Published in final edited form as:

Stat Biosci. 2015 May 1; 7(1): 90–107. doi:10.1007/s12561-013-9104-y.

Semiparametric Bayes local additive models for longitudinal data

Zhaowei Hua^a, Hongtu Zhu^a, and David B. Dunson^b

Zhaowei Hua: zhua@email.unc.edu; Hongtu Zhu: hzhu@bios.unc.edu; David B. Dunson: dunson@stat.duke.edu

^aDepartment of Biostatistics, University of North Carolina at Chapel Hill

^bDepartment of Statistical Science, Duke University

Abstract

In longitudinal data analysis, there is great interest in assessing the impact of predictors on the time-varying trajectory in a response variable. In such settings, an important issue is to account for heterogeneity in the shape of the trajectory among subjects, while allowing the impact of the predictors to vary across subjects. We propose a flexible semiparametric Bayes approach for addressing this issue relying on a local partition process prior, which allows flexible local borrowing of information across subjects. Local hypothesis testing and credible bands are developed for the identification of time windows across which a predictor has a significant impact, while adjusting for multiple comparisons. Posterior computation proceeds via an efficient MCMC algorithm using the exact block Gibbs sampler. The methods are assessed using simulation studies and applied to a yeast cell-cycle gene expression data set.

Keywords

Confidence band; Functional data; Gaussian process; Local partition process; Random effects; Time-varying coefficients

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 *and others* (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:

7 Supplementary Material

The reader is referred to the Supplementary Materials for the Appendix A for the proof of Proposition 1 in Subsection 2.3, the Appendix B, the Supplementary Table, and the Supplementary Figures, referenced in Section 3, Section 4, and Subsection 5.2.

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

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 $\beta_i(t_{ij}) = (\beta_{i1}(t_{ij}), \dots, \beta_{iK}(t_{ij}))'$ 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 (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.

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 *and others* (2007) adopted an analysis-of-variance dependent DP (De Iorio *and others*, 2004) for the problem of sequential classification analysis. Ohlssen *and others* (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 model (1). The DP prior assumes the random effects distribution $P \sim \text{DP}(\alpha P_0)$, 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 occupied clusters 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 *and others*, 2005). A hybrid functional Dirichlet process (hFDP) (Petrone *and others*, 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 $\beta_{ik}(t)$, $k = 1, \dots, K$.

The motivating application for this article 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 *and others*, 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 *and others*, 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 in 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 coefficients 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.

There are three primary goals of this article. The first applies the LPP prior to the random effects distributions in the random time-varying coefficient model (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 when the predictor has a significant impact on an outcome trajectory. Multiplicity issues arising from pointwise inference is accounted for by controlling false discovery rate (FDR).

2 Semiparametric Bayes Local Additive Model

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 $\Theta_{gh} = (\Theta_{g,h,1}, \dots, \Theta_{g,h,p})'$, $g = 0, 1$, $h = 1, \dots, \infty$, 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 $z = (z_1, \dots, z_p)'$ for the allocation to

global clustering or local clustering, with $z_j \sim \text{Ber}(v_j)$ (Ber: Bernoulli) independently, taking the value of 1 for global clustering and 0 for local clustering, $j = 1, \dots, p$. Define a $p \times 2$

local cluster indices matrix $\psi = (\psi_1', \dots, \psi_p')'$, with $\psi_j = (1 - z_j, \varphi_j)$ and $\varphi_j \in \{1, \dots, \infty\}$, $j = 1, \dots, p$. A corresponding hybrid atom $\Theta_\psi = (\Theta_{\psi_1,1}, \dots, \Theta_{\psi_p,p})'$ is obtained by setting the j th element of Θ_ψ equal to $\Theta_{\psi_j,j}$.

