Bayesian Analysis (2023+) **TBA**, Number TBA, pp. 1

Functional Concurrent Regression Mixture Models Using Spiked Ewens-Pitman Attraction Priors

Mingrui Liang*, Matthew D. Koslovsky†, Emily T. Hébert‡, Michael S. Businelle§, and Marina Vannucci*

Abstract. Functional concurrent, or varying-coefficient, regression models are a form of functional data analysis methods in which functional covariates and outcomes are collected concurrently. Two active areas of research for this class of models are identifying influential functional covariates and clustering their relations across observations. In various applications, researchers have applied and developed methods to address these objectives separately. However, no approach currently performs both tasks simultaneously. In this paper, we propose a fully Bayesian functional concurrent regression mixture model that simultaneously performs functional variable selection and clustering for subject-specific trajectories. Our approach introduces a novel spiked Ewens-Pitman attraction prior that identifies and clusters subjects' trajectories marginally for each functional covariate while using similarities in subjects' auxiliary covariate patterns to inform clustering allocation. Using simulated data, we evaluate the clustering, variable selection, and parameter estimation performance of our approach and compare its performance with alternative spiked processes. We then apply our method to functional data collected in a novel, smartphone-based smoking cessation intervention study to investigate individual-level dynamic relations between smoking behaviors and potential risk factors.

Keywords: clustering, Ewens-Pitman Attraction distribution, functional data analysis, spiked nonparametric priors, variable selection.

1 Introduction

Functional data analysis (FDA) methods have become increasingly popular in recent years for modeling data in which covariates, outcomes, or both have functional forms (i.e., scalar-on-function, function-on-scalar, or function-on-function regression, respectively) (Morris, 2015; Reiss et al., 2017). For example, James et al. (2009) built a scalar-on-function regression model to predict annual rainfall, a scalar response, based

arXiv: 2010.00000

*Department of Statistics, Rice University, Houston, TX, USA. ml86@rice.edu marina@rice.edu

†Department of Statistics, Colorado State University, Fort Collins, CO, USA. matt.koslovsky@colostate.edu

‡ Department of Health Promotion and Behavioral Sciences, School of Public Health, University of Texas Health Science Center, Austin, TX 78701, USA emily.t.hebert@uth.tmc.edu

§Department of Family and Preventive Medicine, College of Medicine, University of Oklahoma, Oklahoma City, OK 73104, USA Michael-Businelle@ouhsc.edu

on daily temperature measurements, a functional covariate. Goldsmith et al. (2015) applied function-on-scalar regression methods to predict activity level trajectories given a subject's BMI and age. Zhang et al. (2011) employed function-on-function regression models to investigate relations between forest nitrogen cycling and deforestation using spatial imagery. The special case of function-on-function regression models in which the functional response and functional predictors are collected at the same time is often referred to as a functional concurrent, or varying-coefficient, regression model (Zhang et al., 2011; Maity, 2017). This framework is commonly used to investigate how the relation between an observed covariate and the outcome varies as a function of another covariate, typically time (Tan et al., 2012; Kim et al., 2018; Leroux et al., 2018).

One of the main research objectives in FDA studies is to cluster groups of observations that share similar functional trends into homogeneous subsets (Tarpey and Kinateder, 2003; Ferreira and Hitchcock, 2009). For example, functional clustering methods have been applied in various fields, including genomics (Ma et al., 2008), neuroimaging (Zhang et al., 2014), environmental health (White and Gelfand, 2020), and population demographics (Hu et al., 2020). In the frequentist setting, functional clustering techniques typically belong to one of three categories (Jacques and Preda, 2014). The first is referred to as filtering methods, in which smooth functions are first estimated with finite approximations using regression techniques and the corresponding coefficients are clustered thereafter. For example, Abraham et al. (2003) applied the k-means algorithm to cluster estimated B-spline coefficients. Alternatively, distance-based methods adapt common geometric clustering algorithms, such as k-means, to cluster based on distances or dissimilarities between curves. The third approach is referred to as adaptive, or model-based, and comprises methods in which coefficients from finite approximations of the smooth functions are assumed to follow a cluster-specific probability distribution.

In Bayesian settings, model-based clustering methods are widely used to perform functional clustering. One of the advantages of the Bayesian approach is the ability to relax assumptions regarding the number of clusters in the data using nonparametric prior specifications. Species sampling priors, such as the Dirichlet process (Ferguson, 1973) and its generalization, the Pitman-Yor process (Pitman and Yor, 1997), are commonly used to govern clustering allocation due to their computational simplicity (Ray and Mallick, 2006; Wiesenfarth et al., 2014; Suarez and Ghosal, 2016; White and Gelfand, 2020; Das et al., 2021). The Dirichlet and Pitman-Yor processes are exchangeable processes, meaning that clustering allocation only depends on the cardinality of the realized partition. Alternatively, nonexchangeable nonparametric prior specifications allow researchers to inform clustering allocation using auxiliary information (Blei and Frazier, 2011; Dahl et al., 2017; Quintana et al., 2020). For example, Jin et al. (2021) used a distance-dependent Chinese restaurant process to cluster heterogeneous populations while considering individuals' treatment histories. While model-based functional clustering methods are equipped to easily incorporate additional population- and/or subject-level (functional) covariates, clustering is typically only based on one functional term. Notably, there are a variety of Bayesian nonparametric mixture models that cluster over the entire linear predictor (Dunson et al., 2008b; Kim et al., 2009; Barcella et al., 2016; Ding and Karabatsos, 2021), or together with other parameters in a joint model

(Bigelow and Dunson, 2009), but to our knowledge there are no methods designed to cluster over multiple smooth functions.

In addition to clustering, functional variable selection plays an important role in FDA to induce sparsity and improve interpretability of the model (Ghosal et al., 2020). To perform functional variable selection, group-based selection procedures are typically applied to the parameters corresponding to the finite approximation of each smooth function. In frequentist settings, researchers employ penalized optimization routines to shrink corresponding regression coefficients to zero (e.g., group-LASSO (Yuan and Lin, 2006), smoothly clipped absolute deviations (Wang et al., 2008), or minimax concave penalty (Ma et al., 2011)) (Gertheiss et al., 2013; Islam et al., 2018; Ghosal et al., 2020). In the Bayesian framework, researchers have imposed sparsity-inducing priors such as the spike-and-slab (Scheipl et al., 2012; Goldsmith and Schwartz, 2017) or shrinkage priors such as the functional horseshoe (Shin et al., 2020) to perform functional variable selection in regression settings.

Over the last two decades, researchers have innovatively employed sparsity-inducing and clustering priors in various ways to deliver novel insights from complex data structures. For example, Tadesse et al. (2005) used variable selection techniques to differentiate between covariates that are cluster-specific or shared by all clusters in a nonparameteric mixture model setting. Relatedly, Ding and Karabatsos (2021) applied horseshoe priors to shrink non-influential, cluster-specific covariates within Dirichlet process mixture models. More recently, researchers have designed and implemented spiked process priors that simultaneously perform variable selection and clustering of active terms (Dunson et al., 2008a; Kim et al., 2009; Savitsky and Vannucci, 2010; Canale et al., 2017; Cassese et al., 2019; Koslovsky et al., 2020). These priors can be constructed in two different ways, as described in Cassese et al. (2019). In one, a nonparametric distribution is assigned to the slab component of the traditional spike-and-slab prior, leading to an "outer" formulation. Alternatively, in the "inner" formulation, a spike-and-slab or non-diffuse distribution is assumed as the base measure of a nonparametric prior. The latter is considered less-informative and more robust to prior misspecification (Canale et al., 2017). The versatility of this class of priors has enabled researchers to design novel methods, ranging from identifying risk factors that share similar effects (Savitsky and Vannucci, 2010; Koslovsky et al., 2020) to dynamic variable selection techniques (Cassese et al., 2019). Surprisingly, spiked process priors have rarely been implemented in functional regression settings to simultaneously identify and cluster active smooth functions.

In this paper, we develop a Bayesian functional concurrent regression mixture model that uses novel spiked Ewens-Pitman attraction (spiked-EPA) distribution priors which exploit pairwise similarities in subjects' auxiliary covariate information to inform functional selection and clustering allocation. As such, we provide a comprehensive model-based functional clustering method that clusters multiple covariates' time-varying effects across subjects. By clustering subjects marginally for each smooth function, our approach flexibly allows different clustering patterns for each functional covariate's relation with the functional outcome. We demonstrate the selection, clustering, and estimation performance of our proposed method and compare to alternative methods on

simulated data. We then apply the proposed approach to intensive longitudinal mobile health (mHealth) data collected in a smartphone-based smoking cessation intervention study (Smart-T2, (Hébert et al., 2020)). Current FDA methods designed to identify and reaffirm relations between risk factors and behavioral outcomes in mHealth data suffer from several limitations, as they typically only identify population-level effects and ignore potential subpopulation trends. Our approach differs in that it is able to identify participant-level risk profile trajectories and cluster the trajectories marginally to help discover subgroups of participants using baseline characteristics.

The rest of the paper is organized as follows. In section 2, we introduce our Bayesian functional concurrent regression model and the spiked-EPA prior. In section 3, we compare the selection, clustering, and estimation performance of our proposed model with alternative mixture models on simulated data and perform a sensitivity analysis. In section 4, we apply our model to the Smart-T2 data. In section 5, we provide concluding remarks.

2 Functional Concurrent Regression Mixture Models

Motivated by momentary smoking behavior data collected in the Smart-T2 study, we develop a binary functional concurrent regression mixture model using a spiked-EPA prior which enables simultaneous identification and clustering of active subject-specific functional trends. We first present the functional concurrent regression model and finite approximation for the smooth functions. Then, we introduce our prior specification and discuss posterior inference.

Let $Y_i(\cdot) \in \{0, 1\}$ represent the binary functional outcome (e.g., momentary smoking status) for participant $i = 1, \ldots, N$, and $X_{ip}(\cdot)$, $p = 1, \ldots, P$, represent the P corresponding functional covariates collected on each participant. We assume the outcome and covariates are concurrently observed at subject-specific measurement times, t_{ij}, where $j = 1, \ldots, n_i$. Note that the observation times and the number of observations may vary between participants. To model the binary outcome, we assume a logistic regression modeling framework

$$logit(P(Y_i(t_{ij}) = 1 | \boldsymbol{X}_i(t_{ij}))) = \sum_{p=1}^{P} \mathcal{B}_{ip}(t_{ij}) X_{ip}(t_{ij}), \tag{2.1}$$

where $\boldsymbol{X}_i(\cdot)$ represents the $n_i \times P$-dimensional matrix of observed functional covariates and $\mathcal{B}_{ip}(\cdot)$ are subject- and covariate-specific smooth functions. Thus, $\mathcal{B}_{ip}(t_{ij})$ is interpreted as the expected change in the log odds of smoking for the i^{th} participant for a unit increase in X_{ip} at time t_{ij}, holding all else constant. In general the proposed model can readily capture population-level trends (i.e., $\mathcal{B}_{ip}(\cdot) = \mathcal{B}_p(\cdot)$ for all $i = 1, \ldots, N$) and other covariate forms. Similar *population-level* time-varying effect models are often applied to investigate the dynamic relations between smoking behaviors and risk factors around a scheduled quit attempt (Tan et al., 2012; Koslovsky et al., 2017; Shiyko et al., 2012).

