COVARIATE-GUIDED BAYESIAN MIXTURE OF SPLINE EXPERTS FOR THE ANALYSIS OF MULTIVARIATE TIME SERIES

Haoyi Fu
Department of Biostatistics
University of Pittsburgh
Pittsburgh, PA, USA
haf48@pitt.edu

Lu Tang
Department of Biostatistics
University of Pittsburgh
Pittsburgh, PA, USA

Ori Rosen
Department of Mathematical Sciences
University of Texas at El Paso
El Paso, TX, USA

Alison E. Hipwell
Department of Psychiatry
University of Pittsburgh
Pittsburgh, PA, USA

Theodore J. Huppert
Department of Electrical and Computer Engineering
University of Pittsburgh
Pittsburgh, PA, USA

Robert T. Krafty
Department of Biostatistics and Bioinformatics
Emory University
Atlanta, GA, USA

ABSTRACT

With rapid development of techniques to measure brain activity and structure, statistical methods for analyzing modern brain-imaging play an important role in the advancement of science. Imaging data that measure brain function are usually multivariate time series and are heterogeneous across both imaging sources and subjects, which lead to various statistical and computational challenges. In this paper, we propose a group-based method to cluster a collection of multivariate time series via a Bayesian mixture of smoothing splines. Our method assumes each multivariate time series is a mixture of multiple components with different mixing weights. Time-independent covariates are assumed to be associated with the mixture components and are incorporated via logistic weights of a mixture-of-experts model. We formulate this approach under a fully Bayesian framework using Gibbs sampling where the number of components is selected based on a deviance information criterion. The proposed method is compared to existing methods via simulation studies and is applied to a study on functional near-infrared spectroscopy (fNIRS), which aims to understand infant emotional reactivity and recovery from stress. The results reveal distinct patterns of brain activity, as well as associations between these patterns and selected covariates.

***K*eywords** Bayesian mixture model · Brain-imaging · Functional near-infrared spectroscopy · Model-based clustering · Multivariate time series · Smoothing splines · Face-to-face still-face

1 Introduction

Time series are realizations of random processes. Obtaining estimated time series trajectories may provide insights into many practical problems. Functional near-infrared spectroscopy (fNIRS) is a noninvasive brain imaging technique that measures changes in both oxy- and deoxy-hemoglobin using near-infrared light (Jobsis, 1977). In fNIRS, processed data are nonstationary multivariate time series with a non-constant mean and high variability across time, which pose many statistical challenges in inference and estimation. In the case of fNIRS, different subjects could have distinct patterns of multivariate time series trajectories, which could be associated with certain clinical or demographic characteristics.

The analysis of fNIRS data requires an appropriate method for the analysis of a collection of multivariate time series observed from different subjects, which is often referred to as a replicated multivariate time series setting.

Cluster analysis is often used to address the issue of heterogeneity and identify subgroups from collections of time series observed from different subjects. Time series clustering has been used in diverse scientific areas to discover trajectory patterns, which can uncover valuable information from complex and massive datasets (Liao, 2005). Time series clustering partitions the entire collection of data into different groups such that homogeneous time series are grouped together based on a certain similarity measure. Challenges in time-series clustering include computational issues due to high-dimensionality and the selection of proper similarity measures (Lin *and others*, 2003; Keogh and Pazzani, 2000). Several authors have proposed clustering algorithms for multivariate time series. Kakizawa *and others* (1998) used Kullback-Leibler discrimination information as the minimum discrimination criterion for clustering multivariate Gaussian time series. Wang *and others* (2007) used a modified K-means clustering algorithm for clustering multivariate time series based on univariate structures. A variety of papers have established different model-based clustering methods for clustering multivariate time series, such as multivariate autoregressive models (Maharaj, 1999; He *and others*, 2022), a hidden Markov model (Li *and others*, 2001) and smoothing splines (Krafty *and others*, 2017; Li and Krafty, 2019). Comprehensive review of methods for time series clustering can be found in Liao (2005) and in Maharaj *and others* (2019).

Covariate-dependent structures can often be associated with the mixture components from a clustering of time series. Bertolacci *and others* (2022) presented an analysis of multiple nonstationary time series by using a covariate-dependent infinite mixture with logistic stick-breaking weights, where mixing weights are computed based on covariates. The mixture-of-experts model (Jacobs *and others*, 1991) assigns weights to each expert via a covariate-dependent multinomial logists. Huerta *and others* (2003) addressed the issue of time series model mixing based on covariates using the hierarchical mixture-of-experts (Jordan and Jacobs, 1994).

Smoothing splines, which are nonparametric methods that utilize roughness-based penalties, have been widely used in the analysis of time series (Wang, 2011; Gu, 2013). Bayesian interpretations of smoothing splines were first discussed by Kimeldorf and Wahba (1970). Wahba (1978) showed that the solution to the smoothing splines objective function is equivalent to Bayesian estimation with a partially diffuse prior. Speckman and Sun (2003) adopted a fully Bayesian approach for implementing smoothing splines with a noninformative prior on the variance component, as well as derived necessary and sufficient conditions for the propriety of the posterior. Smoothing splines require estimation of a large number of coefficients, which might be impractical in high-dimensional settings. Gu and Kim (2002) used a subset of reproducing kernel functions to achieve a low-dimensional approximation. Wood *and others* (2002) obtained a subset of basis functions using the eigen-decomposition of the Gaussian kernel. Krafty *and others* (2017) proposed a tensor-product model for the analysis of replicated multivariate time series which decomposes the power spectrum into products of univariate outcomes and frequencies.

Our goal in this paper is to perform a covariate-guided clustering of multivariate time series that can capture trajectory patterns of mixture components and evaluate the relationship between covariates and trajectory patterns. To this end, each mixture component is modeled via smoothing splines, and time-independent covariates are incorporated into the mixture model via the mixing weights. The method is formulated in a fully Bayesian framework. The rest of this paper is organized as follows. In Section 2 we introduce the motivating study. Sections 3 and 4 present the proposed model and priors. Section 5 introduces the sampling scheme. In Section 6 we report simulation results under different settings and Section 7 illustrates our proposed method with application to the motivating study. Section 8 concludes the paper with a discussion.

2 Motivating Study

Our motivating study aims to understand patterns of infant's brain activity before, during and after an emotionally stressful probe called face-to-face still-face (FFSF) (Tronick *and others*, 1978). Participant mothers in this study were recruited from the longitudinal Pittsburgh Girls Study (PGS), a population-based study of 2,450 girls who were recruited in the city of Pittsburgh between the ages of 5 and 8 (Keenan *and others*, 2010). In 2016, a large-scale sub-study of the PGS was initiated to investigate how environmental factors, such as psychological stressors experienced during childhood and adolescence, affect later maternal pregnancy and child health. The study is part of the National Institutes of Health Environmental Influences of Child Health Outcomes (ECHO) program, which examines different impacts of prenatal environmental exposures across biological, chemical, physical and social domains on offspring health and development (Gillman and Blaisdell, 2018). The PGS-ECHO study enrolls PGS participants as they become pregnant or recently deliver a live birth. Participants complete multiple prenatal lab visits and the children are followed from ages 6 to 36 months. The lab protocol includes interviews and interaction tasks to assess contextual stressors, health, mood, lifestyle behaviors and offspring behavioral and emotional development.

Face-to-face interactions between mothers and infants are essential to the development of infants with respect to communication and social skills, as well as the regulation of emotion and temperament (Hipwell *and others*, 2019). The FFSF paradigm is a widely used stress task (a violation of the expectation of social interaction) that allows for biobehavioral measurement of individual differences in infant response and recovery. The FFSF comprises of three phases: interact (or baseline), still-face and recovery (Adamson and Frick, 2003). In phase 1, mothers perform normal interactions with infants without the use of toys; this phase serves as the baseline. In phase 2, mothers adopt a neutral facial expression (still-face with no facial or oral communication) to infants, followed by phase 3, where mothers resume normal interactions with their infants. Prior to the start of the FFSF, an fNIRS cap is fitted on the infant's head to measure the level of and change in brain activation across the three phases.

PGS-ECHO fNIRS still-face data are recorded using a continuous NIRS imaging system (NIRScout; NIRx Medical Technologies, Berlin, Germany) at the sampling rate of 7.8125 Hz and using the NIRStart acquisition software. The data are measured simultaneously at two wavelengths (760 nm and 850 nm). As shown in Figure 1(a), this fNIRS probe consists of 12 channels from 8 sources and 4 detectors.

In the current study, we measured infant brain activity using the above fNIRS probe (roughly 120 seconds of measurements for each phase). At the end of 2021, recorded fNIRS still-face data had been collected from 155 infant subjects. Demographic variables of infants and mothers such as gestational age, infant age, sex, birth weight, head circumference, along with parent reports on the Infant Behavior Questionnaire-Revised (IBQ-R) (Gartstein and Rothbart, 2003) were also collected. By removing infants who did not complete the three phases of the still-face paradigm, who had large outliers based on leverage and who had a very short period of measurements in any of the three still-face phases, there were a total of 82 subjects with complete fNIRS still-face data available for future analysis. The above quality control steps were performed by the NIRS brain AnalyzIR toolbox in MATLAB (Santosa *and others*, 2018). Moreover, additional data pre-processing steps were performed in R software, including data interpolation and rescaling. Finally, processed fNIRS data had a total of 1,500 measurement points for each subject and each channel, where each phase consisted of 500 points. All measurements and sampling times were rescaled to be between 0 and 1, with the interact phase occurring between time 0 to 1/3, still-face between 1/3 to 2/3, and recovery between 2/3 to 1. An example of processed fNIRS time series from two selected subjects and four selected channels is displayed in Figure 2.

The goals of our analysis are to identify distinct patterns of brain activity trajectories from multiple fNIRS channels represented by the relative concentration of oxy-hemoglobin, and to assess the association between trajectory patterns and relevant covariates.

3 Model

In this section, we provide a detailed description of our proposed covariate-guided Bayesian mixture of spline experts model. The proposed model consists of spline components whose mixing weights depend on covariates.

3.1 Mixture of splines model

We propose a tensor-product mixture of splines model for multivariate time series. For each subject $i = 1, \ldots, N$, let $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}', \ldots, \boldsymbol{y}_{ik}', \ldots, \boldsymbol{y}_{iK}')'$ be the nK-vector corresponding to the K-dimensional time series for $k = 1, \ldots, K$, where $\boldsymbol{y}_{ik} = \big[y_{ik}(t_1), \ldots, y_{ik}(t_j), \ldots, y_{ik}(t_n)\big]'$ contains the trajectory of measurements on the kth entry of the time series evaluated over a grid of n time points for $j = 1, \ldots, n$, and $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}', \ldots, \boldsymbol{\epsilon}_{iK}')'$ is the nK-vector of errors. Following the model representation of Krafty *and others* (2017), the tensor-product model for the K-dimensional multivariate time series, conditional on component g, $g = 1, \ldots, G$, can be written as:

$$\{\boldsymbol{y}_i \mid z_{ig} = 1\} = (\boldsymbol{I}_K \otimes \boldsymbol{X})\boldsymbol{\alpha}_g + (\boldsymbol{I}_K \otimes \boldsymbol{W})\boldsymbol{\beta}_g + \boldsymbol{\epsilon}_i, \tag{1}$$

where $\{z_{ig}\}_{g=1}^{G}$ are latent indicators as described in Section 3.3, $\boldsymbol{\alpha}_g = (\boldsymbol{\alpha}_{g1}', \ldots, \boldsymbol{\alpha}_{gK}')'$ is a $2K$-vector of intercepts and slopes, $\boldsymbol{\beta}_g = (\boldsymbol{\beta}_{g1}', \ldots, \boldsymbol{\beta}_{gK}')'$ is a mK-vector of basis function coefficients as described in Section 4.1, $\boldsymbol{I}_K$ is a $K \times K$ identity matrix and $\otimes$ denotes a tensor product. The matrix $\boldsymbol{X}$ is given by $\boldsymbol{X} = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ t_1 & t_2 & \ldots & t_n \end{pmatrix}'$ and the m columns of the matrix $\boldsymbol{W}$ are smoothing splines basis functions as described in Section 4.1. We assume the error vector $\boldsymbol{\epsilon}_i$ follows a $\text{MVN}(\boldsymbol{0}, \boldsymbol{\Psi}_g \otimes \boldsymbol{U})$ distribution, where $\boldsymbol{U} = \boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $\boldsymbol{\Psi}_g = \text{diag}(\boldsymbol{\sigma}_g^2)$ is a $K \times K$ diagonal matrix with the error variances $\boldsymbol{\sigma}_g^2 = (\sigma_{g1}^2, \ldots, \sigma_{gK}^2)'$. We assume each subject has a common grid of time points across all K entries, such that $\boldsymbol{X}$ and $\boldsymbol{W}$ are common to all subjects, although our proposed method can be generalized to the case where subjects are observed at different grids of time points. In addition, we assume $\boldsymbol{E}(\boldsymbol{y}_{ik}, \boldsymbol{y}_{ih}) = \boldsymbol{0}_{n \times n}$ for $k \neq h$.

To simplify notation, we let $\boldsymbol{S} = [\boldsymbol{X}\ \boldsymbol{W}]$ and $\boldsymbol{\theta}_g = (\boldsymbol{\alpha}'_{g1}, \boldsymbol{\beta}'_{g1}, \ldots, \boldsymbol{\alpha}'_{gK}, \boldsymbol{\beta}'_{gK})'$. Equation (1) can then be rewritten as:

$$\{\boldsymbol{y}_i \mid z_{ig} = 1\} = (\boldsymbol{I}_K \otimes \boldsymbol{S})\boldsymbol{\theta}_g + \boldsymbol{\epsilon}_i. \tag{2}$$

3.2 Model for the mixing weights

The mixture-of-experts model (Jacobs *and others*, 1991) is applied to form a covariate-guided structure for our proposed model, where the mixing weights are multinomial logits that are functions of selected covariates. As in Sun *and others* (2007), the mixing weights are expressed as

$$\pi_{ig}(\boldsymbol{V}_i) = \frac{\exp(\boldsymbol{V}'_i\boldsymbol{\delta}_g + \zeta_{ig})}{\sum_{h=1}^{G}\exp(\boldsymbol{V}'_i\boldsymbol{\delta}_h + \zeta_{ih})}, \tag{3}$$

where $\boldsymbol{V}_i = (1, V_{i1}, \cdots, V_{iP})'$ is a vector of length $(P+1)$ containing values of P covariates for subject i, and $\boldsymbol{\delta}_g = (\delta_{g0}, \delta_{g1}, \cdots, \delta_{gP})'$ is the corresponding coefficient vector. For identifiability, we set $\boldsymbol{\delta}_G = \boldsymbol{0}$. Equation (3) differs slightly from the weights in the traditional mixture of experts model in that it includes a random term ζ_{ig} for each subject. This term accounts for unmeasured factors beyond the observed covariates, and enhances model performance and inference of the mixing weights.

3.3 Augmented likelihood

To account for heterogeneity across subjects, we assume that the kth entry of the multivariate time series, $\boldsymbol{y}_{ik}$, comes from a mixture model with G components, i.e.,

$$\boldsymbol{y}_{ik} \sim \sum_{g=1}^{G}\pi_{ig}f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\mu}_{gk}, \sigma^2_{gk}\boldsymbol{I}_n), \tag{4}$$

where $f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\mu}_{gk}, \sigma^2_{gk}\boldsymbol{I}_n)$ is the probability density function of the multivariate normal distribution with mean vector $\boldsymbol{\mu}_{gk} = \boldsymbol{X}\boldsymbol{\alpha}_{gk} + \boldsymbol{W}\boldsymbol{\beta}_{gk}$ and covariance matrix $\sigma^2_{gk}\boldsymbol{I}_n$ for the gth component and the kth entry. The π_{ig} are mixing weights that depend on covariates as described in Section 3.2.