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

$$P = \sum_{z_1=0}^1 \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_p=0}^1 \sum_{\psi_p=(1-z_p,1)}^{(1-z_p,\infty)} \pi_{\psi_1,\dots,\psi_p} \delta_{\Theta_\psi}, \quad (2)$$

where δ_x denotes a degenerate distribution with all its mass at x and $\pi_{\psi_1, \dots, \psi_p}$ is the

$$\text{probability of } \theta_i = \Theta_\psi \text{ with } \sum_{z_1=0}^1 \sum_{\psi_1=(1-z_1,1)}^{(1-z_1,\infty)} \cdots \sum_{z_p=0}^1 \sum_{\psi_p=(1-z_p,1)}^{(1-z_p,\infty)} \pi_{\psi_1,\dots,\psi_p} = 1.$$

Let π_h denote $\Pr(\theta_{ij} = \Theta_{g,h,j}), g = 0, 1, h = 1, \dots, \infty, j = 1, \dots, p$. 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 ,

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

The allocation probability of $\theta_i = \Theta_\psi$ in (2) is then simply

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

The specification is completed by choosing the hyperpriors

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

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 (2) 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

$\Pr(z_1, \dots, z_p) = \prod_{j=1}^p \nu_j^{z_j} (1 - \nu_j)^{1-z_j}$. The joint cluster allocation weight at stage 2 conditional on stage 1 corresponds to $\Pr(\psi_1 = (1 - z_1, h_1), \dots, \psi_p = (1 - z_p, h_p) | z_1, \dots, z_p)$. The overall joint cluster allocation weight corresponds to $\pi_{\psi_1, \dots, \psi_p}$.

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 subject i , we rewrite model (1) as

$$Y_i(t) = \beta_{i0}(t) + \sum_{k=1}^K \beta_{ik}(t) x_{ik} + \varepsilon_i(t), \quad \varepsilon_i(t) \sim N(0, \tau^{-1}), \quad i=1, \dots, n, \quad (3)$$

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

Similar to Huang *and others* (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:

$$\beta_{ik}(t) = \sum_{l=1}^{p_k} \theta_{ikl} b_{kl}(t) = b_k(t)' \theta_{ik}, \quad (4)$$

where $\theta_{ik} = (\theta_{ik1}, \dots, \theta_{ikp_k})'$ and the basis functions $b_k(t) = (b_{k1}(t), \dots, b_{kp_k}(t))'$. Various basis systems can be applied, such as Fourier, wavelets and B-splines. For computational convenience, we let $b_k(t) \equiv b(t) = (b_1(t), \dots, b_p(t))'$ and $p_k \equiv p$ for all k . We specify Gaussian kernels with $b_1(t) = 1$, $b_{j+1}(t) = \exp(-\zeta \|t - \eta_j\|^2)$, $j = 1, \dots, p-1$, where $\eta_1, \dots, \eta_{p-1}$ are equally spaced kernel locations and $\zeta = 25$. Letting $\theta_i = (\theta'_{i1}, \dots, \theta'_{iK})'$ and $B_i(t) = (b(t)'x_{i1}, \dots, b(t)'x_{iK})'$, model (3) can be written as

$$Y_i(t) = b(t)' \theta_{i0} + B_i(t)' \theta_i + \varepsilon_i(t). \quad (5)$$

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:

$$P \sim \text{LPP2}(\alpha_0, \gamma_0, P_0) \\ \tilde{P} \sim \text{LPP2}(\alpha_1, \gamma_1, \otimes \{P_{0k}\}_{k=1}^K), \quad P_{0k} = P_0, \quad (6)$$

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 allow automatic selection from among a large number of pre-specified potential basis functions, while accommodating 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:

$$P_0: \theta_{ik} \sim N_p(0_{p \times 1}, \Lambda^{-1}), \quad i=1, \dots, n, \quad k=0, 1, \dots, K. \quad (7)$$

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ with $\lambda_l \sim G(0.5, 0.5)$. To allow borrowing information across coefficients in 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.

Although we assume a simple diagonal covariance structure in the base measure P_0 , which facilitates efficient computation and shrinkage towards a lower dimensional structure, it is important to note that the random effects distribution P is given a nonparametric prior that can accommodate dependence. In addition, it is trivial to modify the implementation to accommodate different choices of basis functions (e.g., Fourier) and non-diagonal Λ , as we briefly allude to in discussing properties and computation later in the text.

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 (3) – (7).

Considering a generic unknown function

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

with $t \in [0, T]$, $\vartheta = (\vartheta_1, \dots, \vartheta_p)'$ and assuming $\vartheta \sim P$, $P \sim \text{LPP2}(\alpha, \gamma, P_0)$, with P_0 defined as in (7), the induced prior presents the following properties.