2.1 Finite approximation of the smooth functions

A common approach for approximating smooth functions is with Bayesian penalized splines (Lang and Brezger, 2004). We assume the smooth functions are modeled as

$$\mathcal{B}_{ip}(t_{ij}) = \boldsymbol{\mathcal{T}}_{ijp}^{\intercal}\boldsymbol{\zeta}_{ip}, \tag{2.2}$$

where $\boldsymbol{\mathcal{T}}_{ijp}$ is a r_p-dimensional spline basis function for $\mathcal{B}_{ip}(t_{ij})$ and $\boldsymbol{\zeta}_{ip}$ is a r_p-dimensional vector of corresponding spline coefficients. The spline coefficients are typically assumed to follow a second-order random walk prior (i.e., $\boldsymbol{\zeta}_{ip}|s^2 \sim N(\mathbf{0}, s^2\boldsymbol{P})$), where s^2 controls the amount of smoothness and $\boldsymbol{P}$ is the appropriate penalty matrix (Lang and Brezger, 2004). In Scheipl et al. (2012), the authors show how priors of this type can be reparameterized with iid prior formulations that provide deeper insights into the underlying structure of the functional relations and enable variable selection. As pointed out by Ni et al. (2018), the reparameterization additionally enables penalization for the entire smooth function, as opposed to only the non-linear components. In addition to improving inference, we preferred this semiparamtric approach over a fully nonparametric alternative (e.g., the Gaussian process (Petrone et al., 2009; Scarpa and Dunson, 2009)) due to its computational efficiency and interpretability. Following (Scheipl et al., 2012), we deconstruct Equation (Eq.) (2.2) as a summation of a main effect, linear interaction term, and non-linear interaction term. Specifically, we set

$$\mathcal{B}_{ip}(t_{ij}) = \tilde{\beta}_{ip}\boldsymbol{\mathcal{T}}_{ijp}^{\prime\intercal}\boldsymbol{\xi}_{ip} + \bar{\beta}_{ip}t_{ij} + \dot{\beta}_{ip}, \tag{2.3}$$

where the constant term $\dot{\beta}_{ip}$ captures the main effect of $X_p(\cdot)$, $\bar{\beta}_{ip}$ is the linear interaction term for $X_p(\cdot)$ and time, and $\tilde{\beta}_{ip}\boldsymbol{\xi}_{ip}$ is a parameter-expanded vector of coefficients corresponding to the non-linear (penalized) interaction term. For interpretation, the r_p-dimensional vector $\boldsymbol{\xi}_{ip}$ maintains the shape of the non-linear portion of the smooth function and $\tilde{\beta}_{ip}$ controls the term's strength of association while preserving identifibility. The benefit of this approach is that it enables us to perform variable selection on the non-linear interaction term coefficients as a group via the prior specification of $\tilde{\beta}_{ip}$, which we illustrate in section 2.2. Further details of the reparameterization technique are provided in the Supplementary Material.

2.2 Spiked-nonparametric priors

In exploratory researcher settings, researchers often employ sparsity-inducing and clustering priors in various ways to perform dimension reduction, identify potential subpopulations, borrow strength across observations, and generate new hypotheses. Spiked-nonparamteric priors are designed to perform both tasks simultaneously in a unifying framework (Dunson et al., 2008a; Kim et al., 2009; Savitsky and Vannucci, 2010; Canale et al., 2017; Cassese et al., 2019; Koslovsky et al., 2020). In practice, researchers commonly employ spiked-Dirichlet (DP) or Pitman-Yor (PY) processes to simultaneously identify and cluster active terms in the model. For example, our team successfully developed a variable selection method for time-varying effect models using spiked-DP priors that identifies active *population-level* regression coefficients and clusters their magnitudes to help determine which risk factor to target at a given moment

(Koslovsky et al., 2020). Spiked process priors of this type are exchangeable, meaning that clustering allocation only depends on the size or cardinality of the current clustering. Alternatively, nonexchangeable priors, such as the distance dependence Chinese restaurant process (ddCRP, Blei and Frazier (2011)) and the EPA (Dahl et al., 2017), use auxiliary covariate information to inform clustering allocation.

In this work, we propose using a novel spiked-EPA distribution to simultaneously perform variable selection on individual-level smooth functions while using baseline covariate information to inform clustering allocation using an "inner" formulation. To help introduce the model's construction, we illustrate the class of species sampling models in which a sequence of random variables (e.g., $\theta_1, \theta_2, \dots$) is characterized by predictive probability functions of the following form

$$\theta_i|\theta_1, \cdots, \theta_{i-1} \sim w'(i)G_0(\theta_i) + \sum_{k=1}^{q_{i-1}} w_k(i)\delta_{\theta_k^*}(\theta_i),$$

where $\delta_x(\cdot)$ represents a point mass at x, $w'(i)$ and $w_k(i)$ represent weights that sum to one, G_0 is a probability measure, and q_{i-1} is the number of unique clusters (indexed by k) created during the realizations of $\theta_1, \dots, \theta_{i-1}$. For the familiar DP prior with concentration parameter α, the weights are defined as $w'(i) = \frac{\alpha}{\alpha+i-1}$ and $w_k(i) = \frac{n_k}{\alpha+i-1}$, where n_k is the number elements assigned to the k^{th} cluster. When the parameters follow a PY (a two-parameter generalization of the DP) process prior with concentration parameter α and discount parameter δ, we obtain a similar representation with $w'(i) = \frac{\alpha+\delta q_{i-1}}{\alpha+i-1}$ and $w_k(i) = \frac{n_k-\delta}{\alpha+i-1}$.

To construct the proposed spiked-EPA distribution, we first let $\boldsymbol{\theta}_{ip} = (\boldsymbol{\beta}'_{ip}, \boldsymbol{\xi}'_{ip}) = (\tilde{\beta}_{ip}, \bar{\beta}_{ip}, \dot{\beta}_{ip}, \boldsymbol{\xi}'_{ip})'$. Using a similar species sampling representation as above, we assume

$$\boldsymbol{\theta}_{\sigma_i,p}|\boldsymbol{\theta}_{\sigma_1,p}, \cdots, \boldsymbol{\theta}_{\sigma_{i-1},p} \sim w'_p(\sigma_i)G_0(\boldsymbol{\theta}_{\sigma_i,p}) + \sum_{k=1}^{q_{i-1,p}} w_{kp}(\sigma_i)\delta_{\boldsymbol{\theta}^*_{kp}}(\boldsymbol{\theta}_{\sigma_i,p}),$$

where $w_{kp}(\cdot)$ and $w'_p(\cdot)$ are the weight of assigning a subject to a realized cluster k and the weight of sampling a new cluster from the base distribution $G_0(\cdot)$ for the p^{th} smooth function, respectively. Due to the nonexchangeability of the EPA distribution, the clustering allocation order of the individuals may affect the realized clustering. We use $\boldsymbol{\sigma}$ to represent the allocation order. Further, $q_{i-1,p}$ represents the number of clusters created during the allocation of $\sigma_1, \cdots, \sigma_{i-1}$ for smooth function p and $\boldsymbol{\theta}^*_{kp}$ represents the unique realizations of $\boldsymbol{\theta}_{ip}$. A cluster is then characterized as $S_{kp} = \{i : \boldsymbol{\theta}_{ip} = \boldsymbol{\theta}^*_{kp}\}$. The disjoint clusters S_{kp} form a partition, π_p, of all subjects for covariate p. The weights $w_{kp}(\cdot)$ and $w'_p(\cdot)$ are defined following the EPA distribution as

$$w_{kp}(\sigma_i) = \frac{i-1-\delta_p q_{i-1,p}}{\alpha_p+i-1} \times \frac{\sum_{\{j:\ \sigma_j \in S_k,\ j<i\}} \lambda_p(\sigma_i, \sigma_j)}{\sum_{j'=1}^{i-1} \lambda_p(\sigma_i, \sigma_{j'})}, \tag{2.4}$$

$$\text{and } w'_p(\sigma_i) = \frac{\alpha_p + \delta_p q_{i-1,p}}{\alpha_p+i-1},$$

where $\delta_p \in [0, 1)$ is the discount parameter, $\alpha_p > -\delta_p$ is the concentration parameter, and $\lambda_p(\cdot, \cdot)$ is the similarity function, parameterized by the similarity parameter η_p, representing the attraction between two subjects. Note that the difference between the DP, PY, and EPA priors in the species sampling model formulation is captured via the weights, in addition to the ordering dependence. As opposed to the PY and DP priors, the EPA distribution accounts for the attraction $\lambda_p(\cdot, \cdot)$ of σ_i to all existing members in a given cluster. This alteration allows the clustering procedure to be influenced and modeled by additional covariates of interest through the formulation of the similarity function $\lambda_p(\cdot, \cdot)$. In the context of our application, this allows baseline participant information to inform clusters of time-varying effects for potential risk factors of smoking (e.g., urge to smoke) across participants.

A critical component of the implementation of our proposed method is the specification of the priors in the EPA which govern the clustering allocation. Following the recommendations of Dahl et al. (2017), we assume $\alpha_p \sim \text{Gamma}(a_\alpha, b_\alpha)$, $\delta_p \sim \gamma_{\delta_p}\text{Beta}(a_\delta, b_\delta) + (1 - \gamma_{\delta_p})\delta_0(\delta_p)$, where $\delta_0(\cdot)$ is a Dirac delta, or point mass, at zero, and γ_{δ_p} follows a Beta-Binomial$(a_{\gamma_\delta}, b_{\gamma_\delta})$. We assume the similarity function takes the form of an exponential decay function, $\lambda_p(\sigma_i, \sigma_j) = \exp(-\eta_p d_{ij})$, where d_{ij} indicates the Euclidean distance between a set of measures (e.g., baseline covariates, time) for individuals i and j. The similarity parameters η_p are assumed to be positive and follow a Gamma(a_η, b_η) distribution. Hence a smaller Euclidean distance d_{ij} leads to a larger similarity. Lastly, we assume each possible permutation of $\boldsymbol{\sigma}$ has the same possibility (i.e., every allocation order has the same uniform prior). By allowing the EPA hyperparameters to vary across smooth functions, we are able to infer the amount of clustering for each smooth function as well as the effect of the similarity function on clustering.