As is common in mixture models, augmenting the likelihood with latent variables indicating the component from which a time series originates simplifies the computation greatly (Dempster *and others*, 1977). In particular, let $z_{ig} = 1$ if the ith multivariate time series belongs to the gth component and $z_{ig} = 0$, otherwise. Let $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)'$ be all observed multivariate time series and $\boldsymbol{\Theta}_{gk}$ be the aggregation of all parameters for component g and entry k. The parameter vector for all components and all entries is then denoted by $\boldsymbol{\Theta} = (\boldsymbol{\Theta}'_{11}, \ldots, \boldsymbol{\Theta}'_{GK})'$. The augmented likelihood of all N multivariate time series is given by

$$L(\boldsymbol{\Theta} \mid \boldsymbol{y}, Z) = \prod_{i=1}^{N}\prod_{g=1}^{G}\Big[\pi_{ig}\prod_{k=1}^{K}f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{gk})\Big]^{z_{ig}}, \tag{5}$$

where $f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{gk})$ is the probability density function as appeared in the (4). From Bayes' rule, the distribution of the latent indicators z_{ig} is given by

$$p(z_{ig} = 1 \mid \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\Theta}, \pi_{ig}) = \frac{\pi_{ig}\prod_{k=1}^{K}f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{gk})}{\sum_{h=1}^{G}\pi_{ih}\prod_{k=1}^{K}f_{hk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{hk})}. \tag{6}$$

4 Priors

In this section, the priors on the model parameters are introduced.

4.1 Smoothing splines prior

The conditional expectation of a mixture component in model (4) is given by $E(\boldsymbol{y}_{ik} \mid z_{ig} = 1) = \boldsymbol{X}\boldsymbol{\alpha}_{gk} + \boldsymbol{W}\boldsymbol{\beta}_{gk}$. We place a smoothing spline prior on $\boldsymbol{\beta}_{gk}$ and let $\boldsymbol{\mathcal{H}}_{gk} = \boldsymbol{W}\boldsymbol{\beta}_{gk}$, where $\boldsymbol{\mathcal{H}}_{gk} = \big[\mathcal{H}_{gk}(t_1), \ldots, \mathcal{H}_{gk}(t_n)\big]'$ is a zero-mean Gaussian process with variance covariance matrix $\tau^2_{gk}\boldsymbol{\Phi}$ (Wahba, 1980; Wood *and others*, 2002), such that $\text{cov}\big[\mathcal{H}_{gk}(t_r), \mathcal{H}_{gk}(t_h)\big] = \tau^2_{gk}\phi_{rh}$, τ^2_{gk} is a smoothing parameter for component g and entry k, and the (r,h)th element

of $\boldsymbol{\Phi}$ is given by $\phi_{rh} = \frac{1}{2}t_r^2(t_h - \frac{t_r}{3})$ for $t_r \leq t_h$. The matrix $\boldsymbol{\Phi}$ is common to all subjects since all entries of the multivariate time series are observed at common time points.

As seen above, the matrix $\boldsymbol{\Phi}$ is $n \times n$, and to avoid the computational burden for large n, a low-rank approximation is often adopted. To facilitate this approximation, we obtain basis functions via the spectral decomposition of $\boldsymbol{\Phi}$, as has been proposed in Wood *and others* (2002) and used in Rosen *and others* (2009, 2012); Krafty *and others* (2011). In particular, the matrix $\boldsymbol{W}$ consists of m basis functions evaluated at times $t_1, \ldots, t_n$, and $\boldsymbol{\beta}_{gk}$ is an m-dimensional vector of basis function coefficients. These basis functions are obtained by applying the spectral decomposition to $\boldsymbol{\Phi}$ such that $\boldsymbol{\Phi} = \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is the matrix of eigenvectors of $\boldsymbol{\Phi}$, and $\boldsymbol{\Gamma}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{\Phi}$. We then let the design matrix $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{\Gamma}^{1/2}$ and place a normal prior $N(0, \tau_{gk}^2\boldsymbol{I}_n)$ on $\boldsymbol{\beta}_{gk}$, which leads to $\boldsymbol{\mathcal{H}}_{gk}$ or $\boldsymbol{W}\boldsymbol{\beta}_{gk} \sim N(\boldsymbol{0}, \tau_{gk}^2\boldsymbol{\Phi})$ as mentioned above.

By using the low-rank approximation, the number of columns of $\boldsymbol{W}$ is reduced from n to m ($m < n$), which greatly reduces the computational burden without sacrificing the model fit (Wahba, 1980; Wood, 2006). Eubank (1999) indicated that the eigenvalues in the diagonal matrix $\boldsymbol{\Gamma}$ decay rapidly as m increases. Thus, we can achieve a good approximation by selecting a relatively small number m of basis functions. The number of basis functions m is set to 10 in simulation studies as described in Section 6, which has been shown (Krafty *and others* 2011) to explain more than 98% of the total variability.

The prior on $\boldsymbol{\theta}_g$ is thus $\boldsymbol{\theta}_g \sim N(\boldsymbol{0}, \boldsymbol{D}_g)$, where $\boldsymbol{D}_g = \text{diag}(\sigma_{\alpha 1}^2\boldsymbol{1}_2,\ \tau_{g1}^2\boldsymbol{1}_m,\ \ldots\ , \sigma_{\alpha K}^2\boldsymbol{1}_2,\ \tau_{gK}^2\boldsymbol{1}_m)$ is the covariance matrix of $\boldsymbol{\theta}_g$. The vector $(\sigma_{\alpha 1}^2, \ldots, \sigma_{\alpha K}^2)'$ contains fixed prior variances for the regression coefficients $\boldsymbol{\alpha}_{gk}$, common to all components and entries. In particular, we fix the common prior variance $\sigma_\alpha^2 = 100$. The vector $\boldsymbol{\tau}_g^2 = (\tau_{g1}^2, \ldots, \tau_{gK}^2)'$ contains the smoothing parameters for the gth mixture component and $\boldsymbol{1}_m$ is an m-vector of ones. We assume independence between the regression coefficients $\boldsymbol{\alpha}_{gk}$ and the basis function coefficients $\boldsymbol{\beta}_{gk}$.

4.2 Priors on the smoothing parameters

We assume the smoothing parameters $\boldsymbol{\tau}_g^2 = (\tau_{g1}^2, \ldots, \tau_{gK}^2)'$ vary across components g and entries k. Although the most common choice for the prior on a variance parameter is the inverse gamma distribution, Gelman (2006) and Wand *and others* (2011) suggested that a half-t prior on the standard deviation can reflect lack of information on a scale parameter. The half-t is a family of heavy-tailed distributions and has a good shrinkage performance. It can be expressed as a scale mixture of inverse gamma random variables using a latent variable which follows an inverse gamma distribution (Wand *and others*, 2011). Thus, we assume a half-t distribution such that $\tau_{gk} \sim t_{\nu_\tau}^+(0, A_\tau)$, where ν_τ is a degrees of freedom parameter, and A_τ is a scale parameter. We set $\nu_\tau = 3$ and $A_\tau = 10$ for all components and entries.

4.3 Priors on the error variances

We assume $\sigma_{gk} \overset{\text{i.i.d}}{\sim} t_{\nu_\sigma}^+(0, A_\sigma)$ and set $\nu_\sigma = 3$ and $A_\sigma = 10$ for all components and entries.

4.4 Priors on the logistic parameters and the variances of random intercepts

This section provides details on the prior distributions placed on the parameters of the logistic weights (3). For ease of notation, we denote $\boldsymbol{\delta}_g^* = (\boldsymbol{\delta}_g^T, \boldsymbol{\zeta}_g^T)^T$, where $\boldsymbol{\zeta}_g = (\zeta_{1g}, \cdots, \zeta_{Ng})^T$, $g = 1, \ldots, G$. We let $\boldsymbol{V}_i^* = (\boldsymbol{V}_i', \boldsymbol{e}_i')'$ where $\boldsymbol{e}_i$ is a vector of all zeros except for a single 1 in the ith position, and $\boldsymbol{V}^*$ is a matrix consisting of the rows $\boldsymbol{V}_i^{*T}$, $i = 1, \ldots, N$. Gaussian priors are placed on the logistic parameters, i.e., $\boldsymbol{\delta}_g^* \sim N(\boldsymbol{0}, \boldsymbol{B}_g)$, where $\boldsymbol{B}_g = \text{diag}(\sigma_{\delta\text{g}}^2\boldsymbol{1}_{\text{P}+1},\ \kappa_{\zeta\text{g}}^2\boldsymbol{1}_{\text{N}})$, and the priors on the random intercepts satisfy $\boldsymbol{\zeta}_g \sim N(\boldsymbol{0}, \kappa_{\zeta g}^2\boldsymbol{I}_N)$. As for the hyperparameters, we assume $\sigma_{\delta g}^2 = 10$ for all components and covariates, and $\kappa_{\zeta g} \sim t_{\nu_\kappa}^+(0, A_\kappa)$, where $\nu_\kappa = 3$ and $A_\kappa = 10$ for all components.

To sample the logistic parameters, Polson *and others* (2013) proposed a data augmentation scheme incorporating Pólya-Gamma latent variables, which facilitates Gibbs steps. Details on sampling the logistic parameters are provided in the Supplementary Material.

5 Sampling scheme

This section outlines the Gibbs steps for sampling from the conditional posterior distributions of all the model parameters. More details are given in Supplementary Material.

5.1 Gibbs sampling steps

Letting ℓ denote the current Gibbs sampling iteration, parameter values at the $(\ell+1)$th iteration are drawn according to the following steps.

1. Draw $\boldsymbol{\theta}_{gk}^{(\ell+1)}$ from $(\boldsymbol{\theta}_{gk}^{(\ell+1)} \mid \boldsymbol{y}, \boldsymbol{S}, \tau_{gk}^{2(\ell)}, \sigma_{gk}^{2(\ell)}) \sim N(\boldsymbol{u}_{gk}, \sigma_{gk}^2 \boldsymbol{\Lambda}_{gk})$, where $\boldsymbol{u}_{gk}$ and $\boldsymbol{\Lambda}_{gk}$ are mean vectors and covariance matrices.
2. Draw $\sigma_{gk}^{2(\ell+1)}$ from $(\sigma_{gk}^{2(\ell+1)} \mid \boldsymbol{\epsilon}_{igk}^{(\ell+1)}, a_{\sigma_{gk}}^{(\ell+1)}) \sim IG\Big((nN_g^{(\ell)} + \nu_\sigma)/2, \sum_{i=1}^{N} z_{ig}\boldsymbol{\epsilon}_{igk}'\boldsymbol{\epsilon}_{igk}/2 + \nu_\sigma/a_{\sigma_{gk}}\Big)$, where $N_g^{(\ell)}$ is the current number of subjects in the gth component, $\boldsymbol{\epsilon}_{igk}$ is the error vector for the gth component, the ith subject and the kth entry, and $a_{\sigma_{gk}}$ is a latent variable in the IG scale mixture underlying the half-t distribution.
3. Draw $\tau_{gk}^{2(\ell+1)}$ from $(\tau_{gk}^{2(\ell+1)} \mid \boldsymbol{\beta}_{gk}^{(\ell+1)}, a_{\tau_{gk}}^{(\ell+1)}) \sim IG\Big((\nu_\tau + m)/2, \boldsymbol{\beta}_{gk}'\boldsymbol{\beta}_{gk}/2 + \nu_\tau/a_{\tau_{gk}}\Big)$, where $a_{\tau_{gk}}$ is a latent variable as in 2.
4. Draw $\boldsymbol{\delta}_g^{*(\ell+1)}$ from $(\boldsymbol{\delta}_g^{*(\ell+1)} \mid \boldsymbol{V^*}, z_{ig}^{(\ell)}, \omega_{ig}^{(\ell+1)}, \kappa_{\zeta g}^{2(\ell)}) \sim N(\boldsymbol{M}_g, \boldsymbol{\Sigma}_g)$, where $\omega_{ig}^{(\ell+1)}$ is a Pólya-Gamma latent variable in the augmentation described in Section 4.4.
5. Draw $\kappa_{\zeta g}^{2(\ell+1)}$ from $(\kappa_{\zeta g}^{2(\ell+1)} \mid \boldsymbol{\zeta}_g^{(\ell+1)}, a_{\kappa_g}^{(\ell+1)}) \sim IG\Big(\nu_\kappa/2, \boldsymbol{\zeta}_g'\boldsymbol{\zeta}_g/2 + (\nu_\kappa + N)/a_{\kappa_g}\Big)$, where a_{κ_g} is a latent variable as in 2 and 3.
6. The mixing weights $\pi_{ig}^{(\ell+1)}$ are obtained by computing $p(\pi_{ig}^{(\ell+1)} \mid \boldsymbol{V}^*, \boldsymbol{\delta}_g^{*(\ell+1)}, z_{ig}^{(\ell)})$ from Equation (3).
7. Draw $z_{ig}^{(\ell+1)} \sim p(z_{ig}^{(\ell+1)} = 1 \mid \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\theta}_{gk}^{(\ell+1)}, \sigma_{gk}^{2(\ell+1)}, \pi_{ig}^{(\ell+1)})$ according to Equation (6).

5.2 Selecting the number of components

Spiegelhalter *and others* (2002) suggested the use of the deviance information criterion (DIC) for model selection based on the effective number of parameters. Gelman *and others* (2003) introduced an alternative measure of effective number of parameters based on the variance of the log predictive density across MCMC iterations. This measure is robust and more accurate than the original one. Moreover, it has the advantages of always being positive and invariant to reparameterizations (Gelman *and others*, 2003).

In this paper, we use DIC to select the number of components for our proposed mixture model.

6 Simulation studies

To demonstrate the performance of the proposed method, we conduct simulation studies by generating data sets from the proposed model under two scenarios: two-component mixture ($G = 2$) of trivariate time series ($K = 3$) and four-component mixture ($G = 4$) of bivariate time series ($K = 2$). We simulate 100 replicates in each simulation setting with $N = 150$ time series of length $n = 50$. A total of $20,000$ Gibbs sampling iterations are run with a burn-in of $4,000$. In all simulation settings, the hyperparameters are assigned the same values, given in Section 4.

6.1 Two-component trivariate model

In this scenario, we consider the two-component trivariate model. From Equation (1), the gth component of the proposed mixture model is given by

$$\{\boldsymbol{y}(t_j) \mid z_{ig} = 1\} = \boldsymbol{\alpha}_{0g} + \boldsymbol{\alpha}_{1g}t_j + \sum_{q=1}^{m} w_q(t_j)\boldsymbol{\beta}_{gq} + \boldsymbol{\epsilon}_{gt_j}, \quad j = 1, \ldots, n, \;\; g = 1, \ldots, G, \tag{7}$$

where $\boldsymbol{y}(t_j)$ is the trivariate time series evaluated at time t_j, $\boldsymbol{\alpha}_{01} = (1, -3, -2)'$, $\boldsymbol{\alpha}_{02} = (5, 4, 3)'$ and $\boldsymbol{\alpha}_{11} = (-2, 2, 0.5)'$, $\boldsymbol{\alpha}_{12} = (1, -1, -0.5)'$ are independent intercepts and slopes for each component, respectively. The vector $\boldsymbol{\beta}_{gq}$ consists of the qth spline coefficients of all variates for component g, and $w_q(t_j)$ is the qth spline basis function evaluated at time t_j. The $\boldsymbol{\epsilon}_{gt_j}$ are independent zero-mean error terms, distributed as $\boldsymbol{\epsilon}_{gt_j} \sim \text{MVN}\Big(\boldsymbol{0}, \text{diag}(\sigma_{g1}^2, \sigma_{g2}^2, \sigma_{g3}^2)\Big)$, where $\sigma_1^2 = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)' = (3, 5, 4.5)'$ and $\sigma_2^2 = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)' = (4, 3.5, 4)'$. The smoothing parameters are set to $\tau_1^2 = (\tau_{11}^2, \tau_{12}^2, \tau_{13}^2)' = (3.5, 5, 8.5)'$ and $\tau_2^2 = (\tau_{21}^2, \tau_{22}^2, \tau_{23}^2)' = (6, 2.5, 1.5)'$.

