying covariate effects in Cox regression," *J R Statist Soc B*, 65, 473488.

Woolrich, M., Jenkinson, M., Brady, J., and Smith, S. (2004), "Fully bayesian spatiotemporal modeling of fMRI data," *IEEE Trans. Med. Imaging*, 23, 213 – 231.

Wu, C. O., Chiang, C. T., and Hoover, D. R. (1998), "Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data," *J. Amer. Statist. Assoc.*, 93, 13881402.

Wu, H. L. and Liang, H. (2004), "Backfitting random varying-coefficient models with timedependent smoothing covariates," *Scandinavian Journal of Statistics*, 31, 3–19.

Wu, H. L. and Zhang, J. T. (2002), "Local polynomial mixed-effects models for longitudinal data," *J. Amer. Statist. Assoc.*, 97, 883–897.

— (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis*, Hoboken, New Jersey.: John Wiley & Sons, Inc.

Xia, Y. and Li, W. K. (1999), "On the estimation and testing of functional-coefficient linear models," *Statistica Sinica*, 9, 735758.

Yu, Z. and Lin, X. (2008), "Nonparametric regression using local kernel estimating equations for correlated failure time data," *Biometrika*, 95, 123137.

Yu, Z. S. and Lin, X. H. (2010), "Semiparametric regression with time-dependent coefficients for failure time data analysis," *Statistica Sinica*, 20, 853–869.

Yushkevich, P. A., Zhang, H., Simon, T., and Gee, J. C. (2008), "Structure-specific statistical mapping ofwhite matter tracts," *Neuroimage*, 41, 448–461.

Zhang, W. and Lee, S. Y. (2000), "Variable bandwidth selection in varying-coefficient models," *Journal of Multivariate Analysis*, 74, 116134.

Zhang, W., Lee, S. Y., and Song, X. (2002), "Local polynomial fitting in semivarying coefficient models," *J. Multivar. Anal.*, 82, 166188.

Zhang, W. and Steele, F. (2004), "A semiparametric multilevel survival model," *Journal of the Royal Statistical Society C*, 53, 387404.

Zhu, H., Kong, L., Li, R., Styner, M., Gerig, G., Lin, W., and Gilmore, J. H. (2011), "FADTTS: Functional analysis of diffusion tensor tract statistics," *NeuroImage*, 56, 1412 – 1425.

Zhu, H. T., Cheng, Y. S., Ibrahim, J. G., Li, Y. M., Hall, C., and Lin, W. L. (2009), "Intrinsic regression models for positive definitive matrices with applications in diffusion tensor images," *Journal of the American Statistical Association*, 104, 1203–1212.

Zhu, H. T., Styner, M., Tang, N. S., Liu, Z. X., Lin, W. L., and Gilmore, J. (2010), "FRATS: Functional regression analysis of DTI tract statistics," *IEEE Transactions on Medical Imaging*, 29, 1039–1049.

Zhu, H. T., Xu, D., Amir, R., Hao, X., Zhang, H., Alayar, K., Ravi, B., and Peterson, B. (2006), "A statistical framework for the classification of tensor morphologies in diffusion tensor images," *Magnetic Resonance Imaging*, 24, 569–582.

Zhu, H. T., Zhang, H. P., Ibrahim, J. G., and Peterson, B. G. (2007), "Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data (with discussion)," *Journal of the American Statistical Association*, 102, 1085–1102.

Zucker, D. M. and Karr, A. F. (1990), "Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach," *The Ann Statist*, 18, 329353.