Proposition 2.1—Conditional on Λ^{-1} , with Λ a $p \times p$ (potentially non-diagonal) covariance matrix,

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

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

Further,

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

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

$\Lambda^{-1} = \text{diag}\{\lambda_1^{-1}, \dots, \lambda_p^{-1}\}$ and $\lambda_1 = \dots = \lambda_p = \lambda$,

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

which depends only on the set of basis functions and not the base measure P_0 . Furthermore, 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. Otherwise, a long-range dependency structure will be realized with the general non-diagonal Λ or non-local bases such as 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.

2.4 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 k th predictor effect specific to a time point $t \in [0, T]$ are formulated as:

$$H_{0k}(t): |\bar{\beta}_k(t)| \leq \varepsilon, \text{ versus } H_{1k}(t): |\bar{\beta}_k(t)| > \varepsilon,$$

where $\bar{\beta}_k(t) = n^{-1} \sum_{i=1}^n \beta_{ik}(t)$ represents the sample mean of the subject-specific time-varying coefficients for the k th predictor at time t . To test whether the average time-varying

coefficient curve for the k th predictor deviates from a small neighborhood of zero anywhere in time interval $[0, T]$, the global null and alternative hypotheses are formulated as

$$\begin{aligned} H_{0k}: |\bar{\beta}_k(t)| &\leq \varepsilon, \text{ for all } t \in [0, T] \\ H_{1k}: |\bar{\beta}_k(t)| &> \varepsilon, \text{ for any } t \in [0, T]. \end{aligned}$$

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 *and others* (2004). Following the implementation in Wang and Dunson (2010), our strategy is to reject $H_{0k}(t)$ if the posterior alternative hypothesis probability $v_{kt} = \Pr(H_{1k}(t)|Data) > r$ 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 $d_{kt} = 1(v_{kt} > r)$, an indicator of rejecting $H_{0k}(t)$. The posterior expected FNR and FDR are calculated as:

$$\begin{aligned} \overline{FNR} &= \frac{\sum_{k=1}^K \int_0^T (1-d_{kt})v_{kt} dt}{K \int_0^T dt - \sum_{k=1}^K \int_0^T d_{kt} dt + \kappa_0} \approx \frac{\sum_{k=1}^K \sum_{t=t_1}^{t_m} (1-d_{kt})v_{kt} T/m}{KT - \sum_{k=1}^K \sum_{t=t_1}^{t_m} d_{kt} T/m + \kappa_0}, \\ \overline{FDR} &= \frac{\sum_{k=1}^K \int_0^T d_{kt}(1-v_{kt}) dt}{\sum_{k=1}^K \int_0^T d_{kt} dt + \kappa_0} \approx \frac{\sum_{k=1}^K \sum_{t=t_1}^{t_m} d_{kt}(1-v_{kt}) T/m}{\sum_{k=1}^K \sum_{t=t_1}^{t_m} d_{kt} T/m + \kappa_0}, \end{aligned}$$

where $t_1, \dots, t_m$ 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 $r^* = \operatorname{argmin}_r \overline{FNR}\{r \in [0, 1], \overline{FDR} \leq \alpha_T\}$. For each combination of k and $t \in \{t_1, \dots, t_m\}$, we reject $H_{0k}(t)$ if $v_{kt} > r^*$. For a sufficiently fine grid of $m = 10$, the results are robust to m , with the optimal threshold appropriately adapting to the chosen m .

2.5 Bayesian Confidence Band

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

The most commonly-used method is to compute a confidence band based on pointwise credible intervals, which we denote as CR. The constructed confidence band is obtained as $[\beta_k^L(t_l), \beta_k^U(t_l)]$, $l = 1, \dots, L$, 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 *and others* (2007) assumes approximate posterior normality and derives the $1 - \alpha$ sample percentile $c_{1-\alpha}$ of

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

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

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

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 $s_{1-\alpha/2}(t_l)$, $l = 1, \dots, L$. By deriving the maximal deviations $s_{\alpha/2} = \max_{l=1, \dots, L} (\hat{\beta}_k(t_l) - s_{\alpha/2}(t_l))$ and $s_{1-\alpha/2} = \max_{l=1, \dots, L} (s_{1-\alpha/2}(t_l) - \hat{\beta}_k(t_l))$ 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

$$\max_{l=1, \dots, L} |\beta_k^s(t_l) - \hat{\beta}_k(t_l)| \quad s=1, \dots, S,$$