We investigate the performance of the trajectory and logistic parameter (see Equation (3)) estimates. For the former, we calculate the averaged root square error (ARSE) of each mixture component g

$$\text{ARSE}_g = \sqrt{\frac{1}{nK}\sum_{j=1}^{n}\sum_{k=1}^{K}\Big[\mu_{gk}(t_j) - \hat{\mu}_{gk}(t_j)\Big]^2},$$

where $\mu_{gk}(t_j)$ is the expectation of $y_k(t_j)$ according to the gth component, and $y_k(t_j)$ is the kth entry of the time series evaluated at time t_j. The $\hat{\mu}_{gk}(t_j)$ are the estimated posterior means of $\mu_{gk}(t_j)$ for $k = 1, \ldots, K$ and $j = 1, \ldots, n$.

To handle a potential label switching across mixture components, we compute ARSE_g as the minimum value across all components, by using the estimate of the gth component and the truth of each group, $g = 1, \ldots, G$. After obtaining correct component labels by evaluating ARSE, we also report the averaged bias (A-bias) and the variance of the bias (V-bias) of each mixture component g, where

$$\text{A-bias}_g = \frac{1}{nK}\sum_{j=1}^{n}\sum_{k=1}^{K}\Big[\hat{\mu}_{gk}(t_j) - \mu_{gk}(t_j)\Big],$$

and V-bias_g is computed by calculating the sample variance of the bias over entries and time points.

For each replicate, time series trajectories are estimated by three methods: the proposed method, the `R` package `gbmt` (Magrini, 2022) and the `TRAJ` procedure in `SAS` (Nagin *and others*, 2018). Boxplots of ARSE, A-bias and V-bias of each component are given in Figure 3. Notably, `TRAJ` is able to fit a regression spline model by treating basis functions as time-varying covariates, while `gbmt` is only able to fit a cubic model. Our proposed method fits a penalized spline model under the Bayesian framework and is able to outperform both `gbmt` and `TRAJ` in terms of ARSE and V-bias for both components. A-biases are close to zero and comparable for all three methods. These findings demonstrate that all three methods are able to achieve a reasonable fit to group-based trajectories since bias over the entire time series is close to zero. Our proposed method is able to obtain more precise estimates of trajectories as is evident from the smaller V-biases.

To evaluate the performances of the logistic parameters, we compute the root mean squared error (RMSE) for each logistic parameter using the proposed method and `TRAJ`. Notably, `gbmt` is not able to incorporate covariates into the computation of mixing weights. Results of RMSEs of each logistic parameter are given in Table 1. We also compare RMSEs between the proposed method and `TRAJ` under four settings of different combinations of $N = 150, 250$ and $n = 50, 70$. Our proposed method yields smaller RMSEs of the logistic parameters in all cases, especially for the intercept δ_0 and the first covariate δ_1. This is to be expected since `TRAJ` uses a multinomial logistic model, which may result in inflated parameter estimates in cases of unbalanced outcomes or perfect separation, while our proposed method is able to obtain a shrinkage result using the penalization method.

6.2 Four-component bivariate model

In this scenario, we consider the four-component bivariate model whose gth component is given in Equation (7), where the values of the intercepts and slopes are $\boldsymbol{\alpha}_{01} = (1, -2)'$, $\boldsymbol{\alpha}_{02} = (5, 3)'$, $\boldsymbol{\alpha}_{03} = (-3, 5.5)'$, $\boldsymbol{\alpha}_{04} = (4, -1)'$, $\boldsymbol{\alpha}_{11} = (-3, 0)'$, $\boldsymbol{\alpha}_{12} = (2, -3.5)'$, $\boldsymbol{\alpha}_{13} = (2.5, 2)'$ and $\boldsymbol{\alpha}_{14} = (-3, 1.5)'$. By analogy to the two-component trivariate model, the errors $\boldsymbol{\epsilon}_{gt_j}$ are independent zero-mean bivariate Gaussian random variables, distributed as $\boldsymbol{\epsilon}_{gt_j} \sim \text{MVN}\Big(\mathbf{0}, \text{diag}(\sigma^2_{g1}, \sigma^2_{g2})\Big)$, where $\sigma^2_1 = (\sigma^2_{11}, \sigma^2_{12})' = (6, 9)'$, $\sigma^2_2 = (\sigma^2_{21}, \sigma^2_{22})' = (8, 7.5)'$, $\sigma^2_3 = (\sigma^2_{31}, \sigma^2_{32})' = (10, 6.5)'$ and $\sigma^2_4 = (\sigma^2_{41}, \sigma^2_{42})' = (7, 8.5)'$.

The performances of the estimated trajectories and logistic parameters for this scenario are displayed in Figure 4 and Table 2. As in the first scenario, our proposed method outperforms both `gbmt` and `TRAJ` in terms of ARSE and V-bias for all components. Notably, `TRAJ` fails to yield precise estimates in several replicates and thus results in larger mean ARSE and V-bias. In terms of the logistic parameters, the proposed method performs well with smaller RMSEs in almost all cases, especially for δ_0 and δ_1. More simulation results based on different values of N and n under the two scenarios considered above are presented in the Supplementary Material.

7 Real data application

We apply our proposed method to the analysis of the fNIRS still-face study introduced in Section 2. Six covariates are considered in our covariate-guided model, including Infant Behavior Questionnaire-Revised negative emotionality (IBQ-NE) score, Infant Behavior Questionnaire-Revised effortful control (IBQ-EC) score, gestational age (in Days),

infant age (in Months), head circumference (in cm) and sex. All continuous covariates are centered and scaled. We set the number of basis functions at $m = 20$ and run a total of $30,000$ Gibbs iterations with a burn-in period of $6,000$. The values of the hyperparameters are the same as the ones used in the simulation studies.

The IBQ-NE construct combines data from the following subscales: Sadness, Distress to Limitations, Fear, and Falling Reactivity/Rate of Recovery from Distress. IBQ-EC refers to the ability to inhibit a dominant response to perform a subdominant one and has been shown to be protective against a myriad of difficulties (Gartstein *and others*, 2013). Finally, the data consist of 79 subjects with complete fNIRS and covariate values. We present results based on analyzing one set of four-channels. Additional results based on analyzing another set of four channels and all channels are given in the Supplementary Material. The four channels are S1D1, S2D2, S5D3 and S6D4. Channels S1D1 and S5D3 are in the central prefrontal region, while channels S2D2 and S6D4 are in the left and right prefrontal region, respectively. We fit our proposed model with the number of components varying from 2 to 6. Based on values of DIC introduced in Section 5.2, the two-component model is selected as the best model for this four-channel analysis.

Figure 5 presents the estimated trajectories of the two-component model fitted to the four channels. We are interested in brain activation signals in the still-face period while the interact period is used as the reference level. For component 1, a decreasing trajectory is observed for the still-face period in all four channels. In contrast, an increasing trend is observed for the still-face period in all four channels for component 2. After fitting the mixture model and finding above trajectory patterns, we define component 1 as the no response component and component 2 as the response component based on trajectory patterns in the still-face period. Figure 6 displays the logistic parameter estimates for all covariates in the 2-component model, where component 2 is used as the reference. There is evidence that IBQ-NE scores differ between the two components as its 95% credible interval does not include zero. A positive coefficient of IBQ-NE indicates that a higher IBQ-NE score is associated with component 1, which has decreased brain activation levels in the still-face period for all four channels. Though other logistic coefficients have 95% credible intervals that include zero, the negative posterior mean estimate of the IBQ-EC score could still indicate that a high IBQ-EC is associated with an increased brain activation as shown for component 2. These conclusions are consistent with findings in Gartstein *and others* (2013) that IBQ-NE is negatively associated with IBQ-EC. Enlow *and others* (2016) reported a negative association between activity level and IBQ-NE among infants whose families encourage a high level of activities. Furthermore, a negative posterior mean of logistic coefficient of infant age suggests that younger infant tends to have a decreasing brain activation level in the still-face period.

8 Discussion

The proposed covariate-guided Bayesian mixture of spline experts model aims to perform a model-based clustering of multivariate time series from multiple subjects. The mixture components in this model are penalized splines, and the mixing weights incorporate covariates. Our proposed method is compared to two commonly used methods through simulation studies which demonstrate a better performance of our method under different scenarios. We apply our proposed method to a fNIRS still-face study and find distinct patterns of components of time series trajectories, as well as an association between IBQ-NE score and a pattern of decreased brain activity in the still-face period. To the best of our knowledge, this is the first still-face study using fNIRS whose purpose is to identify trajectory components.

Our proposed method has some limitations. First, as in any mixture models, label switching may occur, especially in the real-data application. We have adopted the Equivalence Classes Representatives (ECR) algorithm proposed by Papastamoulis and Iliopoulos (2010) to make the components interpretable, but other methods may be considered. Second, the proposed method assumes independence among the entries of the time series and does not allow spatial dependence. Spatial correlations of fNIRS are correlations among fNIRS channels based on the placements and locations of each source and detector. An extension to a multilevel multivariate model would be possible by considering spatial correlations among time series entries. Lastly, our proposed method uses DIC to select the number of components which might be sub-optimal. Bayesian model averaging and reversible jump MCMC (RJMCMC) methods could be considered, but trans-dimensional sampling methods would pose challenges in providing interpretable components.

9 Software

Software in the form of R codes, together with an example data, is available at `https://github.com/HaoyiFu1993/CBMOSE`.

References

ADAMSON, LAUREN B AND FRICK, JANET E. (2003). The still face: A history of a shared experimental paradigm. *Infancy* **4**(4), 451–473.

BERTOLACCI, MICHAEL, ROSEN, ORI, CRIPPS, EDWARD AND CRIPPS, SALLY. (2022). Adaptspec-x: Covariate-dependent spectral modeling of multiple nonstationary time series. *Journal of Computational and Graphical Statistics* **31**(2), 436–454.

DEMPSTER, ARTHUR P, LAIRD, NAN M AND RUBIN, DONALD B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

ENLOW, MICHELLE BOSQUET, WHITE, MATTHEW T, HAILS, KATHERINE, CABRERA, IVAN AND WRIGHT, ROSALIND J. (2016). The infant behavior questionnaire-revised: Factor structure in a culturally and sociodemographically diverse sample in the united states. *Infant Behavior and Development* **43**, 24–35.

EUBANK, RANDALL L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

GARTSTEIN, MARIA A, BRIDGETT, DAVID J, YOUNG, BRANDI N, PANKSEPP, JAAK AND POWER, THOMAS. (2013). Origins of effortful control: Infant and parent contributions. *Infancy* **18**(2), 149–183.

GARTSTEIN, MARIA A AND ROTHBART, MARY K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant behavior and development* **26**(1), 64–86.

GELMAN, ANDREW. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis* **1**(3), 515–534.

GELMAN, A, CARLIN, JB, STERN, HS, RUBIN, DB *and others*. (2003). Bayesian data analysis.

GILLMAN, MATTHEW W AND BLAISDELL, CAROL J. (2018). Environmental influences on child health outcomes, a research program of the nih. *Current opinion in pediatrics* **30**(2), 260.

GU, CHONG. (2013). *Smoothing spline ANOVA models*, Volume 297. Springer.

GU, CHONG AND KIM, YOUNG-JU. (2002). Penalized likelihood regression: General formulation and efficient approximation. *Canadian Journal of Statistics* **30**(4), 619–628.

HE, LINCHEN, WANG, CHAN, HU, JIYUAN, GAO, ZHAN, FALCONE, EMILIA, HOLLAND, STEVEN M, BLASER, MARTIN J AND LI, HUILIN. (2022). Arzimm: A novel analytic platform for the inference of microbial interactions and community stability from longitudinal microbiome study. *Frontiers in genetics* **13**.

HIPWELL, ALISON E, TUNG, IRENE, NORTHRUP, JESSIE AND KEENAN, KATE. (2019). Transgenerational associations between maternal childhood stress exposure and profiles of infant emotional reactivity. *Development and psychopathology* **31**(3), 887–898.

HUERTA, GABRIEL, JIANG, WENXIN AND TANNER, MARTIN A. (2003). Time series modeling via hierarchical mixtures. *Statistica Sinica*, 1097–1118.

JACOBS, ROBERT A, JORDAN, MICHAEL I, NOWLAN, STEVEN J AND HINTON, GEOFFREY E. (1991). Adaptive mixtures of local experts. *Neural computation* **3**(1), 79–87.

JOBSIS, FRANS F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* **198**(4323), 1264–1267.

JORDAN, MICHAEL I AND JACOBS, ROBERT A. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural computation* **6**(2), 181–214.

KAKIZAWA, YOSHIHIDE, SHUMWAY, ROBERT H AND TANIGUCHI, MASANOBU. (1998). Discrimination and clustering for multivariate time series. *Journal of the American Statistical Association* **93**(441), 328–340.

KEENAN, KATE, HIPWELL, ALISON, CHUNG, TAMMY, STEPP, STEPHANIE, STOUTHAMER-LOEBER, MAGDA, LOEBER, ROLF AND MCTIGUE, KATHLEEN. (2010). The pittsburgh girls study: overview and initial findings. *Journal of Clinical Child & Adolescent Psychology* **39**(4), 506–521.

KEOGH, EAMONN J AND PAZZANI, MICHAEL J. (2000). A simple dimensionality reduction technique for fast similarity search in large time series databases. In: *Pacific-Asia conference on knowledge discovery and data mining*. Springer. pp. 122–133.

KIMELDORF, GEORGE S AND WAHBA, GRACE. (1970). A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics* **41**(2), 495–502.

KRAFTY, ROBERT T, HALL, MARTICA AND GUO, WENSHENG. (2011). Functional mixed effects spectral analysis. *Biometrika* **98**(3), 583–598.

KRAFTY, ROBERT T, ROSEN, ORI, STOFFER, DAVID S, BUYSSE, DANIEL J AND HALL, MARTICA H. (2017). Conditional spectral analysis of replicated multiple time series with application to nocturnal physiology. *Journal of the American Statistical Association* **112**(520), 1405–1416.

LI, CEN, BISWAS, GAUTAM, DALE, MIKE AND DALE, PAT. (2001). Building models of ecological dynamics using hmm based temporal data clustering—a preliminary study. In: *International Symposium on Intelligent Data Analysis*. Springer. pp. 53–62.

LI, ZEDA AND KRAFTY, ROBERT T. (2019). Adaptive bayesian time–frequency analysis of multivariate time series. *Journal of the American Statistical Association* **114**(525), 453–465.

LIAO, T WARREN. (2005). Clustering of time series data—a survey. *Pattern recognition* **38**(11), 1857–1874.