To complete the prior formulation, we specify the base distribution $G_0(\cdot)$, which ultimately controls the variable selection in the model. Here, our objective is to construct a prior that identifies active smooth functions at the subject level and further induces sparsity on the individual components that make up the smooth functions defined in Eq. 2.3. Specifically, for the base distribution of the species sampling prior, we assume $\boldsymbol{\beta}_{ip}$ follow a multivariate spike-and-slab distribution to identify active smooth functions at the individual level. We let

$$p(\boldsymbol{\beta}_{ip}|\gamma_{ip}) = \gamma_{ip}\mathcal{S}(\boldsymbol{\beta}_{ip}) + (1 - \gamma_{ip})\delta_0(\boldsymbol{\beta}_{ip}),$$

where $\mathcal{S}(\boldsymbol{\beta}_{ip})$ is referred to as the slab distribution and $\gamma_{ip} \in \{0, 1\}$ is a global inclusion indicator. When $\gamma_{ip} = 1$, the p^{th} smooth function is active in the model for individual i, and 0 otherwise (i.e., $\tilde{\beta}_{ip} = \bar{\beta}_{ip} = \dot{\beta}_{ip} = 0$). We assume γ_{ip} follows a Beta-Binomial(a_γ, b_γ), where the hyperparameters control the overall sparsity of the model, with prior mean $\frac{a_\gamma}{a_\gamma + b_\gamma}$. We further induce shrinkage on the active group of regression coefficients using horseshoe priors for $\mathcal{S}(\boldsymbol{\beta}_{ip})$ (Carvalho et al., 2010). We assume $\mathcal{S}(\boldsymbol{\beta}_{ip})$ follows a multivariate normal distribution with mean $\mathbf{0}_{3\times 1}$ and variance

$$\boldsymbol{V}_{ip} = \tau_p^2 \begin{pmatrix} \tilde{\lambda}_{ip}^2 & 0 & 0 \\ 0 & \bar{\lambda}_{ip}^2 & 0 \\ 0 & 0 & \dot{\lambda}_{ip}^2 \end{pmatrix}.$$

Here, τ_p^2 represents a global variance shared among each regression coefficient in $\boldsymbol{\beta}_{ip}$, and $\tilde{\lambda}_{ip}^2, \bar{\lambda}_{ip}^2$, and $\dot{\lambda}_{ip}^2$ are local variances. We assume $\tilde{\lambda}_{ip}^2 | \nu_{\tilde{\lambda}_{ip}^2} \sim \mathcal{IG}(1/2, 1/\nu_{\tilde{\lambda}_{ip}^2})$ and $\nu_{\tilde{\lambda}_{ip}^2} \sim \mathcal{IG}(1/2, 1)$, where $\nu_{\tilde{\lambda}_{ip}^2}$ is an auxiliary parameter introduced for efficient sampling following Makalic and Schmidt (2015) and $\mathcal{IG}$ represents an Inverse-Gamma distribution. To complete the model specification, we assume similar prior formulations for $\bar{\lambda}_{ip}^2, \dot{\lambda}_{ip}^2$ and τ_p^2 and their corresponding auxiliary parameters $\nu_{\bar{\lambda}_{ip}^2}$, $\nu_{\dot{\lambda}_{ip}^2}$, and ν_{τ_p}, respectively. Lastly, each element of $\boldsymbol{\xi}_{ip}$, ξ_{ipr}, $r = 1, \ldots, r_p$, is assumed to follow $N(\mu_{ipr}, 1)$, where $\mu_{ipr} = \pm 1$ with equal probability, namely $\mu_{ipr} \sim 0.5 \times \delta_1(\mu_{ipr}) + 0.5 \times \delta_{-1}(\mu_{ipr})$.

A unique aspect of our approach is the combination of spike-and-slab and horseshoe priors, two competing Bayesian methods for inducing sparsity in regression settings, to regularize active smooth functions. Alternatively, we could have assumed the base distribution follows a multivariate horseshoe prior. However, this approach would not explicitly perform variable selection on the smooth functions as it would allow non-zero $\boldsymbol{\beta}_{ip}$ terms, even if the time-varying effect was truly inactive. Similarly, we could have embedded spike-and-slab priors for the individual $\boldsymbol{\beta}_{ip}$ terms instead of the horseshoe priors. While this would have also achieved our goal of regularizing active smooth functions, this formulation required a more complicated Markov chain Monte Carlo (MCMC) algorithm that may result in poorer mixing.

2.3 Posterior inference

For posterior inference, we implement a Metropolis-Hastings within Gibbs sampling scheme (see Algorithm 1. See the Supplementary Materials for further details). Briefly, we use the Pólya-Gamma data augmentation technique of Polson et al. (2013) to efficiently update $\boldsymbol{\theta}^*$ by introducing a latent variable $\omega_{ij} \sim \text{PG}(1, 0)$ for each observation $Y_i(t_{ij})$. As such, the full joint distribution is written as

$$\begin{aligned} &f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}^*, \boldsymbol{\pi}, \boldsymbol{\omega}) p(\boldsymbol{\theta}^*|\boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{\tau}) p(\boldsymbol{\gamma}) p(\boldsymbol{\mu}) p(\boldsymbol{\lambda}|\boldsymbol{\nu}) p(\boldsymbol{\tau}|\boldsymbol{\nu}) p(\boldsymbol{\nu}) \\ &\times p(\boldsymbol{\pi}|\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\eta}, \boldsymbol{\sigma}) p(\boldsymbol{\alpha}) p(\boldsymbol{\delta}|\boldsymbol{\gamma}_\delta) p(\boldsymbol{\gamma}_\delta) p(\boldsymbol{\sigma}) p(\boldsymbol{\eta}) p(\boldsymbol{\omega}). \end{aligned}$$

Note that we use bold characters to represent variables in vector form. Similar in spirit to algorithm 8 in Neal (2000), we apply an auxiliary variable procedure to update the partitions π_p, which explicitly places a positive probability of assignment to the trivial cluster (i.e., $\boldsymbol{\theta}_{kp}^* = \mathbf{0}$) (Savitsky and Vannucci, 2010). The parameters in the EPA distribution (i.e., α_p, η_p, δ_p, and $\boldsymbol{\sigma}_p$) are updated using Metropolis-Hastings steps. When the discount parameter $\delta_p = 0$, a Gibbs update for α_p can be performed, as described in Escobar and West (1995). The mean of $\boldsymbol{\xi}$, $\boldsymbol{\mu}$, is updated using a Gibbs update. At each MCMC iteration, the $\tilde{\beta}_{ip}$ and $\boldsymbol{\xi}_{ip}$ terms are rescaled such that the mean of $|\boldsymbol{\xi}_{ip}|$ is equal to one. Lastly, we use the approach of Makalic and Schmidt (2015) to sample the horseshoe parameters.

After burn-in and thinning, the remaining samples obtained from the MCMC algorithm are used for inference. To determine a functional covariate's overall inclusion in the model for each subject, its marginal posterior probability of inclusion (MPPI) is

estimated by calculating the average of its respective inclusion indicator's MCMC samples (George and McCulloch, 1997). Commonly, covariates are included in the model if their MPPI exceeds 0.50, often referred to as the median model approach (Barbieri et al., 2004).

One of the challenges of using nonparametric Bayesian priors to induce clustering is that the resulting posterior samples do not provide direct inference on cluster allocation. As a result, researchers typically rely on post-hoc algorithms for posterior inference. To determine subjects' cluster assignments for each smooth function, we use the sequentially-allocated latent structure optimization (SALSO) method to minimize the lower bound of the variation of information loss (Meilă, 2003; Dahl et al., 2021). Dahl et al. (2021) demonstrate the superior clustering performance of the SALSO method compared to alternative approaches. This algorithm requires an $N \times N$-dimensional matrix representing the pairwise probabilities that subjects i and j are in the same cluster of a partition, estimated using the posterior samples. This approach, and other existing post-hoc partition summary methods, are not designed to handle spiked processes. In these settings, the pairwise probabilities for marginally active terms (i.e., MPPIs near 0.50) will be dominated by the trivial cluster. Instead, we propose defining the pairwise similarity matrix for each covariate p, $\boldsymbol{h}_p$, using only the samples in which coefficient terms for both subjects are active. Specifically, each element h_{ijp} is defined as

$$h_{ijp} = \left(\sum_{m=m_1+1}^{M} \mathbb{1}(\hat{s}_{ip}^{(m)} = \hat{s}_{jp}^{(m)}) \hat{\gamma}_{ip}^{(m)} \hat{\gamma}_{jp}^{(m)} \right) / \sum_{m=m_1+1}^{M} \hat{\gamma}_{ip}^{(m)} \hat{\gamma}_{jp}^{(m)},$$

where the first m_1 samples are treated as burn-in, $\hat{s}^{(m)}$ and $\hat{\gamma}^{(m)}$ are the posterior samples for the cluster and inclusion indicators at iteration m.

3 Simulation Study

In this section, we evaluate the clustering, variable selection, and estimation performance of our proposed method on simulated data. Notably, there are no existing methods that simultaneously perform functional clustering and selection at the individual level. Thus, to showcase the advantage of incorporating similarity information to inform clustering allocation, we compare the spiked-EPA prior with two exchangeable alternatives for the proposed method, the spiked-Dirichlet process (spiked-DP) prior (Savitsky and Vannucci, 2010) and the spiked-Pitman-Yor process (spiked-PY) (Canale et al., 2017). For all models, we assume an "inner" spiked-process formulation for consistency.

3.1 Data Generation

We simulated $N = 60$ subjects with $n_i = 20$ observations each. Observation times, t_{ij}, were sampled from a Uniform(0,1) distribution, without loss of generality. For each observation, we generated a set of 15 covariates $\boldsymbol{X}(t_{ij})$, which included an intercept term and 14 continuous covariates. Note that while the dimension of the covariate

Algorithm 1 MCMC sampler

Input data $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{\mathcal{T}}$
Initiate $\boldsymbol{\theta}$, $\boldsymbol{\pi}$, $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\boldsymbol{\lambda}$, $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$, $\boldsymbol{\delta}$, $\boldsymbol{\gamma}_\delta$, $\boldsymbol{\eta}$, $\boldsymbol{\sigma}$
for iteration $m = 1, \ldots, M$ **do**
    for iteration $p = 1, \ldots, P$ **do**
        Jointly update $\boldsymbol{\theta}_p$ and $\boldsymbol{\gamma}_p$ with a Between step (Savitsky et al., 2011).
        Jointly update $\boldsymbol{\theta}_p, \boldsymbol{\gamma}_p$, and $\boldsymbol{\pi}_p$ with Neal (2000) algorithm 8.
        Perform a Within Step for $\boldsymbol{\theta}_p$ with a Pólya-Gamma update (Polson et al., 2013).
    end for
    Rescale $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\xi}$ such that the mean of $|\boldsymbol{\xi}|$ is equal to one
    Update $\boldsymbol{\mu}$ with a Gibbs step.
    Update $\boldsymbol{\nu}$ and $\boldsymbol{\tau}$ with a Gibbs step (Makalic and Schmidt, 2015).
    if discount parameters $\boldsymbol{\delta}$ set to **0 then**
        Update $\boldsymbol{\alpha}$ with a Gibbs step (Escobar and West, 1995).
    else
        Update $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ with Metropolis-Hastings steps.
    end if
    Update $\boldsymbol{\eta}$ with Metropolis-Hastings steps.
    Update $\boldsymbol{\sigma}$ with Metropolis-Hastings steps.
end for

Individual i	$\mathcal{B}_{i1}$	$\mathcal{B}_{i2}$	$\mathcal{B}_{i3}$
1-20	$f_2(t_{ij})$	$f_3(t_{ij})$	$f_1(t_{ij})$
21-40	$f_5(t_{ij})$	$f_4(t_{ij})$	$f_1(t_{ij})$
41-60	$f_5(t_{ij})$	$f_3(t_{ij})$	$f_6(t_{ij})$

Table 1: Simulated varying-coefficients: Subject-specific smooth functions defined in the true model for the first three active terms.

space is relatively small, the number of potential models to chose from is quite large (i.e., 2^{PN}) as the effect of each functional covariate is determined at the individual level. The continuous covariates were generated from Normal$(0, \Sigma)$, where $\Sigma_{st} = w^{|s-t|}$ and $w = 0.3$. Seven of the covariates were jittered by Normal$(0, 1)$ at each observation time to mimic time-varying characteristics observed in practice. The first three smooth functions were defined such that the clustering patterns were different for each functional covariate (see Table 1 for details). For example, $\mathcal{B}_{i1}(\cdot)$ for individuals $i \in \{1 \ldots 20\}$ was set similarly (i.e., $\mathcal{B}_{i1}(\cdot) = f_2(\cdot)$), while $\mathcal{B}_{i1}(\cdot) = f_5(\cdot)$ for $i \in \{21 \ldots 60\}$.