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

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

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 and Roberts (2008) and the slice sampling method of Walker (2007). Allowing the data to provide information about the hyperparameters involved in the model through (3)–(7), we specify $a_0 \sim \text{Ga}(a_\alpha, b_\alpha)$, $\gamma_0 \sim \text{Ga}(a_\gamma, b_\gamma)$, $\alpha_1 \sim \text{Ga}(a_\alpha, b_\alpha)$, $\gamma_1 \sim \text{Ga}(a_\gamma, b_\gamma)$, $\tau \sim \text{Ga}(a_\tau, b_\tau)$. The gamma distribution in this article is parameterized by a shape parameter (e.g. a_α) and a rate parameter (e.g. b_α). Introducing auxiliary variables $\{u_{ikl}\}$ to avoid truncated approximations, $k = 0, \dots, K$, $l = 0, \dots, p$, $i = 1, \dots, n$, the complete data joint likelihood of y , u , and z is:

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

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 , $i = 1, \dots, n$, and the $\{u_{ikl}\}$ s are constrained to the interval (0, 1).

Please refer to the appendix B of the supplementary material for the sampling algorithm in detail.

This algorithm is straightforward to implement and exhibits good performance in convergence and mixing. To perform hypothesis testing, we estimate the posterior probability in favor of the alternative hypothesis from the MCMC posterior samples. The posterior sample of $\hat{\beta}_k^s(t)$, $t \in [0, T]$, is obtained by averaging the subject-specific coefficients a posteriori $\hat{\beta}_{ik}^s(t)$ at each iteration $s = 1, \dots, S$ after burn-in,

$\hat{\beta}_k^s(t) = 1/n \sum_{i=1}^n \hat{\beta}_{ik}^s(t)$. It is straightforward that $\hat{\beta}_k^s(t)$ is a direct draw from the posterior conditional distribution of the sample mean $\bar{\beta}_k(t)$. For any $t \in [0, T]$, $\Pr(H_{1k}(t)/\text{Data})$ is estimated as the proportion of iterations in the posterior sample satisfying $\hat{\beta}_k^s(t) \notin [-\varepsilon, \varepsilon]$.

The posterior sample average $\hat{\beta}_k(t) = 1/S \sum_{s=1}^S \hat{\beta}_k^s(t)$ is calculated to estimate the sample mean $\bar{\beta}_k(t)$.

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:

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

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 $\text{Unif}(0, 1)$ and then standardized to have mean 0 and standard deviation 1. The baseline $\beta_{i0}(t)$ for $t \in [0, 1]$, $i = 1, \dots, 100$ 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 equal to $c_0(t, t')$.

We analyze each simulated scenario using 100 repeated simulations by our method. We set $\text{Ga}(1, 1)$ hyperpriors for α_0 , γ_0 , α_1 , and γ_1 , and a $\text{Ga}(0.1, 0.1)$ 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 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. Supplementary Table 1 of the supplementary material 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 $\Pr(H_{1k}(t)/\text{Data}) \leq r$, where r is optimized to minimize FNR under the constraint of $\text{FDR} \leq 0.05$. 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, 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 3, where the coverage rates are calculated as the proportion of confidence bands containing true mean curves. As seen in Table 3, 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.

5 Yeast Cell-Cycle Gene Expression Application

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, 1985), 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 *and others*, 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 take 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 *and others* (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 *and others*, 2007; Wang *and others*, 2007) to the ChIP data of Lee *and others* (2002).

Our first analysis goal is to apply the random effects time-varying coefficient model (3) – (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 $k = 1, \dots, 96$ and $i = 1, \dots, 297$. Wang *and others* (2008) linked the binding probabilities to the gene expression levels using a fixed effect nonparametric time-varying coefficient model:

$$y_i(t) = \beta_0(t) + \sum_{k=1}^{96} \beta_k(t) x_{ik} + \varepsilon_i(t), \quad (8)$$

where $\beta_k(t)$ is the mean transcription effect of the k -th TF on the log expression level of the i -th gene at time point t , and $\varepsilon_i(t)$'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 *and others*, 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 :

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

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 also 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 the cell-cycle regulated gene expression network and profile their active/inactive phases. Wang *and others* (2008) identified the significant TFs through zeroing small regression coefficients using the SCAD penalty. However, this sparse estimation approach does not represent a formal framework for hypothesis testing. 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 DP priors 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 deviations, we expect that the LPP2 prior will produce a more parsimonious representation of the log gene expression trajectories.