LIN, JESSICA, KEOGH, EAMONN AND TRUPPEL, WAGNER. (2003). Clustering of streaming time series is meaningless. In: *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. pp. 56–65.

MAGRINI, ALESSANDRO. (2022). Assessment of agricultural sustainability in european union countries: a group-based multivariate trajectory approach. *AStA Advances in Statistical Analysis*, 1–31.

MAHARAJ, ELIZABETH ANN. (1999). Comparison and classification of stationary multivariate time series. *Pattern Recognition* **32**(7), 1129–1138.

MAHARAJ, ELIZABEH ANN, D'URSO, PIERPAOLO AND CAIADO, JORGE. (2019). *Time Series Clustering and Classification*. chapman and hall/CRC.

NAGIN, DANIEL S, JONES, BOBBY L, PASSOS, VALERIA LIMA AND TREMBLAY, RICHARD E. (2018). Group-based multi-trajectory modeling. *Statistical methods in medical research* **27**(7), 2015–2023.

PAPASTAMOULIS, PANAGIOTIS AND ILIOPOULOS, GEORGE. (2010). An artificial allocations based solution to the label switching problem in bayesian analysis of mixtures of distributions. *Journal of Computational and Graphical Statistics* **19**(2), 313–331.

POLSON, NICHOLAS G, SCOTT, JAMES G AND WINDLE, JESSE. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association* **108**(504), 1339–1349.

ROSEN, ORI, STOFFER, DAVID S AND WOOD, SALLY. (2009). Local spectral analysis via a bayesian mixture of smoothing splines. *Journal of the American Statistical Association* **104**(485), 249–262.

ROSEN, ORI, WOOD, SALLY AND STOFFER, DAVID S. (2012). Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association* **107**(500), 1575–1589.

SANTOSA, HENDRIK, ZHAI, XUETONG, FISHBURN, FRANK AND HUPPERT, THEODORE. (2018). The nirs brain analyzir toolbox. *Algorithms* **11**(5), 73.

SPECKMAN, PAUL L AND SUN, DONGCHU. (2003). Fully bayesian spline smoothing and intrinsic autoregressive priors. *Biometrika* **90**(2), 289–302.

SPIEGELHALTER, DAVID J, BEST, NICOLA G, CARLIN, BRADLEY P AND VAN DER LINDE, ANGELIKA. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)* **64**(4), 583–639.

SUN, ZHUOXIN, ROSEN, ORI AND SAMPSON, ALLAN R. (2007). Multivariate bernoulli mixture models with application to postmortem tissue studies in schizophrenia. *Biometrics* **63**(3), 901–909.

TRONICK, EDWARD, ALS, HEIDELISE, ADAMSON, LAUREN, WISE, SUSAN AND BRAZELTON, T BERRY. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child psychiatry* **17**(1), 1–13.

WAHBA, GRACE. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society: Series B (Methodological)* **40**(3), 364–372.

WAHBA, GRACE. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association* **75**(369), 122–132.

WAND, MATTHEW P, ORMEROD, JOHN T, PADOAN, SIMONE A AND FRÜHWIRTH, RUDOLF. (2011). Mean field variational bayes for elaborate distributions. *Bayesian Analysis* **6**(4), 847–900.

WANG, XIAOZHE, WIRTH, ANTHONY AND WANG, LIANG. (2007). Structure-based statistical features and multivariate time series clustering. In: *Seventh IEEE international conference on data mining (ICDM 2007)*. IEEE. pp. 351–360.

WANG, YUEDONG. (2011). *Smoothing splines: methods and applications*. CRC press.

WOOD, SALLY A, JIANG, WENXIN AND TANNER, MARTIN. (2002). Bayesian mixture of splines for spatially adaptive nonparametric regression. *Biometrika* **89**(3), 513–528.

WOOD, SIMON N. (2006). *Generalized additive models: an introduction with R*. chapman and hall/CRC.

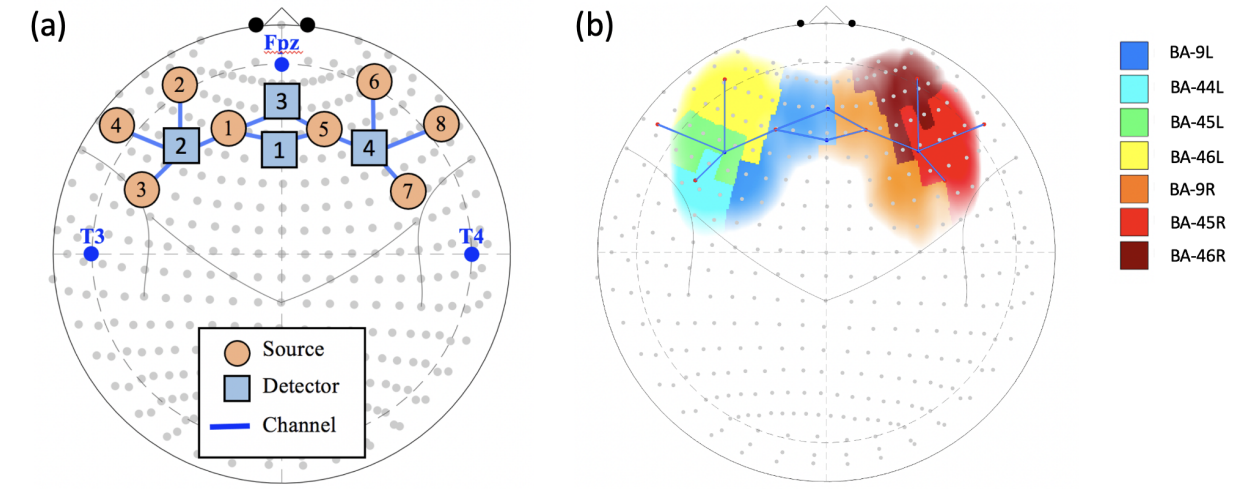

Figure 1: fNIRS probe configuration. (a) Positioning of 8 sources, 4 detectors and 12 channels. A channel is connected by one source and one detector (blue line). (b) Brodmann areas covered by fNIRS probe.

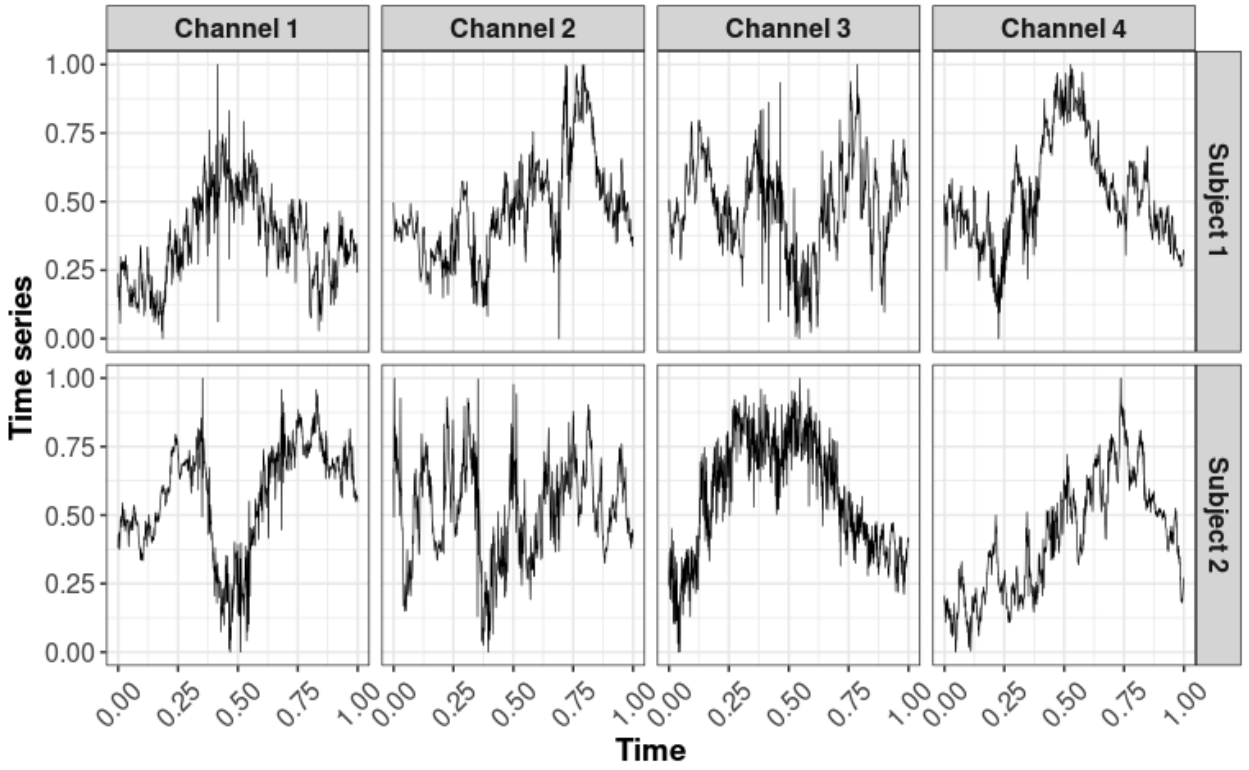

Figure 2: An example of processed fNIRS time series from two selected subjects and four selected channels. The measurements are the relative concentration of oxy-hemoglobin.

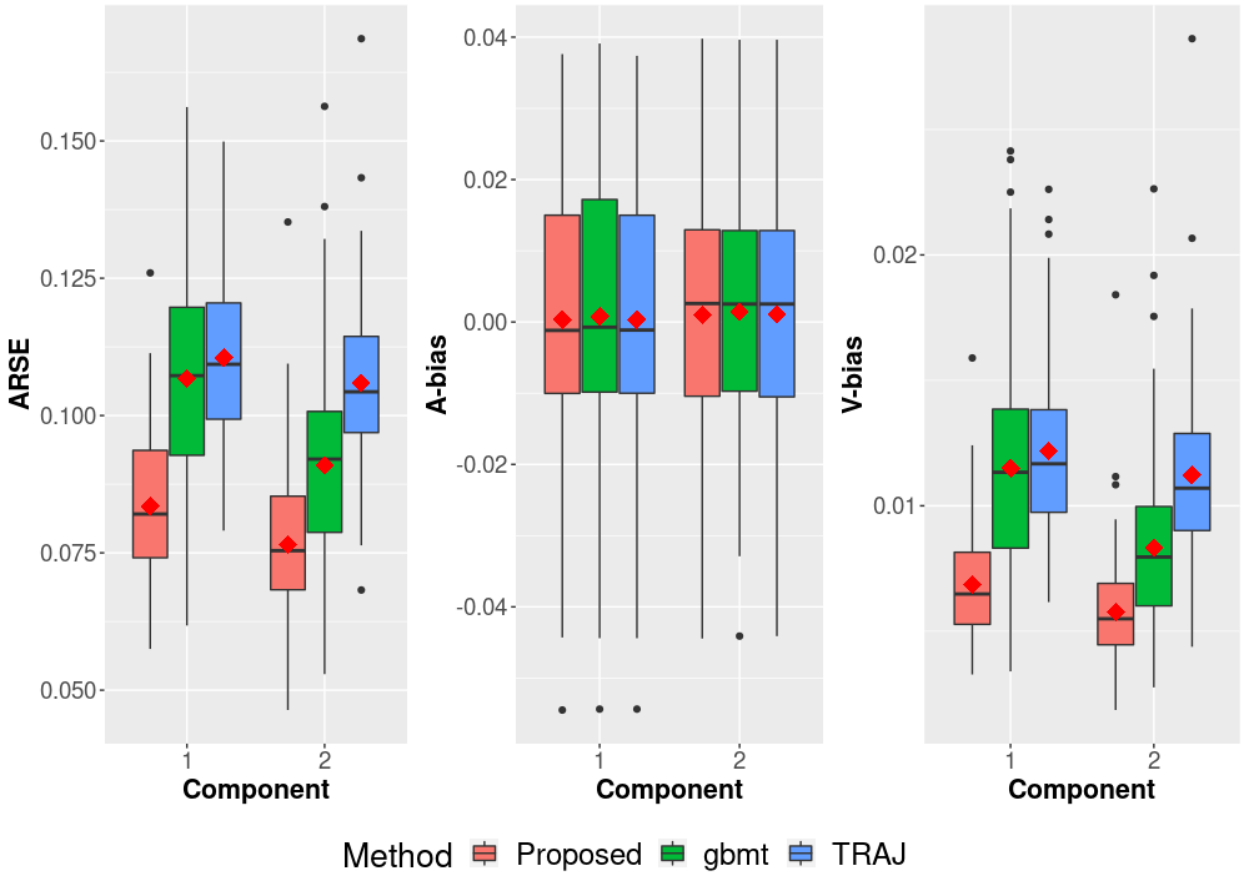

Figure 3: Boxplots of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 150 two-component trivariate time series of length 50. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. The diamond markers denote the mean statistics of each method and component.

Table 1: Root mean square errors (RMSEs) of each logistic parameter for the two-component trivariate model from 100 replicates of N two-component trivariate time series of length n. RMSEs of the proposed method were compared to `TRAJ` procedure in `SAS`. Parameters δ_0, δ_1, δ_2 and δ_3 are intercept, first, second and third logistic parameters, respectively. The true values of logistic parameters are $5, -3.5, 1, 0.1$, respectively

n	N	Method	δ_0	δ_1	δ_2	δ_3
50	150	Proposed	0.89	0.52	0.29	0.32
		TRAJ	1.57	0.87	0.36	0.34
70	150	Proposed	0.86	0.50	0.29	0.31
		TRAJ	1.55	0.86	0.36	0.34
50	250	Proposed	0.77	0.40	0.22	0.23
		TRAJ	0.96	0.50	0.23	0.24
70	250	Proposed	0.77	0.41	0.22	0.23
		TRAJ	0.97	0.51	0.24	0.24

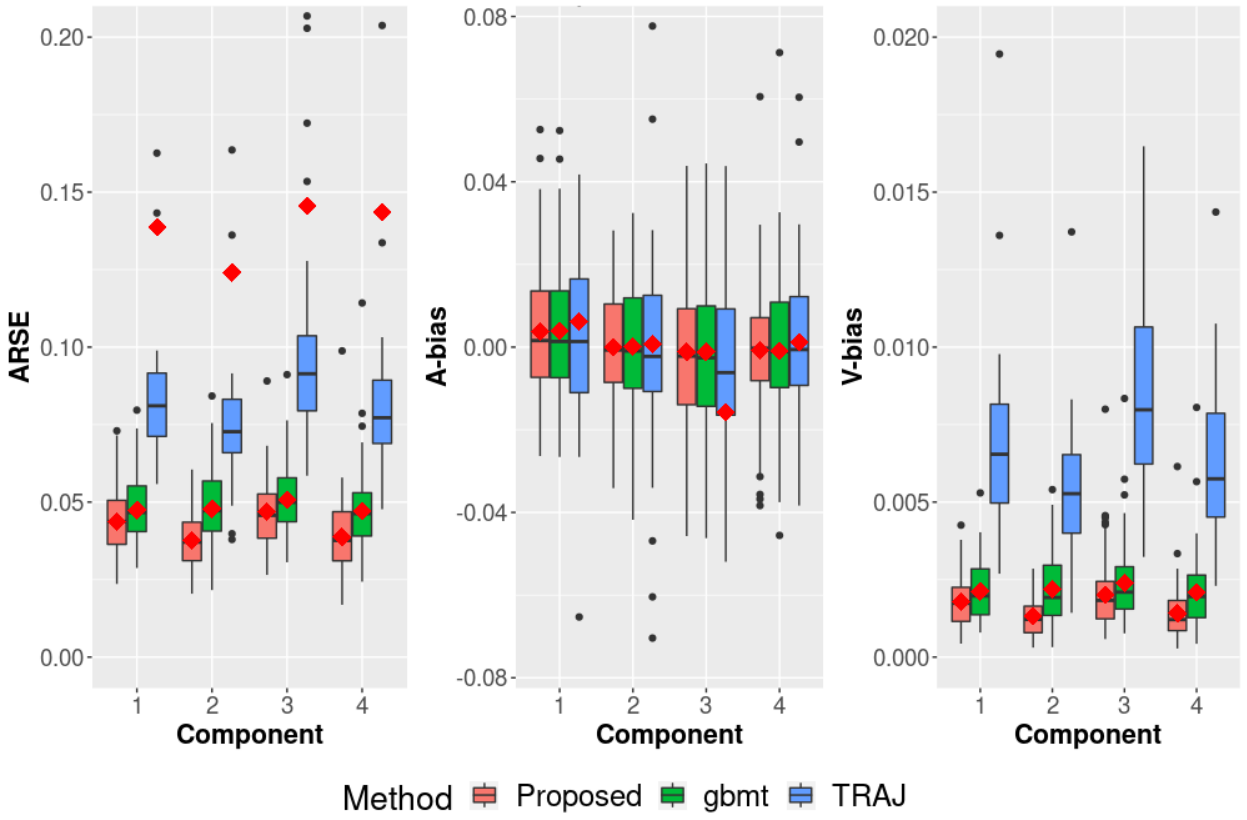

Figure 4: Boxplots of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 150 four-component bivariate time series of length 50. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. The diamond markers denote the mean statistics of each method and component. All boxplots are zoomed in for better visualization.