The true smooth functions were defined as

- $f_1(t_{ij}) = \pi t_{ij} \cos(5\pi t_{ij}) - 1.2 t_{ij}$
- $f_2(t_{ij}) = \pi \sin(3\pi t_{ij}) + 1.4 t_{ij} - 1.6$
- $f_3(t_{ij}) = \pi \cos(2\pi t_{ij}) + 1.6$

- $f_4(t_{ij}) = -\pi\cos(2\pi t_{ij}) + 1.6t_{ij}$
- $f_5(t_{ij}) = \pi\sin(5t_{ij}) - \cos(\pi t_{ij})$
- $f_6(t_{ij}) = 0$.

Smooth functions $\mathcal{B}_4(\cdot) - \mathcal{B}_{15}(\cdot)$ were inactive in the true model for all individuals. An $N \times 4$ pairwise similarity matrix $\boldsymbol{Z}$ was constructed from two binary and two continuous covariates following the block-pattern described in Table 1. Both binary covariates were generated from Bernoulli distributions with mean 0.9, 0.1, and 0.5, and continuous covariates were generated from normal distributions with mean -3, 0, and 3 and variance 1, for individuals 1-20, 21-40, and 41-60, respectively. The pairwise distance matrix was then constructed using the Euclidean distance between individuals' covariate patterns (i.e., $d_{ij} = ||\boldsymbol{z}_i - \boldsymbol{z}_j||_2$). As such, the pairwise distance d_{ij} between individuals i and j was smaller if they were within the same block (e.g., $i, j \in \{1, \ldots, 20\}$) versus if they were in different blocks (e.g., $i \in \{1, \ldots, 20\}$ and $j \in \{21, \ldots, 40\}$). The former leads to a larger similarity metric and stronger attraction. Note that marginally, the similarity matrix does not match the true clustering pattern.

3.2 Model Specification and Inference

We applied the spiked-EPA, spiked-DP, and spiked-PY functional concurrent regression mixture models to 30 replicate data sets. In each simulation, the MCMC algorithm was run for 2,000 iterations, using the first 500 iterations as burn-in. Convergence of the models was determined using Geweke's diagnostics (Geweke et al., 1991), traceplots of the total number of active terms for all smooth functions, and traceplots of the coefficient estimates. Traceplots and corresponding Geweke's diagnostics can be found in the Supplementary Material.

For the spiked-EPA model, we set the discount parameters $\delta_p = 0$, which enabled Gibbs updates for the concentration parameters α_p. For the remaining hyperparameters, we set $a_\alpha = 5$, $b_\alpha = 1$, and $a_\eta = b_\eta = 1$ to impose a mean of 5 and 1 for the concentration and similarity parameters, respectively. We also set $a_\gamma = b_\gamma = 1$, imposing a non-informative prior probability of inclusion for each subject-specific smooth function. We assumed an exponential similarity function to inform clustering allocation in the spiked-EPA model, as described in section 2.2. The hyperparameters for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ were set to the same values for the spiked-DP and spiked-PY. For the spiked-PY model, we additionally set the hyperparameters for the discount parameter as $a_\delta = b_\delta = 2$ such that the distribution is symmetric and centered at 0.5.

To determine cluster allocation for each subject-specific smooth function, we applied the SALSO method to the active posterior samples, as described in section 2.3. We then evaluated clustering performance based on the variation of information (VI) (Meilă, 2007) and the adjusted Rand index (ARI) (Hubert and Arabie, 1985), which both measure closeness between the true and estimated clustering allocations. VI is always non-negative and values closer to 0 indicate better clustering performance. ARI takes on values between 0 and 1, with 0 indicating that two allocations do not agree on any pair

of items and 1 indicating that the allocations are the same. Inclusion in the model was determined using the median model approach (Barbieri et al., 2004). Variable selection performance was evaluated via true positive rate (TPR), false positive rate (FPR), and Matthew's correlation coefficient (MCC). These metrics are defined as

$$TPR = \frac{TP}{FN + TP}$$

$$FPR = \frac{FP}{FP + TN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}},$$

where TN, TP, FN, and FP represent the true negatives, true positives, false negatives, and false positives, respectively. Lastly, we evaluated the methods' ability to recover the true smooth functions by measuring the mean squared error ($\text{MSE}_{\mathcal{B}}$) between the estimates and the truth for the first three smooth functions at each observation time averaged across subjects. That is,

$$MSE_{\mathcal{B}} = \frac{1}{3\sum_{i=1}^{N} n_i} \sum_{p=1}^{3} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(\hat{\mathcal{B}}_{ip}(t_{ij}) - \mathcal{B}_{ip}(t_{ij}) \right)^2,$$

where $\hat{\mathcal{B}}_{ip}(\cdot)$ is the estimated smooth function. A functional covariate's inclusion in the model was determined based on their corresponding MPPIs.

3.3 Results

In this section, we discuss results of the clustering, variable selection, and estimation performance of the three models. The results of the active smooth functions (i.e., $p = 1, \ldots, 3$) are summarized in Table 2. We observed that the spiked-EPA model achieved the best functional clustering performance ($VI \leq 0.05$ and $ARI \geq 0.98$) and functional variable selection performance ($TPR \geq 0.98$) for the first two smooth functions. Note that the first two smooth functions were active for all subjects, and therefore we do not report FPR or MCC. In comparison, the spiked-DP and spiked-PY model achieved similar functional selection performance, with poorer functional clustering performance ($VI \leq 0.45$ and $ARI \geq 0.80$). All three models had difficulty estimating the third smooth function due to the complexity of $f_1(\cdot)$. However, the spiked-EPA model still demonstrated much better performance in terms of selection and clustering compared to the exchangeable alternatives. In contrast, the spiked-DP and spiked-PY models were not able to determine active and inactive smooth functions, resulting in lower TPR, higher FPR, and lower clustering performance. All three methods obtained $FPR < 0.08$ for the inactive functional terms.

These results demonstrate how accommodating similarities in subjects' auxiliary information can help improve clustering and ultimately selection performance, even when the auxiliary information does not fully capture the true clustering patterns. As a result

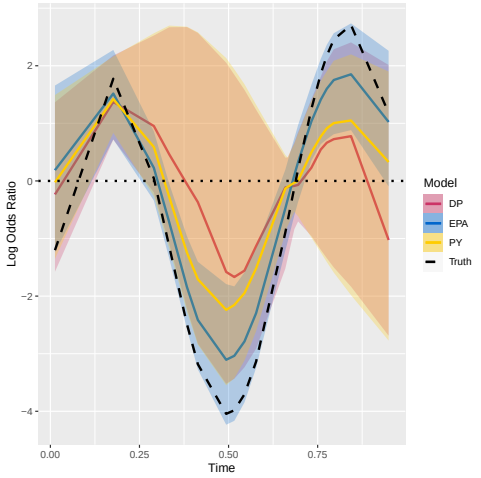
Smooth Function		**spiked-EPA**	**spiked-DP**	**spiked-PY**
	Runtime	10.54 (1.02)	9.63 (2.90)	8.64 (0.84)
	$\text{MSE}_{\mathcal{B}}$	0.31 (0.49)	0.73 (1.62)	0.80 (1.70)
$p = 1$	TPR	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
(Intercept)	FPR	-	-	-
	MCC	-	-	-
	VI	0.00 (0.00)	0.20 (0.20)	0.23 (0.30)
	ARI	1.00 (0.00)	0.92 (0.09)	0.91 (0.13)
$p = 2$	TPR	0.98 (0.03)	0.99 (0.02)	0.99 (0.02)
(Fully active)	FPR	-	-	-
	MCC	-	-	-
	VI	0.05 (0.10)	0.42 (0.34)	0.44 (0.35)
	ARI	0.98 (0.04)	0.81 (0.21)	0.80 (0.21)
$p = 3$	TPR	0.78 (0.21)	0.69 (0.25)	0.67 (0.27)
(Partially active)	FPR	0.26 (0.22)	0.51 (0.22)	0.56 (0.26)
	MCC	0.54 (0.25)	0.20 (0.15)	0.12 (0.19)
	VI	0.45 (0.41)	1.04 (0.12)	1.07 (0.18)
	ARI	0.74 (0.33)	0.10 (0.11)	0.11 (0.11)

Table 2: Simulation Results: Mean (SD) of the variation of information (VI), adjusted Rand Index (ARI), true positive rate (TPR), false positive rate (FPR), Matthew's correlation coefficient (MCC), and mean squared error for estimated and true smooth functions assessed at each observation time ($\text{MSE}_{\mathcal{B}}$) for spiked-EPA, spiked-DP, and spiked-PY models. Results are averaged over 30 replicate data sets. Runtime is in hours.

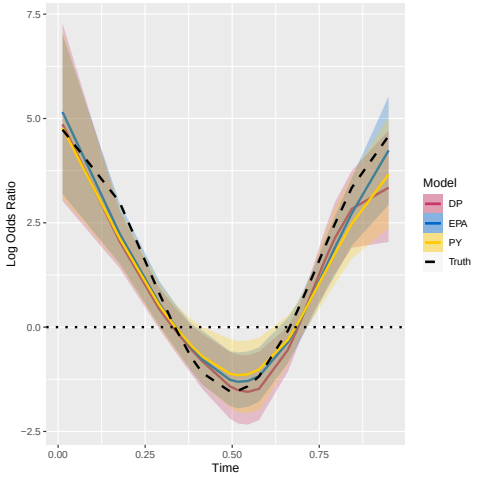
of the improved clustering and variable selection performance, the spiked-EPA model obtained more accurate estimates for the smooth functions in terms of the $\text{MSE}_{\mathcal{B}}$. To explore how the clustering and selection performance led to differences in estimation accuracy between the three models, Figure 1 presents the smooth functions estimated for subject $i = 9$ on a randomly selected data set. For the spiked-EPA model, we observe that the true smooth functions remain within the 95% credible intervals (CIs) of the estimated trajectories. In comparison, both of the alternative models misclassified the functional trajectories for the first functional covariate, which had a negative impact on estimation performance. Additionally, for the third functional covariate the spiked-DP method misclassified all subjects as inactive, while the spiked-PY method classified almost all subjects as active. As a result, the spiked-PY method's estimated smooth function for X_3 departed from the truth, while that of the spiked-DP method's estimated smooth function was biased towards zero. Regardless of the approach, the misclassification of the subjects' clustering allocation resulted in poorer estimation of the subject-specific trends downstream. Due to the larger parameter space and nonexchangeability of the spiked-EPA model, the runtime took roughly 25% longer than the spiked-PY model.