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 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 (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 $\varphi_0^*=1$ 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 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 favoring our approach over the Dirichlet process.

Wang *and others* (2008) used estimated time-varying effect curves and confidence bands 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} , $k = 1, \dots, 96$ if any $v_{kt} \geq 0.886$. Figure 1 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 *and others* (2008). For example, Wang *and others* (2008) didn't identify the fact that FKH2 is associated with genes expressed in G1 and S (Simon *and others*, 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 *and others*, 2001). Although Wang *and others* (2008) found approximately similar peak effect time points for SWI4 and SWI6, Figure 1 clearly displays that they have similar active time windows.

We use our proposed Mdiff method to construct 95% Bayesian credible bands for the time-varying transcription effects. As presented in Supplementary Figure 1 of the supplementary material, MBP1, SWI4, SWI6, MCM1, NDD1, FKH2, ACE2, and SWI5 appear to have periodic transcriptional effect patterns over the collected two cell cycles, while Wang *and others* (2008) only showed certain periodic patterns for SWI4 and FKH2. Since the estimated confidence bands are wide, it is not possible to make definitive conclusions about specific time patterns. However, the frequentist coverage of our Mdiff simultaneous credible bands, reported in the simulation studies in Table 3, was close to the nominal 95% level. Hence, we do not expect that our intervals are overly-wide and instead it seems that there is not sufficient information in these data to make definitive inferences on specific time trajectories. However, we have obtained interesting results for the hypothesis testing methods, with the estimated curves being suggestive of patterns to be followed up in future studies.

We repeated the analysis using DP priors for $P \sim DP(\alpha_0 P_0)$ and $P \sim \tilde{DP}(\alpha_1 P_0)$. 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 Supplementary Figure 3 of the supplementary material. This is likely due to the tendency of the Dirichlet process to overly favor clustering of subjects.

6 Discussion

This article 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Hongzhe Li for providing the yeast cell-cycle gene data set and the ChIP binding data set.

9 Funding

This work was supported in part by NSF grant BCS-08-26844 and NIH grants RR025747-01, P01CA142538-01, MH086633, and AG033387 to Dr. Zhu, and T ES17240 to Dr. Dunson. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or the NIH.

References

- Behseta, Sam; Kass, Robert E.; Wallstrom, Garrick L. Hierarchical models for assessing variability among functions. *Biometrika*. 2005; 92:419–434.
- Botts, Carsten H.; Daniels, Michael J. A flexible approach to Bayesian multiple curve fitting. *Computational Statistics & Data Analysis*. 2008; 52:5100–5120. [PubMed: 21127724]
- Bush CA, MacEachern SN. A semiparametric Bayesian model for randomised block designs. *Biometrika*. 1996; 83:275–285.
- Chen G, Jensen S, Stoeckert C. Clustering of Genes Into Regulons Using Integrated Modeling (CORIM). *Genome Biology*. 2007; 8(1):R4. [PubMed: 17204163]
- Crainiceanu, Ciprian M.; Ruppert, David; Carroll, Raymond J.; Joshi, Adarsh; Goodner, Billy. Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*. 2007; 16:265–288.
- Cruz-Mesía RDL, Fernando AQ, Müller Peter. Semiparametric Bayesian classification with longitudinal markers. *Journal of Royal Statistical Society C*. 2007; 56:119–137.
- De Iorio M, Müller P, Rosner GL, MacEachern SN. An ANOVA model for dependent random measures. *J Amer Statist Assoc*. 2004; 99:205–215.
- Dunson DB. Nonparametric Bayes local partition models for random effects. *Biometrika*. 2009; 96:249–262. [PubMed: 23710074]