Table 2: Root mean square errors (RMSEs) of each logistic parameter for the four-component bivariate model from 100 replicates of 150 four-component bivariate time series of length 50. RMSEs of the proposed method were compared to `TRAJ` procedure in `SAS`. Parameters δ_0, δ_1, δ_2 and δ_3 are intercept, first, second and third logistic parameters, respectively. The fourth component was used as the reference component. The true values of logistic parameters are $5, -3.5, 1, 0.1$ (first component), $-4, 2.5, -2, -0.2$ (second component), $3, -2, 0.8, 0.2$ (third component). C1, C2, C3 and C4 denote first, second, third and fourth component, respectively.

n	N	Method	Comparison	δ_0	δ_1	δ_2	δ_3
50	150	Proposed	C1 vs C4	0.81	0.53	0.30	0.39
			C2 vs C4	1.11	0.46	0.42	0.36
			C3 vs C4	0.89	0.42	0.28	0.34
		TRAJ	C1 vs C4	1.20	0.74	0.35	0.41
			C2 vs C4	3.81	2.27	1.33	0.49
			C3 vs C4	2.07	1.33	0.76	0.32

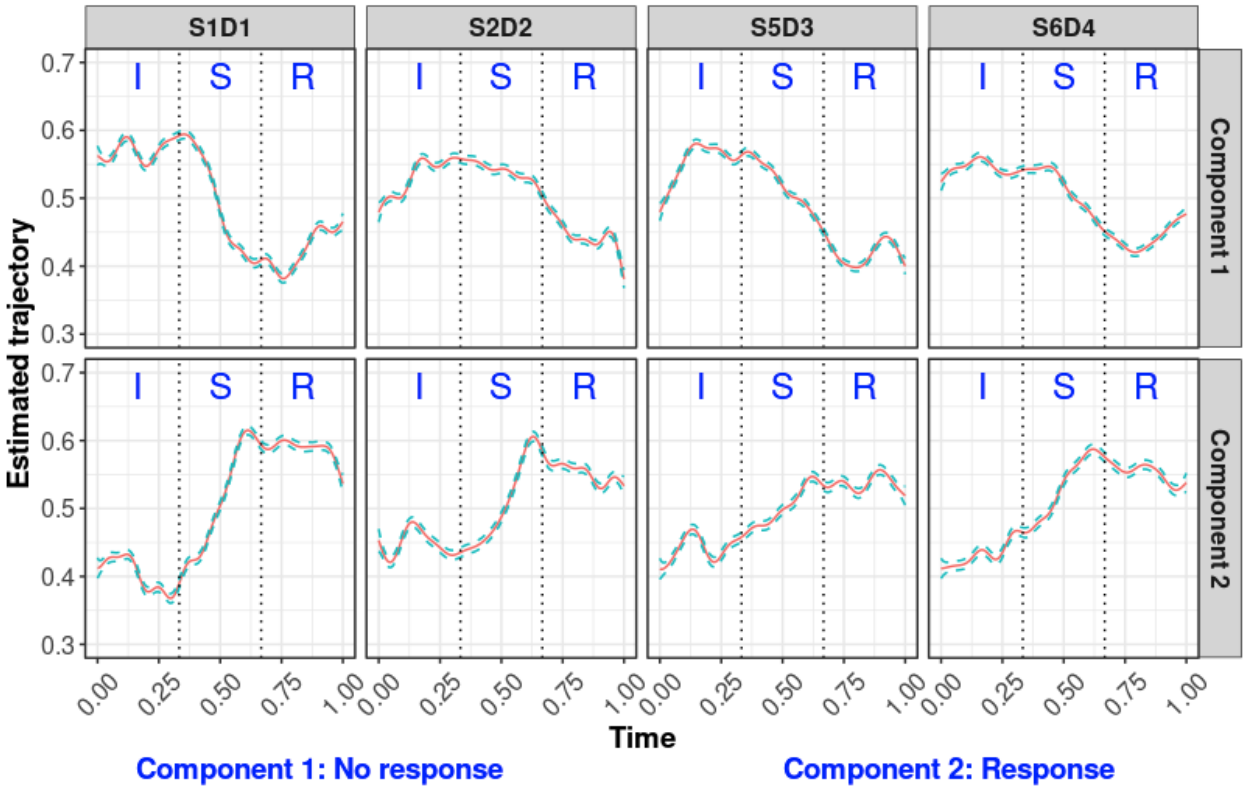

Figure 5: Estimated trajectories of the two-component model with four selected channels. **I**: Interact **S**: Still-face **R**: Recovery. Red curves are posterior mean and two green dashed curves are 95% pointwise credible intervals.

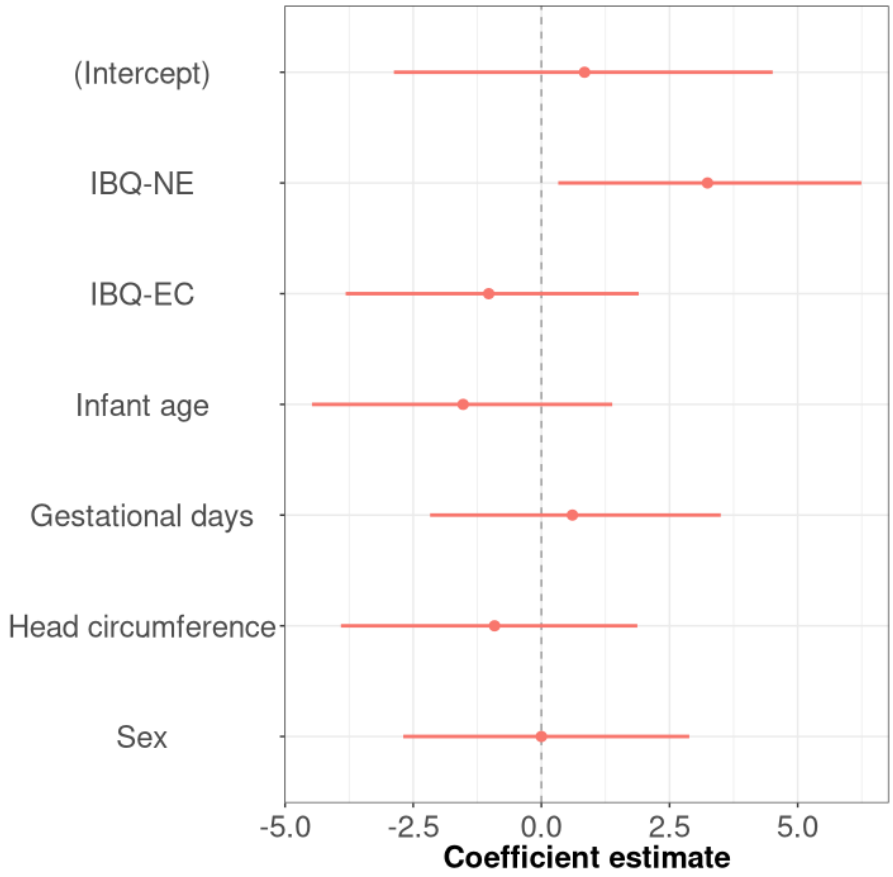

Figure 6: Logistic coefficient estimates and 95% credible intervals for each covariate of the two-component model.

10 Supplemental material

Appendix A: Details of the sampling scheme

As described in Section 5 of the paper, Gibbs sampling is used to facilitate Bayesian inference. We denote by $\boldsymbol{\Theta}_{gk} = (\boldsymbol{\theta}'_{gk}, \tau^2_{gk}, \sigma^2_{gk}, \boldsymbol{\delta}^{*\prime}_g, \kappa^2_{\zeta g})'$ the parameters for the gth component and the kth entry, and the parameters in this vector are drawn from the corresponding conditional posterior distributions. Let ℓ be the current Gibbs sampling iteration; detailed Gibbs sampling steps for drawing the parameters at the $(\ell+1)$th iteration are given below.

1. **Sampling the basis function coefficients**
For each component g and time series entry k, based on the augmented likelihood in Section 3.3 and the priors on $\boldsymbol{\theta}_{gk} = (\boldsymbol{\alpha}'_{gk}, \boldsymbol{\beta}'_{gk})'$ described in Section 4.1, the conditional posterior distribution of $(\boldsymbol{\theta}^{(\ell+1)}_{gk} \mid \boldsymbol{y}, \boldsymbol{S}, \tau^{2(\ell)}_{gk}, \sigma^{2(\ell)}_{gk})$ is:

$$\begin{aligned}
p(\boldsymbol{\theta}^{(\ell+1)}_{gk} \mid \boldsymbol{y}, \boldsymbol{S}, \tau^{2(\ell)}_{gk}, \sigma^{2(\ell)}_{gk}) &\propto p(\boldsymbol{y} \mid \boldsymbol{S}, \boldsymbol{\theta}^{(\ell+1)}_{gk}, \sigma^{2(\ell)}_{gk}) \cdot p(\boldsymbol{\theta}^{(\ell+1)}_{gk} \mid \tau^{2(\ell)}_{gk}) \\
&\propto \prod_{i=1}^{N} \Big\{ (\sigma^2_{gk})^{-n/2} \exp\big[-\frac{1}{2\sigma^2_{gk}} (\boldsymbol{y}_{ik} - \boldsymbol{S}\boldsymbol{\theta}_{gk})'(\boldsymbol{y}_{ik} - \boldsymbol{S}\boldsymbol{\theta}_{gk}) \big] \Big\}^{z_{ig}} \\
&\times |\boldsymbol{D}_{gk}|^{-1/2} \exp\Big(-\frac{1}{2} \boldsymbol{\theta}_{gk} \boldsymbol{D}^{-1}_{gk} \boldsymbol{\theta}_{gk} \Big) \\
&\propto \exp\Big\{ -\frac{1}{2\sigma^2_{gk}} \big[\sum_{i=1}^{N} z_{ig} (\boldsymbol{y}_{ik} - \boldsymbol{S}\boldsymbol{\theta}_{gk})'(\boldsymbol{y}_{ik} - \boldsymbol{S}\boldsymbol{\theta}_{gk}) + \boldsymbol{\theta}'_{gk} \sigma^2_{gk} \boldsymbol{D}^{-1}_{gk} \boldsymbol{\theta}_{gk} \big] \Big\} \\
&\propto \exp\Big[-\frac{1}{2\sigma^2_{gk}} (\boldsymbol{\theta}_{gk} - \boldsymbol{u}_{gk})' (\boldsymbol{\Lambda}_{gk})^{-1} (\boldsymbol{\theta}_{gk} - \boldsymbol{u}_{gk}) \Big] \\
&\sim N(\boldsymbol{u}_{gk}, \sigma^2_{gk} \boldsymbol{\Lambda}_{gk}),
\end{aligned}$$

where $\boldsymbol{\Lambda}_{gk} = (N^{(\ell)}_g \boldsymbol{S}'\boldsymbol{S} + \sigma^2_{gk} \boldsymbol{D}^{-1}_{gk})^{-1}$, $\boldsymbol{u}_{gk} = \boldsymbol{\Lambda}_{gk} \sum_{i=1}^{N} z_{ig} \boldsymbol{S}' \boldsymbol{y}_{ik}$, $N^{(\ell)}_g$ is the current number of subjects in the gth component, $\boldsymbol{D}_{gk} = \text{diag}(\sigma^2_\alpha \mathbf{1}_2, \tau^2_{gk} \mathbf{1}_m)$ is the prior covariance matrix for $\boldsymbol{\theta}_{gk}$. Hence, for each component g and entry k, we draw $\boldsymbol{\theta}^{(\ell+1)}_{gk}$ from $(\boldsymbol{\theta}^{(\ell+1)}_{gk} \mid \boldsymbol{y}, \boldsymbol{S}, \tau^{2(\ell)}_{gk}, \sigma^{2(\ell)}_{gk}) \sim N(\boldsymbol{u}_{gk}, \sigma^2_{gk} \boldsymbol{\Lambda}_{gk})$.