3.4 Sensitivity analysis

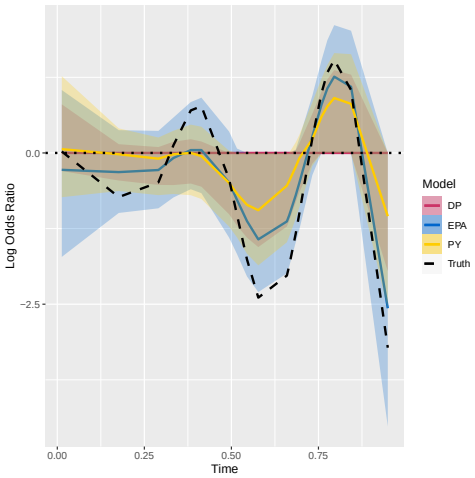
To assess the proposed model's sensitivity to hyperparameter specification, we set each of the hyperparameters to default values and then evaluated the effect of manipulating one term at a time on the selection and clustering performance. For the default parameterization, we used the settings described in the simulation study. Additionally,



(a) Estimated smooth functions for $\mathcal{B}_{91}$



(b) Estimated smooth functions for $\mathcal{B}_{92}$



(c) Estimated smooth functions for $\mathcal{B}_{93}$

Figure 1: Simulation Results: Estimated smooth functions for $\mathcal{B}_1$ to $\mathcal{B}_3$ for subject 9 in the simulation study using spiked-DP, spiked-PY, and spiked-EPA models. Solid lines and colored ribbons represent the estimated smooth functions and their corresponding 95% CI. Black dashed lines represent the true log odds ratios as a function of time.

we explored alternative similarity functions in the EPA distribution, including a logistic similarity function, $\lambda_p(i,j) = \frac{exp(-d_{ij}+\eta_p)}{1+exp(-d_{ij}+\eta_p)}$, and a window similarity function, $\lambda_p(i,j) = \mathbb{1}(d_{ij} < \eta_p) + 0.01\mathbb{1}(d_{ij} \geq \eta_p)$. For the logistic similarity function, any Euclidean distances between two subjects less (greater) than η_p would lead to a similarity

measure greater (less) than 0.5 with a maximum (minimum) at 1 (0). For the window function, we fixed η_p to the mean of the Euclidean distances between each pairwise combination of subjects. As such, the similarity measure takes on values of 1 or 0.01 based on whether the distance between two subjects, d_{ij}, is less than or greater than η_p. For each of the model specifications, we present results on the same 30 replicated data sets generated in the simulation study. The MCMC chains were run for 2,000 iterations, treating the first 500 as burn-in.

The results of the sensitivity analysis are shown in Table 3. We observed that the proposed model's sensitivity to hyperparameter specification varied for inactive and active functional covariates, as well as the local structure of the active smooth functions. For inactive terms, the results were not sensitive to hyperparameter specification. For the first two active smooth functions, the proposed method obtained similar selection and clustering performance, regardless of the value specified for the concentration parameter. However for the third smooth function, we observed that as a_α increased, the selection and clustering performance declined. The proposed model was relatively robust to the prior probability of inclusion. However, stronger attraction in the similarity function resulted in a decrease in selection and clustering performance for all three active smooth function. As expected, the performance of the spiked-EPA approached the performance of the spiked-DP and spiked-PY models as b_η increased. Lastly, compared to the exponential and logistic similarity functions, we observed a marginal reduction in selection and clustering performance when using the window similarity function. We attribute this to the window function capturing limited attraction information. Overall, the proposed method was more sensitive to the hyperparameters for the EPA portion of the model than the spike-and-slab portion.

4 Application

Mobile health (mHealth) methods have revolutionized the design and delivery of behavioral health interventions targeting smoking cessation and may help reduce smoking-related health disparities. A popular mHealth technique in behavioral health research is the use of ecological momentary assessment (EMA) methods to capture information about psychological, emotional, and environmental factors that may relate to a behavioral outcome in near real-time. An emerging strategy for delivering automated personalized support via mobile technology at critical moments throughout an assessment period is the just-in-time adaptive intervention (JITAI) (Spruijt-Metz and Nilsen, 2014). JITAIs are adaptive interventions that use dynamically changing information collected on an individual to determine when and how to deliver treatment in real-time. Historically, studies have used health behavior theory (Shiffman et al., 2002; Timms et al., 2014) or population-level trends of smoking antecedents (Piasecki et al., 2013) to determine JITAI decision rules. However, current health behavior theories are potentially inadequate for guiding the dynamic and granular nature of JITAIs as they may overlook important individual- and subpopulation-level smoking lapse trends that may be useful for targeting high-risk moments.

In this analysis, we demonstrate how our proposed method can be used to gain insights into individual-level patterns of the dynamic relations between risk factors and

Smooth Function		$a_\alpha = 1$		$a_\alpha = 10$		$a_\gamma = 4$		$b_\gamma = 9$		$b_\gamma = 99$	
$p = 1$	VI	0.00	(0.00)	0.01	(0.06)	0.01	(0.04)	0.01	(0.04)	0.01	(0.06)
	ARI	1.00	(0.00)	1.00	(0.02)	1.00	(0.01)	1.00	(0.01)	1.00	(0.02)
$p = 2$	VI	0.06	(0.19)	0.07	(0.20)	0.09	(0.18)	0.04	(0.08)	0.04	(0.18)
	ARI	0.97	(0.10)	0.97	(0.10)	0.95	(0.11)	0.99	(0.03)	0.98	(0.04)
	TPR	1.00	(0.01)	0.96	(0.07)	0.98	(0.06)	0.98	(0.02)	0.96	(0.08)
$p = 3$	VI	0.20	(0.20)	0.55	(0.45)	0.44	(0.43)	0.42	(0.40)	0.39	(0.40)
	ARI	0.92	(0.09)	0.66	(0.38)	0.73	(0.35)	0.74	(0.35)	0.76	(0.34)
	TPR	0.85	(0.19)	0.64	(0.26)	0.78	(0.20)	0.75	(0.29)	0.76	(0.29)
	FPR	0.17	(0.18)	0.25	(0.16)	0.33	(0.26)	0.31	(0.25)	0.28	(0.18)
	MCC	0.70	(0.20)	0.39	(0.27)	0.45	(0.31)	0.45	(0.34)	0.50	(0.26)

Smooth Function		$b_\eta = 5$		$a_\eta = 5$		logistic		window	
$p = 1$	VI	0.01	(0.04)	0.02	(0.07)	0.01	(0.04)	0.10	(0.18)
	ARI	1.00	(0.01)	0.99	(0.02)	1.00	(0.01)	0.96	(0.08)
$p = 2$	VI	0.07	(0.11)	0.14	(0.29)	0.08	(0.13)	0.31	(0.33)
	ARI	0.98	(0.04)	0.94	(0.15)	0.97	(0.05)	0.88	(0.13)
	TPR	0.98	(0.03)	0.96	(0.08)	0.98	(0.02)	0.97	(0.07)
$p = 3$	VI	0.65	(0.41)	0.60	(0.40)	0.37	(0.40)	0.71	(0.33)
	ARI	0.56	(0.38)	0.64	(0.36)	0.80	(0.30)	0.57	(0.31)
	TPR	0.70	(0.27)	0.68	(0.28)	0.83	(0.12)	0.71	(0.26)
	FPR	0.32	(0.21)	0.11	(0.12)	0.25	(0.19)	0.36	(0.23)
	MCC	0.40	(0.23)	0.56	(0.28)	0.59	(0.17)	0.38	(0.24)

Table 3: Sensitivity Results: Evaluation of the proposed method's sensitivity to hyper-parameter specification on simulated data. Results are averaged over 30 replicated data sets with standard deviations in parentheses. VI: variation of information; ARI: adjusted Rand Index; TPR: true positive rate; FPR: false positive rate; MCC: Matthew's correlation coefficient

smoking behaviors in the critical moments after a quit attempt using data collected in the Smart-T2 study, which was designed to investigate the utility of a novel, smartphone-based smoking cessation JITAI (Hébert et al., 2020). The Smart-T intervention uses a lapse risk estimator to identify moments of heightened risk for lapse, and tailors treatment messages in real-time based upon the level of imminent smoking lapse risk and currently present lapse triggers (Businelle et al., 2016). In this study, participants were followed over a five-week period from one week prior to a scheduled quit attempt to four weeks after. Throughout the assessment period, participants completed daily diaries and received four random EMAs from a study provided smartphone that asked each participant questions about their recent smoking behaviors and various factors that may contribute to an increased risk of smoking.

In this analysis, we investigated the time-varying relations between potential risk factors and smoking behaviors after the scheduled quit attempt for a subset of 46 participants who were randomized to the Smart-T treatment or usual care. Participants with less than 5 observations post-quit were removed prior to analysis. The median number of EMAs per participant was 126 (99-132 IQR). Reported momentary smoking, the outcome of interest, was defined as whether or not a participant reported smoking in the 4 hours prior to the current EMA, capturing momentary smoking behaviors during waking hours. At each EMA, a participant was asked questions on their current smoking

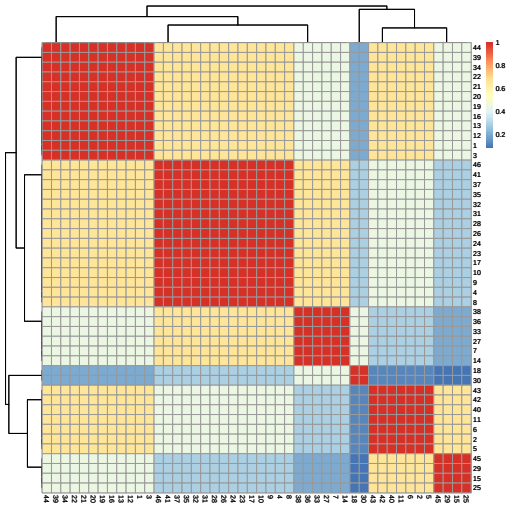


Figure 2: Heatmap for the attraction (i.e., $\lambda(i,j) = \exp(-0.5d_{ij})$) between subjects when using baseline HSI to construct the similarity matrix.

status and potential risk factors. Thus to maintain temporality, we assessed the relations between momentary smoking and measurements collected in the previous EMA. As such, for a particular risk factor, the regression coefficients can be interpreted as the log odds of momentary smoking by the next assessment at given time. Similar temporal assumptions are commonly made in smoking behavior research (Shiffman et al., 1996; Minami et al., 2014; Bolman et al., 2018; Koslovsky et al., 2018, 2020; Liang et al., 2021).