- Escobar MD, West M. Bayesian density estimation and inference using mixtures. *J Amer Statist Assoc.* 1995; 90:577–588.
- Fan J, Zhang W. Statistical methods with varying coefficient models. *Statistics and Its Interface.* 2008; 1:179–195. [PubMed: 18978950]
- Ferguson, Thomas S. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics.* 1973; 1:209–230.
- Ferguson, Thomas S. Prior distributions on spaces of probability measures. *The Annals of Statistics.* 1974; 2:615–629.
- Hastie TJ, Tibshirani RJ. Varying-coefficient models. *J Roy Statist Soc B.* 1993; 55:757–796.
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika.* 1998; 85:809–822.
- Huang JZ, Wu CO, Zhou L. Varying-coefficient models and basis function approximation for the analysis of repeated measurements. *Biometrika.* 2002; 89:111–128.
- Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. *Biometrics.* 1998; 54:921–938. [PubMed: 9750242]
- Lee D, Shaddick G. Time-varying coefficient models for the analysis of air pollution and health outcome data. *Biometrics.* 2007; 63:1253–1261. [PubMed: 17425638]
- Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al. Transcriptional Regulatory Networks in *S. cerevisiae*. *Science.* 2002; 298:799–804. [PubMed: 12399584]
- Lin DY, Ying Z. Semiparametric and nonparametric regression analysis of longitudinal data (with discussions). *J Amer Statist Assoc.* 2001; 96:103–126.
- LOAY. On a class of Bayesian nonparametric estimates. 1. Density estimates. *The Annals of Statistics.* 1984; 12:351–357.
- Luan YH, Li HZ. Clustering of time-course gene expression data using a mixed-effects model with B-splines. *Bioinformatics.* 2003; 19:474–482. [PubMed: 12611802]
- Morris JS, Carroll RJ. Wavelet-based functional mixed models. *Journal of Royal Statistical Society B.* 2006; 68:179–199.
- Müller P, Parmigiani G, Robert C, Rousseau J. Optimal sample size for multiple testing: The case of gene expression microarrays. *Journal of the American Statistical Association.* 2004; 99:990–1001.
- Müller P, Rosner GL. A Bayesian population model with hierarchical mixture priors applied to blood count data. *J Amer Statist Assoc.* 1997; 92:1279–1292.
- Nasmyth K. At least 1400 base pairs of 59-flanking DNA is required for the correct expression of the HO gene in yeast. *Cell.* 1985; 42:213–223. [PubMed: 3893741]
- Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons. *Statistics in Medicine.* 2007; 26:2088–2112. [PubMed: 16906554]
- Papaspiliopoulos, O. Technical Report. Universitat Pompeu Fabra; 2008. A note on posterior sampling from Dirichlet mixture models.
- Papaspiliopoulos O, Roberts G. Retrospective Markov chain Monte Carlo Methods for Dirichlet process hierarchical models. *Biometrika.* 2008; 95:169–186.
- Petrone, Sonia; Guindani, Michele; Gelfand, Alan E. Hybrid Dirichlet mixture models for functional data. *Journal of Royal Statistical Society B.* 2009; 71:755–782.
- Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakola TS, et al. Serial regulation of transcriptional regulators in the Yeast cell cycle. *Cell.* 2001; 106:697–708. [PubMed: 11572776]
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast *saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of Cell.* 1998; 9:3273–3297.
- Thompson, Wesley K.; Rosen, Ori. A Bayesian model for sparse functional data. *Biometrics.* 2008; 64:54–63. [PubMed: 17573864]
- Walker SG. Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation.* 2007; 36:45–54.

- Wang, Lifeng; Chen, Guang; Li, Hongzhe. Group scad regression analysis for microarray time course gene expression data. *Bioinformatics*. 2007; 23:1486–1494. [PubMed: 17463025]
- Wang LM, Dunson DB. Semiparametric Bayes multiple testing: Applications to tumor data. *Biometrics*. 2010; 66:493–501. [PubMed: 19673866]
- Wang, Lifeng; Li, Hongzhe; Huang, Jianhua Z. Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *J Amer Statist Assoc*. 2008; 103:1556–1569.
- Wu, HL.; Zhang, JT. *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc; 2006.

Figure 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.

Table 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.

	LLP2		DP		Fixed Effects	
	$\beta_0(t)$	$\beta_1(t)$	$\beta_0(t)$	$\beta_1(t)$	$\beta_0(t)$	$\beta_1(t)$
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

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 3

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