2. **Sampling the error variances**
Gelman (2006) proposed using the half-t distribution as the prior on scale parameters. We follow Wand *and others* (2011) and express the half-t prior of Section 4.3 as a scale mixture of inverse Gamma distributions as follows

$$(\sigma^2_{gk} \mid a_{\sigma_{gk}}) \sim IG\Big(\frac{\nu_\sigma}{2}, \frac{\nu_\sigma}{a_{\sigma_{gk}}}\Big), a_{\sigma_{gk}} \sim IG\Big(\frac{1}{2}, \frac{1}{A^2_\sigma}\Big).$$

Therefore, the conditional posterior distribution of the latent variable $a_{\sigma_{gk}}$ is

$$p(a^{(\ell+1)}_{\sigma_{gk}} \mid \sigma^{2(\ell)}_{gk}) \propto \exp\Big[-\frac{1}{a_{\sigma_{gk}}} \Big(\frac{\nu_\sigma}{\sigma^2_{gk}} + \frac{1}{A^2_\sigma} \Big) \Big] \times (a_{\sigma_{gk}})^{-(\frac{1}{2}+1+\frac{\nu_\sigma}{2})},$$

which is $IG\Big(\frac{\nu_\sigma+1}{2}, \frac{\nu_\sigma}{\sigma^2_{gk}} + \frac{1}{A^2_\sigma}\Big)$. Denoting by $\boldsymbol{\epsilon}_{igk}$ the error vector of time series $\boldsymbol{y}_{ik}$ for component g, we have $\boldsymbol{\epsilon}_{igk} = \boldsymbol{y}_{ik} - \boldsymbol{S}\boldsymbol{\theta}_{gk}$, where $\boldsymbol{\epsilon}_{igk} \sim N(\mathbf{0}, \sigma^2_{gk} \boldsymbol{I}_n)$. The conditional distribution of the error variance is

$$\begin{aligned}
p(\sigma^{2(\ell+1)}_{gk} \mid \boldsymbol{\epsilon}^{(\ell+1)}_{igk}, a^{(\ell+1)}_{\sigma_{gk}}) &\propto p(\boldsymbol{\epsilon}^{(\ell+1)}_{igk} \mid \sigma^{2(\ell+1)}_{gk}) \cdot p(a^{(\ell+1)}_{\sigma_{gk}} \mid \sigma^{2(\ell+1)}_{gk}) \cdot p(\sigma^{2(\ell+1)}_{gk}) \\
&\propto \prod_{i=1}^{N} \Big[(\sigma^2_{gk})^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2_{gk}} \boldsymbol{\epsilon}'_{igk} \boldsymbol{\epsilon}_{igk} \Big) \Big]^{z_{ig}} \times (\sigma^2_{gk})^{-(\frac{\nu_\sigma}{2}+1)} \exp\Big(-\frac{\nu_\sigma}{\sigma^2_{gk} a_{\sigma_{gk}}} \Big) \\
&\propto (\sigma^2_{gk})^{-(\frac{n}{2} N^{(\ell)}_g + \frac{\nu_\sigma}{2} + 1)} \exp\Big[-\frac{1}{\sigma^2_{gk}} \Big(\frac{\sum_{i=1}^{N} z_{ig} \boldsymbol{\epsilon}'_{igk} \boldsymbol{\epsilon}_{igk}}{2} + \frac{\nu_\sigma}{a_{\sigma_{gk}}} \Big) \Big],
\end{aligned}$$

which is $IG\Big(\frac{nN^{(\ell)}_g + \nu_\sigma}{2}, \frac{\sum_{i=1}^{N} z_{ig} \boldsymbol{\epsilon}'_{igk} \boldsymbol{\epsilon}_{igk}}{2} + \frac{\nu_\sigma}{a_{\sigma_{gk}}}\Big)$. The sampling scheme proceeds by first sampling $(a^{(\ell+1)}_{\sigma_{gk}} \mid \sigma^{2(\ell)}_{gk})$ and then $(\sigma^{2(\ell+1)}_{gk} \mid \boldsymbol{\epsilon}^{(\ell+1)}_{igk}, a^{(\ell+1)}_{\sigma_{gk}})$.

3. **Sampling the smoothing parameters**

The smoothing parameters τ^2_{gk} are drawn by analogy to the error variances. We first draw $(a^{(\ell+1)}_{\tau_{gk}} \mid \tau^{2(\ell)}_{gk}) \sim IG\Big(\frac{\nu_\tau+1}{2}, \frac{\nu_\tau}{\tau^2_{gk}} + \frac{1}{A^2_\tau}\Big)$. The conditional posterior distribution of the smoothing parameters is

$$p(\tau^{2(\ell+1)}_{gk} \mid \boldsymbol{\beta}^{(\ell+1)}_{gk}, a^{(\ell+1)}_{\tau_{gk}}) \propto p(\boldsymbol{\beta}^{(\ell+1)}_{gk} \mid \tau^{2(\ell+1)}_{gk}) \cdot p(a^{(\ell+1)}_{\tau_{gk}} \mid \tau^{2(\ell+1)}_{gk}) \cdot p(\tau^{2(\ell+1)}_{gk})$$
$$\propto (\tau^2_{gk})^{-\frac{m+\nu_\tau}{2}} \exp\Big[-\frac{1}{\tau^2_{gk}}\Big(\frac{\nu_\tau}{a_{\tau_{gk}}} + \frac{\boldsymbol{\beta}'_{gk}\boldsymbol{\beta}_{gk}}{2}\Big)\Big],$$

which is $IG\Big(\frac{\nu_\tau+m}{2}, \frac{\boldsymbol{\beta}'_{gk}\boldsymbol{\beta}_{gk}}{2} + \frac{\nu_\tau}{a_{\tau_{gk}}}\Big)$. The sampling scheme proceeds by first sampling $(a^{(\ell+1)}_{\tau_{gk}} \mid \tau^{2(\ell)}_{gk})$ and then $(\tau^{2(\ell+1)}_{gk} \mid \boldsymbol{\beta}^{(\ell+1)}_{gk}, a^{(\ell+1)}_{\tau_{gk}})$.

4. **Sampling the logistic parameters**

Let $\boldsymbol{\delta}^*_g = (\boldsymbol{\delta}^T_g, \boldsymbol{\zeta}^T_g)^T$ be the aggregation of the logistic parameters and all random intercepts for the gth component. Based on the logits of Section 3.2 and the corresponding priors described in Section 4.4, the conditional posterior distribution of $(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \kappa^{2(\ell)}_{\zeta g})$ is

$$p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \kappa^{2(\ell)}_{\zeta g}) \propto p(z^{(\ell)}_{ig} = 1 \mid \boldsymbol{V}^*, \boldsymbol{\delta}^{*(\ell+1)}_g) \cdot p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \kappa^{2(\ell)}_{\zeta g})$$
$$= \prod_{i=1}^{N} \Big[\frac{\exp(\boldsymbol{V}^{*\prime}_i \boldsymbol{\delta}^*_g)}{\sum_{h=1}^{G} \exp(\boldsymbol{V}^{*\prime}_i \boldsymbol{\delta}^*_h)}\Big]^{z_{ig}} p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \kappa^{2(\ell)}_{\zeta g}),$$

where $\boldsymbol{V}^* = (\boldsymbol{V}^*_1, \ldots, \boldsymbol{V}^*_N)'$ is a $N \times (P+1)$ matrix with $\boldsymbol{V}^*_i$ representing all covariates (including intercepts) for subject i. To sample from the posterior distribution of $p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \kappa^{2(\ell)}_{\zeta g})$, we adopt the Póya-Gamma data augmentation strategy of Polson *and others* (2013) by introducing a latent variable ω_{ig} coming from the Pólya-Gamma distribution. Thus, the conditional posterior distributions of the logistic parameters are

$$p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \omega^{(\ell+1)}_{ig}, \kappa^{2(\ell)}_{\zeta g}) \propto p(z^{(\ell)}_{ig} = 1 \mid \boldsymbol{V}^*, \omega^{(\ell+1)}_{ig}, \boldsymbol{\delta}^{*(\ell+1)}_g) \cdot p(\omega^{(\ell+1)}_{ig} \mid \boldsymbol{V}^*, \boldsymbol{\delta}^{*(\ell)}_g)$$
$$\cdot\, p(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \kappa^{2(\ell)}_{\zeta g})$$
$$\propto \exp\Big(-\frac{\omega_{ig}\eta^2_{ig}}{2}\Big) \cdot p(\omega_{ig} \mid 1, 0) |\boldsymbol{B}_g|^{-P/2} \exp\Big(-\frac{1}{2}\boldsymbol{\delta}^{*\prime}_g \boldsymbol{B}^{-1}_g \boldsymbol{\delta}^*_g\Big),$$

where $\eta_{ig} = \boldsymbol{V}^{*\prime}_i \boldsymbol{\delta}^*_g - C_{ig}$ and $C_{ig} = \log \sum_{h \neq j} \exp(\boldsymbol{V}^{*\prime}_i \boldsymbol{\delta}^*_h)$, $p(\omega_{ig} \mid 1, 0)$ is the Pólya-gamma distribution $PG(b, c)$ with $b = 1$ and $c = 0$, $\boldsymbol{B}_g$ is the prior covariance matrix of Section 4.4 and $\boldsymbol{B}_g = \text{diag}(\sigma^2_{\delta g}\mathbf{1}_{P+1}, \kappa^2_{\zeta g}\mathbf{1}_N)$. By assuming the conjugate prior $N(\mathbf{0}, \boldsymbol{B}_g)$ on $\boldsymbol{\delta}^*_g$, the posterior distribution of the Pólya-gamma latent variable is

$$(\omega^{(\ell+1)}_{ig} \mid \boldsymbol{V}^*, \boldsymbol{\delta}^{*(\ell)}_g) \sim PG(1, \eta_{ig}).$$

Thus, the conditional distributions of the logistic parameters (including the random intercepts) are

$$(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \omega^{(\ell+1)}_{ig}, \kappa^{2(\ell)}_{\zeta g}) \sim N(\boldsymbol{M}_g, \boldsymbol{\Sigma}_g),$$

where $\boldsymbol{\Sigma}_g = (\boldsymbol{V}^{*\prime}\boldsymbol{\Omega}_g\boldsymbol{V}^* + \boldsymbol{B}^{-1}_g)^{-1}$, $\boldsymbol{M}_g = \boldsymbol{\Sigma}_g\big[\boldsymbol{V}^{*\prime}(\boldsymbol{\Omega}_g\boldsymbol{C}_g + \boldsymbol{\xi}_g)\big]$, $\boldsymbol{\Omega}_g = \text{diag}(\omega_{1g}, \cdots, \omega_{Ng})$, $\boldsymbol{C}_g = (C_{1g}, \cdots, C_{Ng})'$, and $\boldsymbol{\xi}_g = (\xi_{1g}, \cdots, \xi_{Ng})'$, with $\xi_{ig} = z_{ig} - \frac{1}{2}$. Thus, $\boldsymbol{\delta}^{*(\ell+1)}_g$ is drawn by first sampling $(\omega^{(\ell+1)}_{ig} \mid \boldsymbol{V}^*, \boldsymbol{\delta}^{*(\ell)}_g)$ and then $(\boldsymbol{\delta}^{*(\ell+1)}_g \mid \boldsymbol{V}^*, z^{(\ell)}_{ig}, \omega^{(\ell+1)}_{ig}, \kappa^{2(\ell)}_{\zeta g})$.

5. **Sampling the variances of the random intercepts**

By analogy with sampling the error variances and sthe moothing parameters, we first draw $(a^{(\ell+1)}_{\kappa_g} \mid \kappa^{2(\ell)}_{\zeta g}) \sim IG\Big(\frac{\nu_\kappa+1}{2}, \frac{\nu_\kappa}{\kappa^2_{\zeta g}} + \frac{1}{A^2_\kappa}\Big)$. The conditional posterior distributions of the variances of the random intercepts are

$$p(\kappa^{2(\ell+1)}_{\zeta g} \mid \boldsymbol{\zeta}^{(\ell+1)}_g, a^{(\ell+1)}_{\kappa_g}) \propto p(\boldsymbol{\zeta}^{(\ell+1)}_g \mid \kappa^{2(\ell+1)}_{\zeta g}) \cdot p(a^{(\ell+1)}_{\kappa_g} \mid \kappa^{2(\ell+1)}_{\zeta g}) \cdot p(\kappa^{2(\ell+1)}_{\zeta g})$$
$$\propto (\kappa^2_{\zeta g})^{-\frac{N+\nu_\kappa}{2}+1} \exp\Big[-\frac{1}{\kappa^2_{\zeta g}}\Big(\frac{\nu_\kappa}{a_{\kappa_g}} + \frac{\boldsymbol{\zeta}^T_g\boldsymbol{\zeta}_g}{2}\Big)\Big],$$

which is $IG\Big(\frac{\nu_\kappa+N}{2}, \frac{\boldsymbol{\zeta}^T_g\boldsymbol{\zeta}_g}{2} + \frac{\nu_\kappa}{a_{\kappa_g}}\Big)$. The sampling scheme proceeds by first sampling $(a^{(\ell+1)}_{\kappa_g} \mid \kappa^{2(\ell)}_{\zeta g})$ and then $(\kappa^{2(\ell+1)}_{\zeta g} \mid \boldsymbol{\zeta}^{(\ell+1)}_g, a^{(\ell+1)}_{\kappa_g})$.

6. **Computing the mixing weights**

   After drawing the $\boldsymbol{\delta}_g^*$, the mixing weights $\pi_{ig}^{(\ell+1)}$ for each component, given the design matrix V_i^*, are computed by

$$p(\pi_{ig}^{(\ell+1)} \mid V_i^*, \boldsymbol{\delta}_g^{*(\ell+1)}) = \frac{\exp(\boldsymbol{V}_i^{*T}\boldsymbol{\delta}_g^*)}{\sum_{h=1}^{G}\exp(\boldsymbol{V}_i^{*T}\boldsymbol{\delta}_h^*)}.$$

7. **Sampling the latent indicators**

   After sampling all parameters and computing the mixing weights, the final Gibbs step is to allocate subjects to different components by drawing the latent indicators z_{ig}. As in Section 3.3, the conditional posterior of these indicators is

$$p(z_{ig}^{(\ell+1)} = 1 \mid \boldsymbol{y}, \boldsymbol{S}, \boldsymbol{\Theta}^{(\ell+1)}, \pi_{ig}^{(\ell+1)}) = \frac{\pi_{ig}\prod_{k=1}^{K} f_{gk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{gk})}{\sum_{h=1}^{G}\pi_{ih}\prod_{k=1}^{K} f_{hk}(\boldsymbol{y}_{ik} \mid \boldsymbol{\Theta}_{hk})},$$

   and the indicators are drawn from the multinomial distribution.

Appendix B: Additional simulation results

Appendix B adds more simulation results in addition to simulation results in the paper itself. To further demonstrate the performance of the proposed method, we conduct simulation studies under two scenarios: two-component mixture of trivariate time series and four-component mixture of bivariate time series. The model formula is displayed in Section 6.1 of the paper. We investigate the performance of our proposed method in terms of estimated trajectories and logistic parameters.

Mean(SD) of the ARSE, A-bias and V-bias for each component of the two-component trivariate model are given in Table 3. To demonstrate the performance of the proposed method in various settings, we look at combinations of the number of multivariate time series ($N = 150, 250$) and the length of each time series ($n = 50, 70$), and compare our proposed method to two existing methods: `gbmt` package in R (Magrini, 2022) and `TRAJ` procedure in `SAS` (Nagin *and others*, 2018). The case of $n = 50$ and $N = 150$ in Table 3 corresponds to Figure 3 in the main paper. The performance of the logistic parameters (RMSEs) with different values of n and N are given in Table 1 of the paper.

The Mean(SD) of the ARSE, A-bias and V-bias for each component of the $N = 150$ four-component mixture of bivariate time series of length $n = 50$ are given in Table 4, which corresponds to Figure 4 in the main paper. RMSEs of the logistic parameters for this setting are listed in Table 2 of the paper. Tables 5 - 10 present performance measures of the estimated trajectories and logistic parameters for combinations of different lengths of time series n and numbers of time series N, under the scenario of the four-component bivariate model.

As expected, our proposed method outperforms the two existing methods in terms of the estimated trajectories for each component under different settings (different values of n and N, for both the two-component trivariate and the four-component bivariate scenarios). The proposed method is able to achieve smaller ARSE and V-bias, while all three methods are able to obtain estimated trajectories with a very small bias. Notably, for the four-component bivariate scenario, `TRAJ` gives larger values of mean ARSE, A-bias and V-bias, which result from imprecise estimates of several replicates due to convergence issues. In terms of the logistic parameters, our proposed method outperforms `TRAJ` in almost all comparisons, especially for the intercept δ_0 and the slope of the first covariate δ_1. Our proposed method yields shrinkage estimates for the logistic parameters due to using a Bayesian method, while the multinomial logistic regression used in `TRAJ` gives inflated parameter estimates in case of perfect separations and unbalanced designs.

Appendix C: Additional real-data results

Appendix C describes more real-data results in addition to those in Section 7 of the main paper. Our motivating study is described in Section 2 of the paper. Figure 7 shows the estimated trajectories of the three-component model for another set of four channels (S1D3, S3D2, S5D4, and S7D4). Based on the selection criterion DIC introduced in Section 5.2 of the main paper, the three-component model was selected as the best model. We named the second component as the mixture response component because it involves both increased and decreased brain activity or hemoglobin level for the still-face period for different channels. In addition, Figure 8 displays the logistic coefficient estimates and 95% credible intervals corresponding to each covariate. The last component (third component) is always used as the reference. We reach the same conclusion with positive estimates of IBQ-NE scores and negative estimates of IBQ-EC scores for both components (component 1 vs. 3, component 2 vs. 3).