The functional covariates we investigated in this analysis were treatment assignment (binary; *usual care* or Smart-T intervention), drinking alcohol within the past hour (binary; *no* or yes), interacting with other smokers (binary; *no* or yes), urge to smoke (continuous), cigarette availability (continuous), positive affect (i.e., happiness and calmness) (continuous), negative affect (i.e., irritability, frustration/anger, sadness, worry, misery) (continuous), feeling restless (continuous), feeling bored (continuous), feeling anxious (continuous), motivation to avoid smoking (continuous), age (continuous), and sex (binary; *male* or female). Terms in italics were treated as the reference group when defining indicator variables. We controlled for a population-level effect of sex in the model (i.e., no selection or clustering). We used baseline heaviness of smoking index (HSI, continuous) to define the similarity metrics between subjects that informs the clustering allocation of the EPA, as presented in Figure 2. The pairwise distance matrix was then constructed using the Euclidean distance as defined in section 3. Lastly, all continuous covariates were standardized, and $t = 0$ represents the beginning of the scheduled quit attempt for each individual.

For the proposed method, we set the EPA hyperparameters to $a_\alpha = 1, b_\alpha = 5$, $a_\eta = b_\eta = 1$, and assumed a non-informative prior probability of inclusion for each

subject-specific functional covariate, characterized by $a_\gamma = b_\gamma = 1$. We compared the results of our method with the approach of Koslovsky et al. (2020) (PGBVS), which performs variable selection for time-varying effect models. Note that although PGBVS and the proposed model both use spiked-nonparametric priors for variable selection, PGBVS clusters similar regression coefficients for different covariates, whereas the proposed method marginally clusters individual-level time-varying effects. Further, although PGBVS identifies and estimates subject-specific main effects and linear interaction terms, it is not designed to perform variable selection for random effects for the non-linear interaction terms in the model. Lastly, PGBVS relies on a spiked Dirichlet process prior which does not incorporate similarity information when determining cluster assignment.

For posterior inference, we initiated the MCMC algorithms with the null model (i.e., $\boldsymbol{\beta} = \mathbf{0}$). For both the spiked-EPA and the PGBVS model, we ran the MCMC algorithm for 10,000 iterations, using the first 5,000 as burn-in. Trace plots of the regression coefficients and the total number of active terms in the model indicated model convergence and appropriate mixing of the posterior samples. Inference on clustering allocation was evaluated using the SALSO method for active terms, as described in section 2.3. For both models, selection was determined using the median model approach (Barbieri et al., 2004).

4.1 Results

Both PGBVS and the proposed method obtained similar findings to previous studies that investigated the temporal relation between risk factors and smoking behaviors around a quit attempt. These findings include associations between urge to smoke, negative affect, cigarette availability, and momentary smoking after the quit attempt (Vasilenko et al., 2014; Koslovsky et al., 2017). Compared to PGBVS, which performs selection for functional covariates at the population level, the proposed method is able to identify and cluster functional covariates at the individual level. Using our proposed method, we observed varying-levels of subjects with active smooth functions for each functional covariate. We found that the effect of age, Smart-T treatment, cigarette availability, anxiousness, and motivation to quit smoking were active for over 65% of the subjects in this study. For these potential risk factors, active participants were clustered together, except for age, where two subjects were clustered separately. We also identified three clusters for the time-varying intercept term. Roughly 35% of the participants had active smooth functions for urge to smoke, and the effect of negative affect was only associated with momentary smoking for 6 participants. No associations between momentary smoking by the next assessment and interacting with other smokers, positive affect, restlessness, being bored, and drinking alcohol were identified for any participants by the proposed model. By implementing a spiked-EPA distribution to simultaneously perform functional variable selection and clustering, our approach leverages information from the similarity matrix to inform clustering. For example, those with active urge to smoke effects had higher baseline HSI compared to those with inactive effects. In practice, this information may help develop and evaluate *personalized* smoking cessation intervention strategies based on baseline subject- or group-specific characteristics.

Figure 3 presents the average time-varying odds ratios (ORs) and corresponding 95% CIs for all active subjects within each cluster estimated by the proposed method. Both PGBVS (Figure 4) and the proposed method identified similar potential risk factors for momentary smoking by the next assessment. However, the estimated trends often varied between models. Notably, since our approach allows for marginal subject-specific time-varying effect selection, it may identify a potential risk factor for a subset of the participants, whereas population-level models may overlook the relation. For example, the proposed method identified two active clusters for age (one with a positive relation with smoking by the next assessment and one with a negative relation) in addition to the trivial cluster. However, PGBVS did not identify age at the population level. A unique aspect of the proposed method is that it clusters functional covariates marginally (i.e., the cluster allocations may be different for each functional covariate across subjects). We observed significant differences in how participants were clustered to active and inactive groups across covariates. For covariates with at least one active participant, the maximum pairwise ARI was less than 0.15. These results further demonstrate the need for designing personalized intervention strategies that are adaptive to individual's unique risk profiles over time.

4.2 Sensitivity Analysis

We performed a sensitivity analysis for the concentration and attraction strength hyperparameter specifications on the application results. We set the model specification in the application analysis as the default setting and evaluated the effect of changing the hyperparameters on the results. Since the true clustering and individual-level smooth functions are never known in practice, we compared the clustering allocations across hyperparameter specifications to the results obtained above. Overall, the results were relatively robust to hyperparameter specification, with the exception of urge to smoke and anxiousness. We observed a positive (negative) relation between the prior mean of the concentration (attraction strength) parameter and the number of clusters observed in the model, as expected.

5 Discussion

In this work, we have proposed a fully Bayesian approach for simultaneous functional variable selection and clustering for functional concurrent regression mixture models using novel spiked-EPA priors. Unlike previously developed Bayesian model-based functional clustering methods, our approach clusters subjects across multiple smooth functions. Instead of clustering across the entire linear predictor, as is done in typical mixture model formulations, we investigate clusters of semiparametric smooth functions marginally for each functional covariate. Additionally, we embed functional variable selection and penalized priors to induce sparsity globally for each functional covariate at the subject-level and shrink non-influential spline coefficients for each smooth function. Our approach leverages auxiliary covariate information to improve clustering and selection performance. As such, our proposed method establishes a generalized

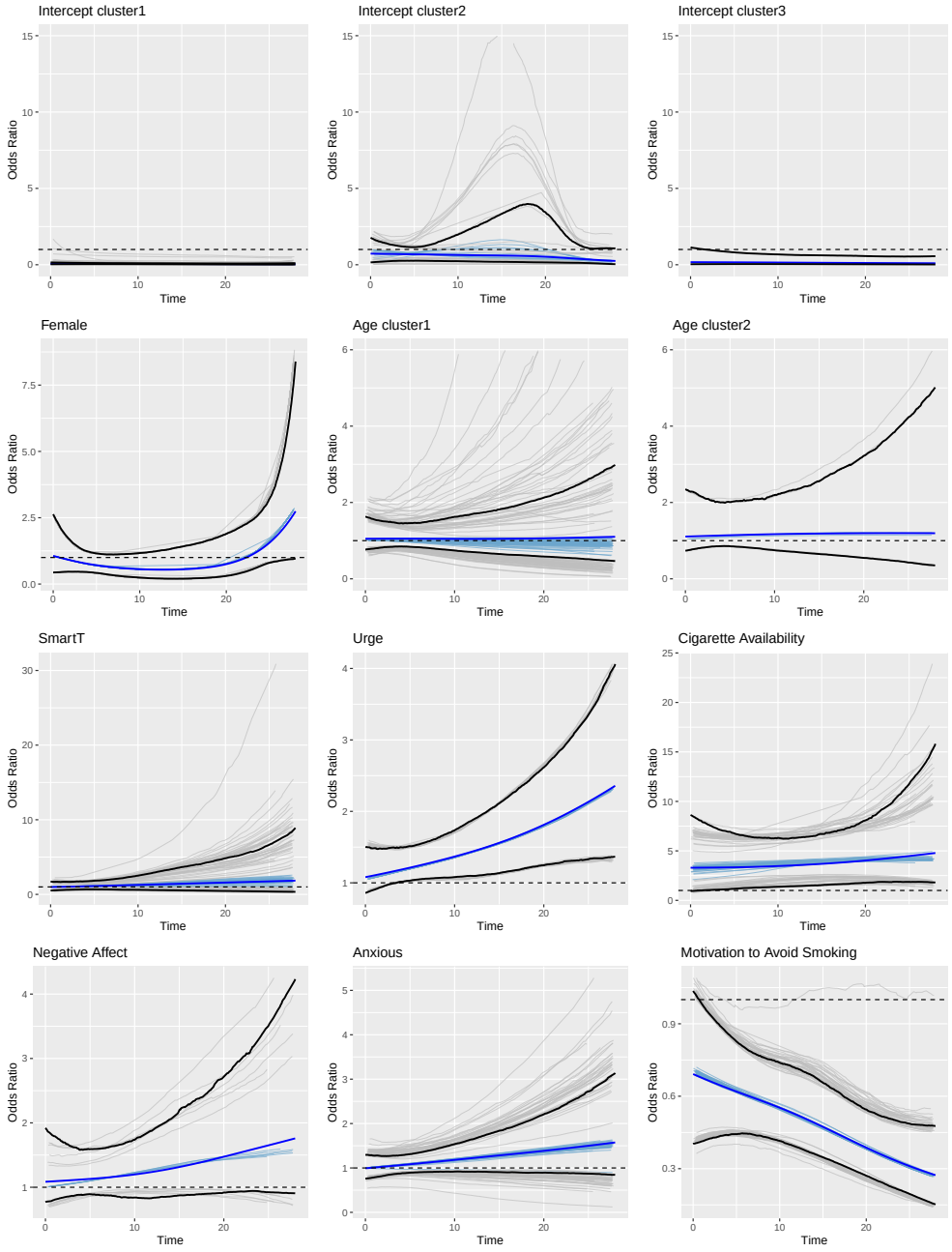


Figure 3: Application Results: Time-varying odds ratios (ORs) of momentary smoking by the next assessment for functional covariates selected with the spiked-EPA model. Thin blue/gray curve indicates the subject-level mean trajectories and 95% CIs. Thick blue/black curves represent cluster-specific mean trajectories and CIs. Black dashed line indicate an $OR = 1$.