In addition to the four-channel analyses, we also present results from all channels (twelve channels). Figures 9, 10, 11 present the estimated trajectories of the first, second and third component for the three-component model with all twelve

Table 3: Mean (standard deviation) of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of N two-component trivariate time series of length n. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. C1 and C2 denote first and second components. Means were calculated by averaging over estimates of 100 replicates. Standard deviations are Monte Carlo standard deviations from estimates of 100 replicates. Each value was reported $\times 10^2$.

n	N	Method	ARSE C1	A-bias C1	V-bias C1	ARSE C2	A-bias C2	V-bias C2
50	150	Proposed	8.35 (1.26)	0.03 (1.83)	0.68 (0.22)	7.65 (1.38)	0.10 (1.76)	0.58 (0.22)
		gbmt	10.67 (1.91)	0.03 (1.83)	1.15 (0.42)	9.08 (1.72)	0.10 (1.76)	0.83 (0.33)
		TRAJ	11.06 (1.48)	0.03 (1.83)	1.22 (0.33)	10.59 (1.52)	0.10 (1.76)	1.12 (0.34)
70	150	Proposed	7.16 (1.04)	0.24 (1.38)	0.51 (0.16)	6.53 (1.07)	-0.11 (1.42)	0.42 (0.14)
		gbmt	9.91 (1.96)	0.24 (1.38)	1.01 (0.40)	8.19 (1.65)	-0.11 (1.42)	0.68 (0.29)
		TRAJ	9.34 (1.10)	0.24 (1.38)	0.87 (0.22)	8.95 (1.13)	-0.11 (1.42)	0.80 (0.20)
50	250	Proposed	6.81 (1.02)	0.07 (1.33)	0.46 (0.14)	6.22 (0.94)	0.02 (1.31)	0.38 (0.12)
		gbmt	9.79 (1.91)	0.07 (1.33)	0.98 (0.40)	8.00 (1.53)	0.02 (1.31)	0.65 (0.26)
		TRAJ	8.70 (1.18)	0.07 (1.33)	0.76 (0.21)	8.20 (1.03)	0.02 (1.31)	0.67 (0.17)
70	250	Proposed	5.65 (0.84)	0.08 (1.00)	0.32 (0.10)	5.27 (0.82)	-0.06 (1.42)	0.27 (0.09)
		gbmt	9.15 (1.96)	0.08 (1.00)	0.87 (0.38)	7.43 (1.60)	-0.06 (1.42)	0.56 (0.26)
		TRAJ	7.18 (0.94)	0.08 (1.00)	0.52 (0.14)	6.80 (0.77)	-0.06 (1.42)	0.45 (0.10)

Table 4: Mean (standard deviation) of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 150 four-component bivariate time series of length 50. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. C1, C2, C3 and C4 denote first, second, third and fourth component, respectively. Means were calculated by averaging over estimates of 100 replicates. Standard deviations are Monte Carlo standard deviations from estimates of 100 replicates. Each value was reported $\times 10^2$.

n	N	Method	ARSE C1	A-bias C1	V-bias C1	ARSE C2	A-bias C2	V-bias C2
50	150	Proposed	4.38 (1.04)	0.38 (1.59)	0.18 (0.08)	3.76 (0.87)	-0.01 (1.37)	0.13 (0.06)
		gbmt	4.75 (1.04)	0.38 (1.59)	0.21 (0.09)	4.79 (1.17)	0.01 (1.65)	0.22 (0.11)
		TRAJ	13.87 (15.59)	0.62 (9.91)	3.39 (9.24)	12.41 (13.42)	0.08 (9.74)	2.41 (5.37)
n	N	Method	ARSE C3	A-bias C3	V-bias C3	ARSE C4	A-bias C4	V-bias C4
50	150	Proposed	4.69 (1.14)	-0.11 (1.83)	0.20 (0.12)	3.88 (1.15)	-0.09 (1.56)	0.14 (0.08)
		gbmt	5.08 (1.12)	-0.12 (1.83)	0.24 (0.12)	4.70 (1.32)	-0.09 (1.78)	0.21 (0.12)
		TRAJ	14.55 (14.82)	-1.58 (10.31)	3.24 (7.35)	14.36 (17.01)	0.12 (9.92)	3.99 (10.70)

Table 5: Mean (standard deviation) of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 150 four-component bivariate time series of length 70. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. C1, C2, C3 and C4 denote first, second, third and fourth component, respectively. Means were calculated by averaging over estimates of 100 replicates. Standard deviations are Monte Carlo standard deviations from estimates of 100 replicates. Each value was reported $\times 10^2$.

n	N	Method	ARSE C1	A-bias C1	V-bias C1	ARSE C2	A-bias C2	V-bias C2
70	150	Proposed	3.82 (0.95)	0.44 (1.30)	0.14 (0.07)	3.22 (0.85)	-0.07 (0.95)	0.10 (0.06)
		gbmt	4.05 (0.97)	0.44 (1.30)	0.16 (0.08)	4.11 (1.12)	-0.08 (1.15)	0.17 (0.10)
		TRAJ	13.51 (17.04)	-0.30 (9.34)	3.86 (10.73)	10.00 (10.11)	-0.25 (6.21)	1.64 (4.02)
n	N	Method	ARSE C3	A-bias C3	V-bias C3	ARSE C4	A-bias C4	V-bias C4
70	150	Proposed	4.12 (0.90)	-0.29 (1.69)	0.15 (0.06)	3.52 (0.85)	0.24 (1.22)	0.12 (0.06)
		gbmt	4.38 (1.01)	-0.29 (1.69)	0.17 (0.07)	4.13 (0.99)	0.27 (1.40)	0.16 (0.09)
		TRAJ	13.03 (17.04)	0.30 (10.49)	3.86 (10.73)	11.77 (13.78)	0.39 (8.49)	2.57 (6.45)

Table 6: Root mean square errors (RMSEs) of each logistic parameter for the four-component bivariate model from 100 replicates of 150 four-component bivariate time series of length 70. RMSEs of the proposed method were compared to `TRAJ` procedure in `SAS`. Parameters δ_0, δ_1, δ_2 and δ_3 are intercept, first, second and third logistic parameters, respectively. The fourth component was used as the reference component. The true values of logistic parameters are $5, -3.5, 1, 0.1$ (first component), $-4, 2.5, -2, -0.2$ (second component), $3, -2, 0.8, 0.2$ (third component). C1, C2, C3 and C4 denote first, second, third and fourth component, respectively.

n	N	Method	Comparison	δ_0	δ_1	δ_2	δ_3
70	150	Proposed	C1 vs C4	0.81	0.51	0.29	0.41
			C2 vs C4	1.42	0.73	0.58	0.36
			C3 vs C4	1.05	0.58	0.37	0.31
		TRAJ	C1 vs C4	1.13	0.66	0.31	0.45
			C2 vs C4	3.12	1.66	0.99	0.55
			C3 vs C4	1.15	0.74	0.48	0.35

Table 7: Mean (standard deviation) of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 250 four-component bivariate time series of length 50. The proposed method was compared to `R` package `gbmt` and `TRAJ` procedure in `SAS`. C1, C2, C3 and C4 denote first, second, third and fourth component, respectively. Means were calculated by averaging over estimates of 100 replicates. Standard deviations are Monte Carlo standard deviations from estimates of 100 replicates. Each value was reported $\times 10^2$.

n	N	Method	ARSE C1	A-bias C1	V-bias C1	ARSE C2	A-bias C2	V-bias C2
50	250	Proposed	3.42 (0.78)	0.18 (1.19)	0.11 (0.05)	2.86 (0.61)	-0.14 (0.98)	0.08 (0.04)
		gbmt	3.57 (0.85)	0.18 (1.19)	0.12 (0.06)	3.68 (0.79)	-0.15 (1.20)	0.13 (0.06)
		TRAJ	11.66 (15.03)	0.53 (10.63)	2.50 (6.30)	8.90 (9.38)	1.47 (7.81)	1.05 (2.16)
n	N	Method	ARSE C3	A-bias C3	V-bias C3	ARSE C4	A-bias C4	V-bias C4
50	250	Proposed	3.93 (0.92)	-0.06 (1.48)	0.14 (0.07)	3.28 (0.76)	-0.10 (1.19)	0.10 (0.05)
		gbmt	4.16 (0.95)	-0.06 (1.49)	0.16 (0.07)	3.83 (0.83)	-0.13 (1.36)	0.14 (0.06)
		TRAJ	10.80 (9.92)	0.49 (5.78)	1.83 (3.70)	10.17 (12.54)	-0.17 (8.83)	1.84 (5.20)

Table 8: Root mean square errors (RMSEs) of each logistic parameter for the four-component bivariate model from 100 replicates of 250 four-component bivariate time series of length 50. RMSEs of the proposed method were compared to TRAJ procedure in SAS. Parameters δ_0, δ_1, δ_2 and δ_3 are intercept, first, second and third logistic parameters, respectively. The fourth component was used as the reference component. The true values of logistic parameters are $5, -3.5, 1, 0.1$ (first component), $-4, 2.5, -2, -0.2$ (second component), $3, -2, 0.8, 0.2$ (third component). C1, C2, C3 and C4 denote first, second, third and fourth component, respectively.

n	N	Method	Comparison	δ_0	δ_1	δ_2	δ_3
50	250	Proposed	C1 vs C4	0.63	0.41	0.26	0.29
			C2 vs C4	1.00	0.46	0.40	0.27
			C3 vs C4	0.63	0.33	0.23	0.24
		TRAJ	C1 vs C4	0.91	0.56	0.30	0.28
			C2 vs C4	1.40	0.86	0.61	0.35
			C3 vs C4	2.24	1.40	0.85	0.27

Table 9: Mean (standard deviation) of the averaged root square error (ARSE), the averaged bias (A-bias) and the variance of bias (V-bias) of estimated trajectories for each component from 100 replicates of 250 four-component bivariate time series of length 70. The proposed method was compared to R package gbmt and TRAJ procedure in SAS. C1, C2, C3 and C4 denote first, second, third and fourth component, respectively. Means were calculated by averaging over estimates of 100 replicates. Standard deviations are Monte Carlo standard deviations from estimates of 100 replicates. Each value was reported $\times 10^2$.

n	N	Method	ARSE C1	A-bias C1	V-bias C1	ARSE C2	A-bias C2	V-bias C2
70	250	Proposed	2.94	-0.04	0.08	2.61	-0.01	0.06
			(0.60)	(1.06)	(0.04)	(0.57)	(0.87)	(0.03)
		gbmt	3.10	-0.04	0.09	3.18	0.01	0.10
			(0.63)	(1.06)	(0.04)	(0.70)	(1.05)	(0.05)
		TRAJ	13.52	-1.58	3.71	11.51	-0.19	2.54
			(17.70)	(11.09)	(8.80)	(14.85)	(9.98)	(7.10)
n	N	Method	ARSE C3	A-bias C3	V-bias C3	ARSE C4	A-bias C4	V-bias C4
70	250	Proposed	3.30	-0.02	0.10	2.85	-0.07	0.08
			(0.76)	(1.21)	(0.05)	(0.73)	(0.97)	(0.04)
		gbmt	3.51	-0.01	0.12	3.26	-0.09	0.10
			(0.79)	(1.21)	(0.06)	(0.80)	(1.06)	(0.05)
		TRAJ	13.07	1.48	2.90	10.68	0.65	2.16
			(15.21)	(10.52)	(7.11)	(12.93)	(8.11)	(5.66)

Table 10: Root mean square errors (RMSEs) of each logistic parameter for the four-component bivariate model from 100 replicates of 250 four-component bivariate time series of length 70. RMSEs of the proposed method were compared to TRAJ procedure in SAS. Parameters δ_0, δ_1, δ_2 and δ_3 are intercept, first, second and third logistic parameters, respectively. The fourth component was used as the reference component. The true values of logistic parameters are $5, -3.5, 1, 0.1$ (first component), $-4, 2.5, -2, -0.2$ (second component), $3, -2, 0.8, 0.2$ (third component). C1, C2, C3 and C4 denote first, second, third and fourth component, respectively.

n	N	Method	Comparison	δ_0	δ_1	δ_2	δ_3
70	250	Proposed	C1 vs C4	0.64	0.40	0.26	0.28
			C2 vs C4	0.92	0.42	0.41	0.28
			C3 vs C4	0.63	0.31	0.23	0.23
		TRAJ	C1 vs C4	0.82	0.50	0.27	0.28
			C2 vs C4	1.47	0.86	0.61	0.36
			C3 vs C4	1.60	0.96	0.57	0.25

channels, respectively. The three-component model was selected as the best model for the twelve-channel analysis based on the adjusted DIC. We named the three components no response, mixture response, and response component, respectively. Figure 12 displays the logistic coefficient estimates and 95 % credible intervals corresponding to each covariate.

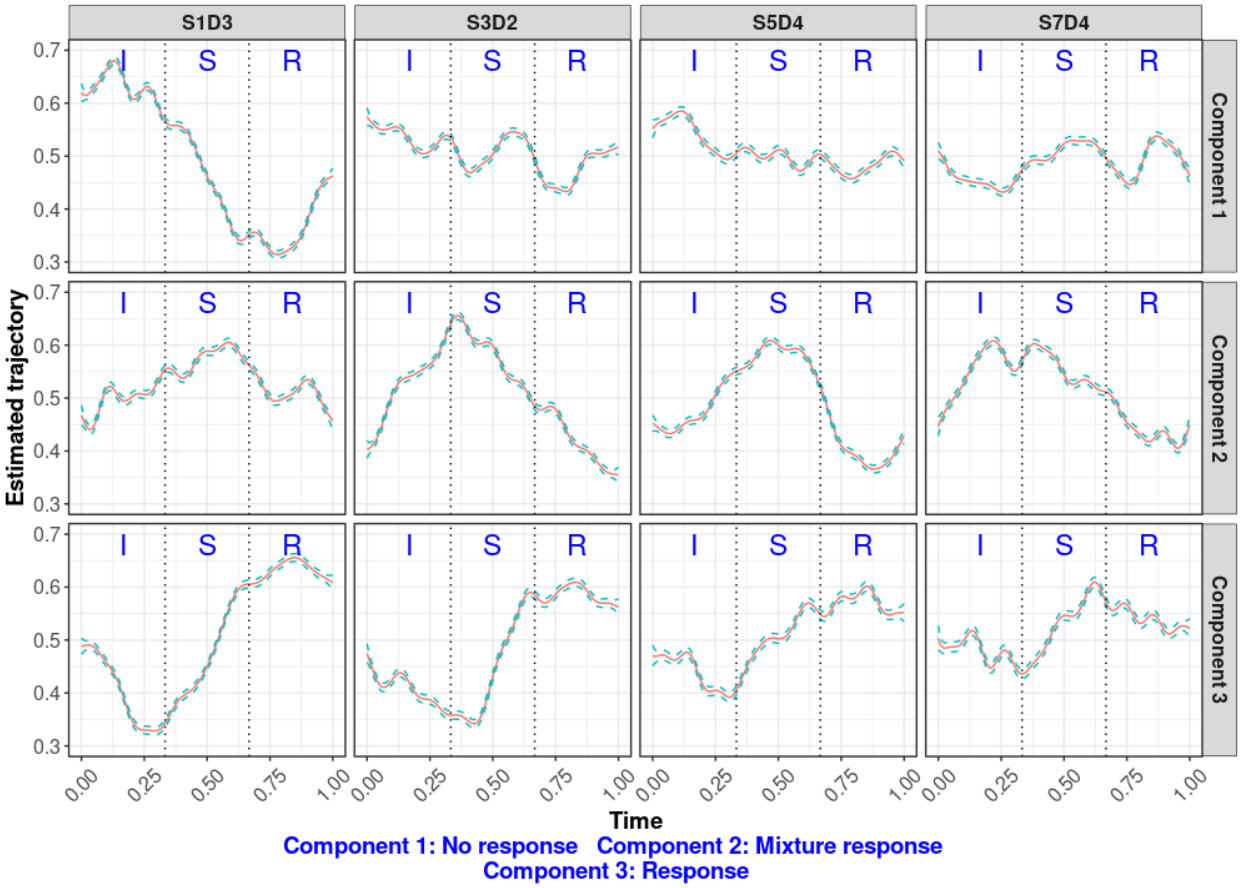

Figure 7: Estimated trajectories of the three-component model with four selected channels. **I**: Interact **S**: Still-face **R**: Recovery. Red curves are posterior means and the two green dashed curves are 95% pointwise credible intervals.