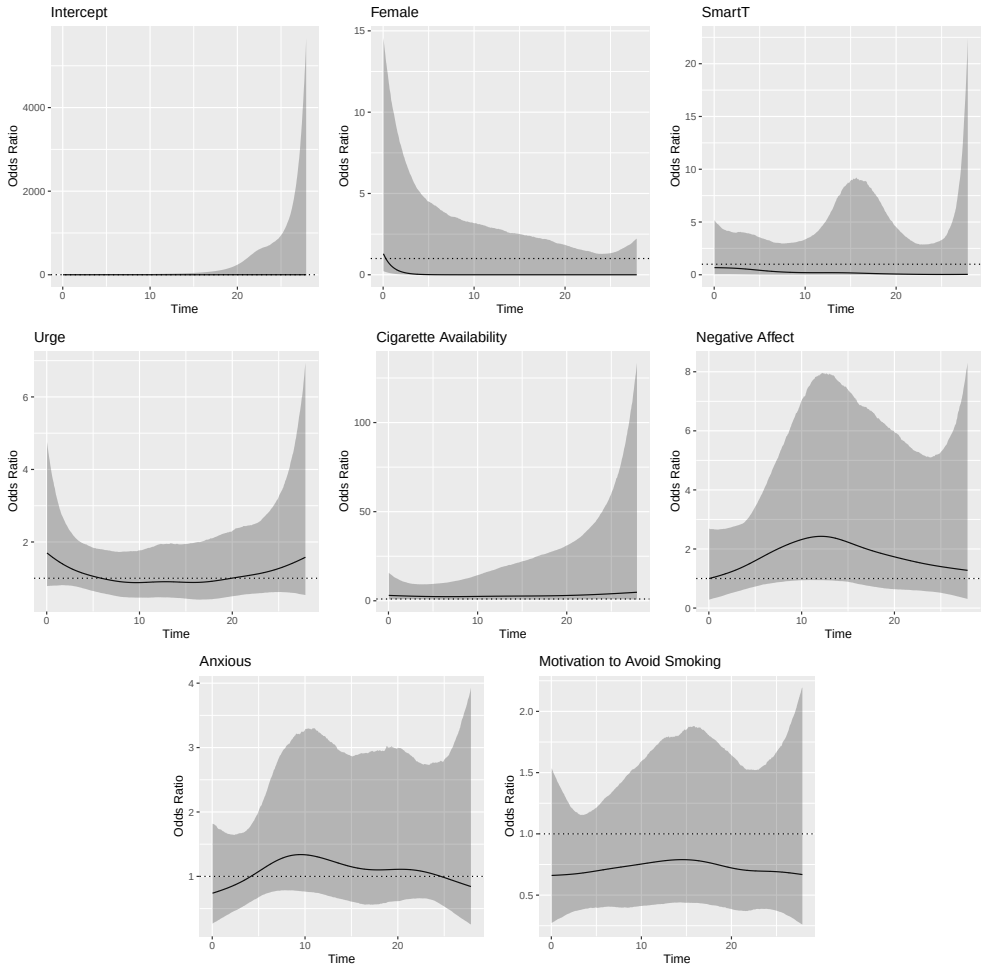


Figure 4: Application Results: *Population-level* time-varying odds ratios (ORs) of momentary smoking by the next assessment for functional covariates selected in both models, estimated with PGBVS. Shaded area indicates 95% CIs. Black curves represent mean trajectories. Black dotted line indicates $OR = 1$.

Risk Factor	Default	$b_\alpha = 1$	$b_\alpha = 20$	$a_\eta = 5$	$b_\eta = 5$
Intercept	(32,13,1)	(32,13,1)	(32,14)	(31,15)	(31,2,13)
Age	(44,2)	(46)	(46)	(41,3,2)	(46)
Smart-T	(46)	(42,3,1*)	(46)	(45,1*)	(44,2)
Urge	(17,29*)	(35,11*)	(46*)	(46*)	(35,11*)
Cigarette Availability	(30,16*)	(30,16*)	(29,17*)	(28,18*)	(30,16*)
Interacting with smokers	(46*)	(46*)	(46*)	(2,44*)	(46*)
Negative Affect	(6,40*)	(46*)	(46*)	(46*)	(1,45*)
Anxiousness	(37,9*)	(10,1,35*)	(40,6*)	(10,36*)	(36,10*)
Motivation to Quit Smoking	(43,3*)	(43,3*)	(43,3*)	(43,3*)	(43,3*)

Table 4: Sensitivity Results: Number of participants assigned to each cluster for a given risk factor across different hyperparameter specifications for the proposed spiked-EPA model for the Smart-T2 data. Only risk factors with at least one active smooth function are presented. * - indicates the trivial cluster.

approach for simultaneously clustering multiple functional covariates and performing variable selection using spiked process priors.

Motivated by momentary smoking behavior data collected in the application study, the proposed method is designed for binary outcomes. However, it can be easily extended to accommodate continuous outcomes via the connection between the Pólya-Gamma data augmentation technique of Polson et al. (2013) and a Gaussian likelihood function. Similar to other mixture models, the proposed method may encounter identifiability issues as the likelihood function is invariant to the cluster labels. This complicates inference for subject-specific functional trends as the cluster labels change across MCMC samples. To address this issue, we take the common post-hoc strategy of obtaining a point estimate for cluster allocation (e.g., via the SALSO method using the variation of information loss (Wade et al., 2018; Dahl et al., 2021)) and then estimating cluster specific functional trajectories (see Supplementary Material for more details). A known limitation of Dirichlet process mixture models is that they tend to overestimate the number of clusters in the data (Miller and Harrison, 2013). Compared to the PY and DP prior, the EPA distribution allows for more flexibility as clustering allocation is a function of the concentration, discount, and similarity parameter as well as the observed distance matrix. While this may help control the number of clusters identified in the model, it may also have the opposite effect. For example in the proposed simulation, there were two true clusters in the data, but there were three groups based on the similarity matrix. As a result the spiked-EPA model often suggested more clusters than the spiked-DP and spiked-PY, but still maintained improved clustering performance. To help protect against excess clusters, we recommend the variation of information loss to provide posterior clustering estimates as it has shown to be more conservative with respect to the number of estimated cluster compared to alternative loss functions (Dahl et al., 2021).

It is important to note that our method is designed for exploratory research settings in which researchers are interested in generating hypotheses for subject-level dynamic risk profiles and potential subpopulations. While spike-and-slab priors may be used to

identify potential risk factors in practice, this does not guarantee that the corresponding regression coefficients' 95% credible intervals do not contain zero. In fact, this is often the case for the results captured by the proposed method and PGBVS in the application. As such, we emphasize the importance of performing confirmatory analyses before generalizing the results of this study (or any exploratory analysis). Further, we strongly recommend performing thorough sensitivity analyses to understand the impact of prior specification on inference.

In an accompanying `R` package available at `https://github.com/mliang4/EPAbvs`, we provide code to simulate data similar to our simulation study and implement our proposed method. In our implementation, we have used Gibbs-type updates to fully sample partitions. This provides an explicit probability distribution for a given partition which enables inference on upstream parameters in the EPA distribution. In situations where inference on upstream parameters is not of interest, Airoldi et al. (2014) and Cassese et al. (2019) suggest using a graph-based sampler with data-pairing labels instead of cluster assignments, as in Blei and Frazier (2011). This sampler allows for large moves in the state-space and relatively fast mixing. Finally, while we are the first to address the limitations of existing post-hoc partition summary methods for spiked processes in practice, we acknowledge that our naive solution lacks theoretical justification and may reduce the effective sample size of the sampled chains. Future work could investigate more robust post-hoc partition summary methods for spiked process priors.

Supplementary Material

The Supplementary Material contains details related the proposed MCMC algorithm, simulation and application results, parameter estimation, and reparameterization scheme .

References

Abraham, C., Cornillon, P.-A., Matzner-Løber, E., and Molinari, N. (2003). "Unsupervised curve clustering using B-splines." *Scandinavian Journal of Statistics*, 30(3): 581–595. 2

Airoldi, E. M., Costa, T., Bassetti, F., Leisen, F., and Guindani, M. (2014). "Generalized species sampling priors with latent beta reinforcements." *Journal of the American Statistical Association*, 109(508): 1466–1480. 23

Barbieri, M. M., Berger, J. O., et al. (2004). "Optimal predictive model selection." *The Annals of Statistics*, 32(3): 870–897. 9, 12, 18

Barcella, W., Iorio, M. D., Baio, G., and Malone-Lee, J. (2016). "Variable selection in covariate dependent random partition models: An application to urinary tract infection." *Statistics in Medicine*, 35(8): 1373–1389. 2

Bigelow, J. L. and Dunson, D. B. (2009). "Bayesian semiparametric joint models for functional predictors." *Journal of the American Statistical Association*, 104(485): 26–36. 3

Blei, D. M. and Frazier, P. I. (2011). "Distance Dependent Chinese Restaurant Processes." *Journal of Machine Learning Research*, 12(8). 2, 6, 23

Bolman, C., Verboon, P., Thewissen, V., Boonen, V., Soons, K., and Jacobs, N. (2018). "Predicting smoking lapses in the first week of quitting: An ecological momentary assessment study." *Journal of Addiction Medicine*, 12(1): 65. 17

Businelle, M. S., Ma, P., Kendzor, D. E., Frank, S. G., Vidrine, D. J., and Wetter, D. W. (2016). "An ecological momentary intervention for smoking cessation: Evaluation of feasibility and effectiveness." *Journal of Medical Internet Research*, 18(12): e321. 16

Canale, A., Lijoi, A., Nipoti, B., and Prünster, I. (2017). "On the Pitman–Yor process with spike and slab base measure." *Biometrika*, 104(3): 681–697. 3, 5, 9

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). "The horseshoe estimator for sparse signals." *Biometrika*, 97(2): 465–480. 7

Cassese, A., Zhu, W., Guindani, M., Vannucci, M., et al. (2019). "A Bayesian nonparametric spiked process prior for dynamic model selection." *Bayesian Analysis*, 14(2): 553–572. 3, 5, 23

Dahl, D. B., Day, R., and Tsai, J. W. (2017). "Random partition distribution indexed by pairwise information." *Journal of the American Statistical Association*, 112(518): 721–732. 2, 6, 7

Dahl, D. B., Johnson, D. J., and Mueller, P. (2021). "Search Algorithms and Loss Functions for Bayesian Clustering." *arXiv preprint arXiv:2105.04451*. 9, 22

Das, K., Ghosh, P., and Daniels, M. J. (2021). "Modeling multiple time-varying related groups: A dynamic hierarchical Bayesian approach with an application to the Health and Retirement Study." *Journal of the American Statistical Association*, 1–11. 2

Ding, D. and Karabatsos, G. (2021). "Dirichlet process mixture models with shrinkage prior." *Stat*, 10(1): e371. 2, 3

Dunson, D. B., Herring, A. H., and Engel, S. M. (2008a). "Bayesian selection and clustering of polymorphisms in functionally related genes." *Journal of the American Statistical Association*, 103(482): 534–546. 3, 5

Dunson, D. B., Herring, A. H., and Siega-Riz, A. M. (2008b). "Bayesian inference on changes in response densities over predictor clusters." *Journal of the American Statistical Association*, 103(484): 1508–1517. 2

Escobar, M. D. and West, M. (1995). "Bayesian density estimation and inference using mixtures." *Journal of the American Statistical Association*, 90(430): 577–588. 8, 10

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 209–230. 2

Ferreira, L. and Hitchcock, D. B. (2009). "A comparison of hierarchical methods for clustering functional data." *Communications in Statistics-Simulation and Computation*, 38(9): 1925–1949. 2

George, E. I. and McCulloch, R. E. (1997). "Approaches for Bayesian variable selection." *Statistica Sinica*, 339–373. 9

Gertheiss, J., Maity, A., and Staicu, A.-M. (2013). "Variable selection in generalized functional linear models." *Stat*, 2(1): 86–101. 3

Geweke, J. F. et al. (1991). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." Technical report, Federal Reserve Bank of Minneapolis. 11

Ghosal, R., Maity, A., Clark, T., and Longo, S. B. (2020). "Variable selection in functional linear concurrent regression." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(3): 565–587. 3

Goldsmith, J. and Schwartz, J. E. (2017). "Variable selection in the functional linear concurrent model." *Statistics in Medicine*, 36(14): 2237–2250. 3

Goldsmith, J., Zipunnikov, V., and Schrack, J. (2015). "Generalized multilevel function-on-scalar regression and principal component analysis." *Biometrics*, 71(2): 344–353. 2

Hébert, E. T., Ra, C. K., Alexander, A. C., Helt, A., Moisiuc, R., Kendzor, D. E., Vidrine, D. J., Funk-Lawler, R. K., and Businelle, M. S. (2020). "A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial." *Journal of medical Internet research*, 22(3): e16907. 4, 16

Hu, G., Geng, J., Xue, Y., and Sang, H. (2020). "Bayesian spatial homogeneity pursuit of functional data: an application to the us income distribution." *arXiv preprint arXiv:2002.06663*. 2

Hubert, L. and Arabie, P. (1985). "Comparing partitions." *Journal of Classification*, 2(1): 193–218. 11

Islam, M. N., Stallings, J., Staicu, A.-M., Crouch, D., Pan, L., and Huang, H. (2018). "Functional Variable Selection for EMG-based Control of a Robotic Hand Prosthetic." *arXiv preprint arXiv:1805.03098*. 3

Jacques, J. and Preda, C. (2014). "Functional data clustering: A survey." *Advances in Data Analysis and Classification*, 8(3): 231–255. 2

James, G. M., Wang, J., and Zhu, J. (2009). "Functional linear regression that's interpretable." *The Annals of Statistics*, 37(5A): 2083–2108. 1

Jin, W., Ni, Y., Rubin, L. H., Spence, A. B., and Xu, Y. (2021). "A Bayesian nonparametric approach for inferring drug combination effects on mental health in people with HIV." *Biometrics*. 2

Kim, J. S., Staicu, A.-M., Maity, A., Carroll, R. J., and Ruppert, D. (2018). "Additive function-on-function regression." *Journal of Computational and Graphical Statistics*, 27(1): 234–244. 2

Kim, S., Dahl, D. B., and Vannucci, M. (2009). "Spiked Dirichlet process prior for

Bayesian multiple hypothesis testing in random effects models." *Bayesian Analysis (Online)*, 4(4): 707. 2, 3, 5

Koslovsky, M. D., Hébert, E. T., Businelle, M. S., Vannucci, M., et al. (2020). "A Bayesian time-varying effect model for behavioral mHealth data." *The Annals of Applied Statistics*, 14(4): 1878–1902. 3, 5, 6, 17, 18

Koslovsky, M. D., Hébert, E. T., Swartz, M. D., Chan, W., Leon-Novelo, L., Wilkinson, A. V., Kendzor, D. E., and Businelle, M. S. (2017). "The time-varying relations between risk factors and smoking before and after a quit attempt." *Nicotine & Tobacco Research*. 4, 18

Koslovsky, M. D., Swartz, M. D., Chan, W., Leon-Novelo, L., Wilkinson, A. V., Kendzor, D. E., and Businelle, M. S. (2018). "Bayesian variable selection for multistate Markov models with interval-censored data in an ecological momentary assessment study of smoking cessation." *Biometrics*, 74(2): 636–644. 17

Lang, S. and Brezger, A. (2004). "Bayesian P-splines." *Journal of Computational and Graphical Statistics*, 13(1): 183–212. 5

Leroux, A., Xiao, L., Crainiceanu, C., and Checkley, W. (2018). "Dynamic prediction in functional concurrent regression with an application to child growth." *Statistics in Medicine*, 37(8): 1376–1388. 2

Liang, M., Koslovsky, M. D., Hébert, E. T., Kendzor, D. E., Businelle, M. S., and Vannucci, M. (2021). "Bayesian Continuous-Time Hidden Markov Models with Covariate Selection for Intensive Longitudinal Data with Measurement Error." *Psychological Methods*, In Press. 17

Ma, P., Zhong, W., Feng, Y., and Liu, J. S. (2008). "Bayesian functional data clustering for temporal microarray data." *International Journal of Plant Genomics*, 2008. 2

Ma, S., Huang, J., Wei, F., Xie, Y., and Fang, K. (2011). "Integrative analysis of multiple cancer prognosis studies with gene expression measurements." *Statistics in Medicine*, 30(28): 3361–3371. 3

Maity, A. (2017). "Nonparametric functional concurrent regression models." *Wiley Interdisciplinary Reviews: Computational Statistics*, 9(2): e1394. 2

Makalic, E. and Schmidt, D. F. (2015). "A simple sampler for the horseshoe estimator." *IEEE Signal Processing Letters*, 23(1): 179–182. 8, 10

Meilă, M. (2003). "Comparing clusterings by the variation of information." In *Learning Theory and Kernel Machines*, 173–187. Springer. 9

— (2007). "Comparing clusterings—an information based distance." *Journal of Multivariate Analysis*, 98(5): 873–895. 11

Miller, J. W. and Harrison, M. T. (2013). "A simple example of Dirichlet process mixture inconsistency for the number of components." *Advances in neural information processing systems*, 26. 22

Minami, H., Yeh, V. M., Bold, K. W., Chapman, G. B., and McCarthy, D. E. (2014).

"Relations among affect, abstinence motivation and confidence, and daily smoking lapse risk." *Psychology of Addictive Behaviors*, 28(2): 376. 17

Morris, J. S. (2015). "Functional regression." *Annual Review of Statistics and Its Application*, 2: 321–359. 1

Neal, R. M. (2000). "Markov chain sampling methods for Dirichlet process mixture models." *Journal of Computational and Graphical Statistics*, 9(2): 249–265. 8, 10

Ni, Y., Stingo, F. C., Ha, M. J., Akbani, R., and Baladandayuthapani, V. (2018). "Bayesian Hierarchical Varying-Sparsity Regression Models with Application to Cancer Proteogenomics." *Journal of the American Statistical Association*, 1–13. 5

Petrone, S., Guindani, M., and Gelfand, A. E. (2009). "Hybrid Dirichlet mixture models for functional data." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(4): 755–782. 5

Piasecki, T. M., Trela, C. J., Hedeker, D., and Mermelstein, R. J. (2013). "Smoking antecedents: Separating between-and within-person effects of tobacco dependence in a multiwave ecological momentary assessment investigation of adolescent smoking." *Nicotine & Tobacco Research*, 16(Suppl_2): S119–S126. 15

Pitman, J. and Yor, M. (1997). "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability*, 855–900. 2

Polson, N. G., Scott, J. G., and Windle, J. (2013). "Bayesian inference for logistic models using Pólya–Gamma latent variables." *Journal of the American Statistical Association*, 108(504): 1339–1349. 8, 10, 22

Quintana, F. A., Mueller, P., Jara, A., and MacEachern, S. N. (2020). "The dependent Dirichlet process and related models." *arXiv preprint arXiv:2007.06129*. 2

Ray, S. and Mallick, B. (2006). "Functional clustering by Bayesian wavelet methods." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2): 305–332. 2

Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). "Methods for scalar-on-function regression." *International Statistical Review*, 85(2): 228–249. 1

Savitsky, T. and Vannucci, M. (2010). "Spiked Dirichlet process priors for Gaussian process models." *Journal of Probability and Statistics*, 2010. 3, 5, 8, 9

Savitsky, T., Vannucci, M., and Sha, N. (2011). "Variable selection for nonparametric Gaussian process priors: Models and computational strategies." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 26(1): 130–149. 10

Scarpa, B. and Dunson, D. B. (2009). "Bayesian hierarchical functional data analysis via contaminated informative priors." *Biometrics*, 65(3): 772–780. 5

Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). "Spike-and-slab priors for function selection in structured additive regression models." *Journal of the American Statistical Association*, 107(500): 1518–1532. 3, 5

Shiffman, S., Gwaltney, C. J., Balabanis, M. H., Liu, K. S., Paty, J. A., Kassel, J. D., Hickcox, M., and Gnys, M. (2002). "Immediate antecedents of cigarette smoking: an analysis from ecological momentary assessment." *Journal of Abnormal Psychology*, 111(4): 531. 15

Shiffman, S., Paty, J. A., Gnys, M., Kassel, J. A., and Hickcox, M. (1996). "First lapses to smoking: within-subjects analysis of real-time reports." *Journal of Consulting and Clinical Psychology*, 64(2): 366. 17

Shin, M., Bhattacharya, A., and Johnson, V. E. (2020). "Functional horseshoe priors for subspace shrinkage." *Journal of the American Statistical Association*, 115(532): 1784–1797. 3

Shiyko, M. P., Lanza, S. T., Tan, X., Li, R., and Shiffman, S. (2012). "Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers." *Prevention science*, 13(3): 288–299. 4

Spruijt-Metz, D. and Nilsen, W. (2014). "Dynamic models of behavior for just-in-time adaptive interventions." *IEEE Pervasive Computing*, 13(3): 13–17. 15

Suarez, A. J. and Ghosal, S. (2016). "Bayesian clustering of functional data using local features." *Bayesian Analysis*, 11(1): 71–98. 2

Tadesse, M. G., Sha, N., and Vannucci, M. (2005). "Bayesian variable selection in clustering high-dimensional data." *Journal of the American Statistical Association*, 100(470): 602–617. 3

Tan, X., Shiyko, M. P., Li, R., Li, Y., and Dierker, L. (2012). "A time-varying effect model for intensive longitudinal data." *Psychological Methods*, 17(1): 61. 2, 4

Tarpey, T. and Kinateder, K. K. (2003). "Clustering Functional Data." *Journal of Classification*, 20(1). 2

Timms, K. P., Rivera, D. E., Collins, L. M., and Piper, M. E. (2014). "Continuous-time system identification of a smoking cessation intervention." *International Journal of Control*, 87(7): 1423–1437. 15

Vasilenko, S. A., Piper, M. E., Lanza, S. T., Liu, X., Yang, J., and Li, R. (2014). "Time-varying processes involved in smoking lapse in a randomized trial of smoking cessation therapies." *Nicotine & Tobacco Research*, 16(Suppl 2): S135–S143. 18

Wade, S., Ghahramani, Z., et al. (2018). "Bayesian cluster analysis: Point estimation and credible balls (with discussion)." *Bayesian Analysis*, 13(2): 559–626. 22

Wang, L., Li, H., and Huang, J. Z. (2008). "Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements." *Journal of the American Statistical Association*, 103(484): 1556–1569. 3

White, P. A. and Gelfand, A. E. (2020). "Multivariate functional data modeling with time-varying clustering." *TEST*, 1–17. 2

Wiesenfarth, M., Hisgen, C. M., Kneib, T., and Cadarso-Suarez, C. (2014). "Bayesian

nonparametric instrumental variables regression based on penalized splines and Dirichlet process mixtures." *Journal of Business & Economic Statistics*, 32(3): 468–482. 2

Yuan, M. and Lin, Y. (2006). "Model selection and estimation in regression with grouped variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1): 49–67. 3

Zhang, J., Clayton, M. K., and Townsend, P. A. (2011). "Functional concurrent linear regression model for spatial images." *Journal of Agricultural, Biological, and Environmental Statistics*, 16(1): 105–130. 2

Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). "A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses." *NeuroImage*, 95: 162–175. 2