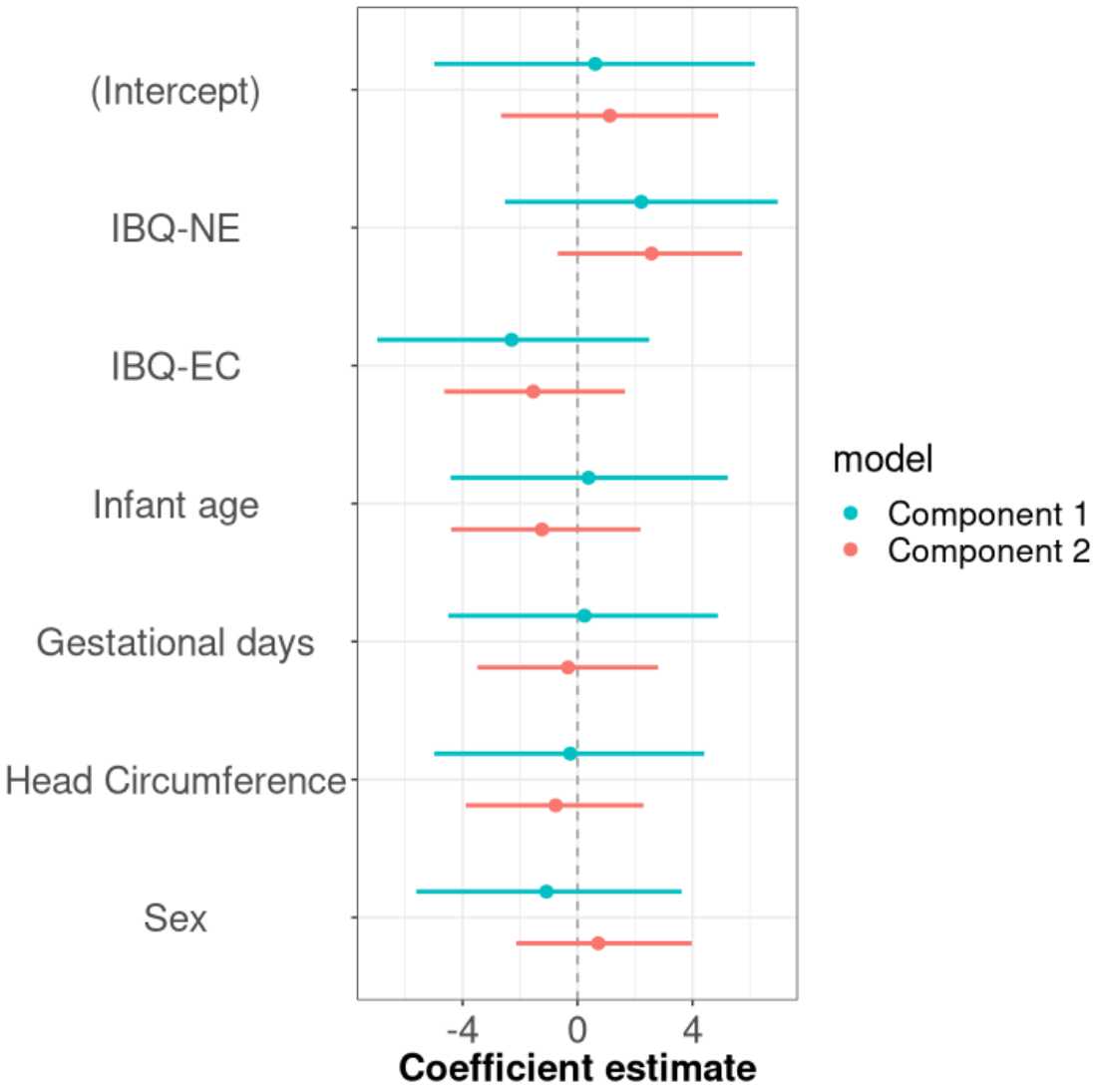

Figure 8: Logistic coefficient estimates and 95% credible intervals corresponding to each covariate of the three-component model.

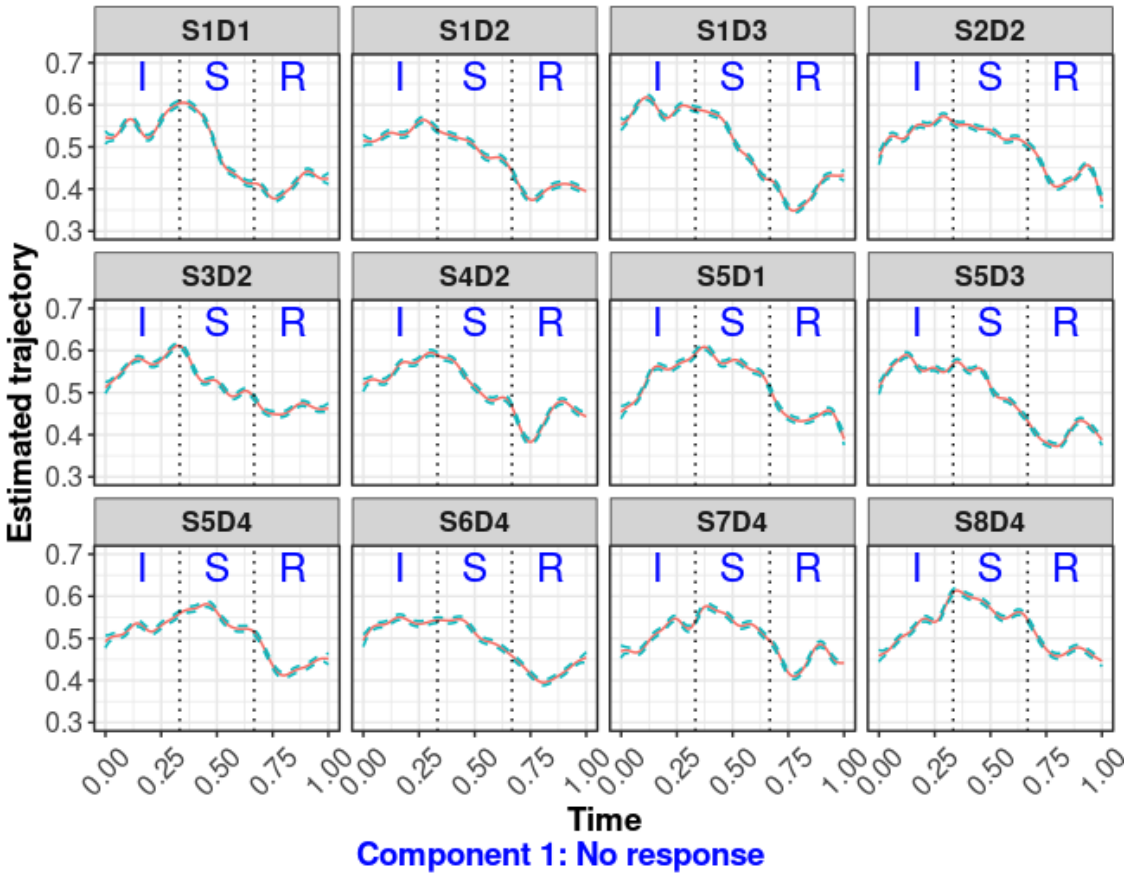

Figure 9: Estimated trajectories of the first component for the three-component model with all twelve channels. **I**: Interact **S**: Still-face **R**: Recovery. Red curves are posterior mean and two green dashed curves are 95% pointwise credible intervals.

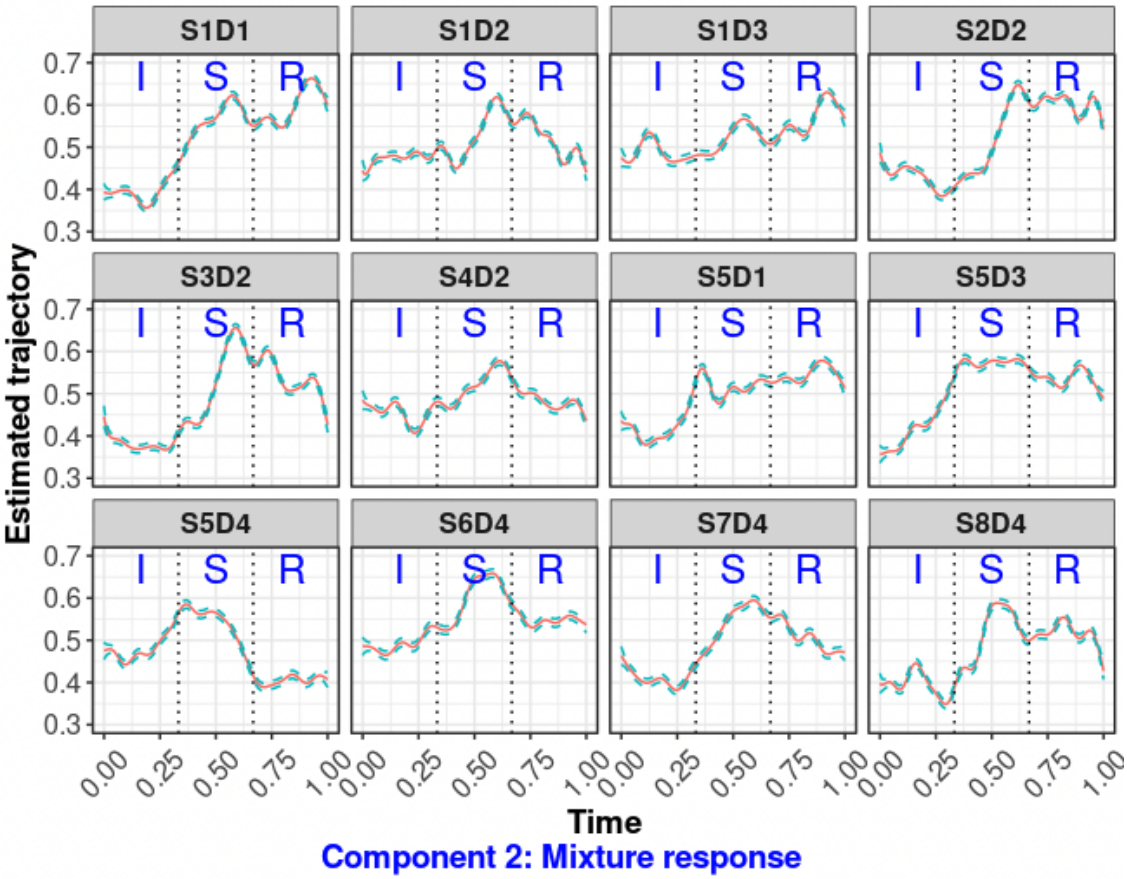

Figure 10: Estimated trajectories of the second component for the three-component model with all twelve channels. **I**: Interact **S**: Still-face **R**: Recovery. Red curves are posterior mean and two green dashed curves are 95% pointwise credible intervals.

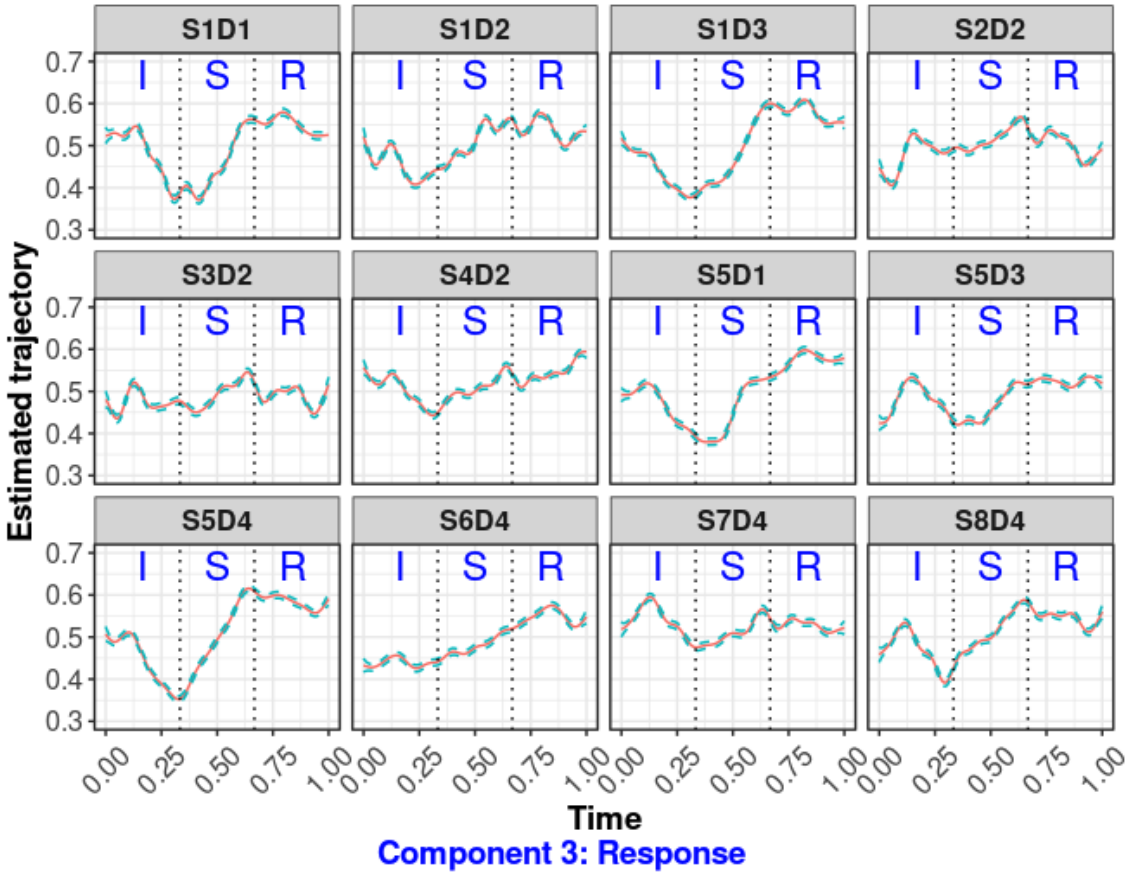

Figure 11: Estimated trajectories of the third component for the three-component model with all twelve channels. **I**: Interact **S**: Still-face **R**: Recovery. Red curves are posterior mean and two green dashed curves are 95% pointwise credible intervals.

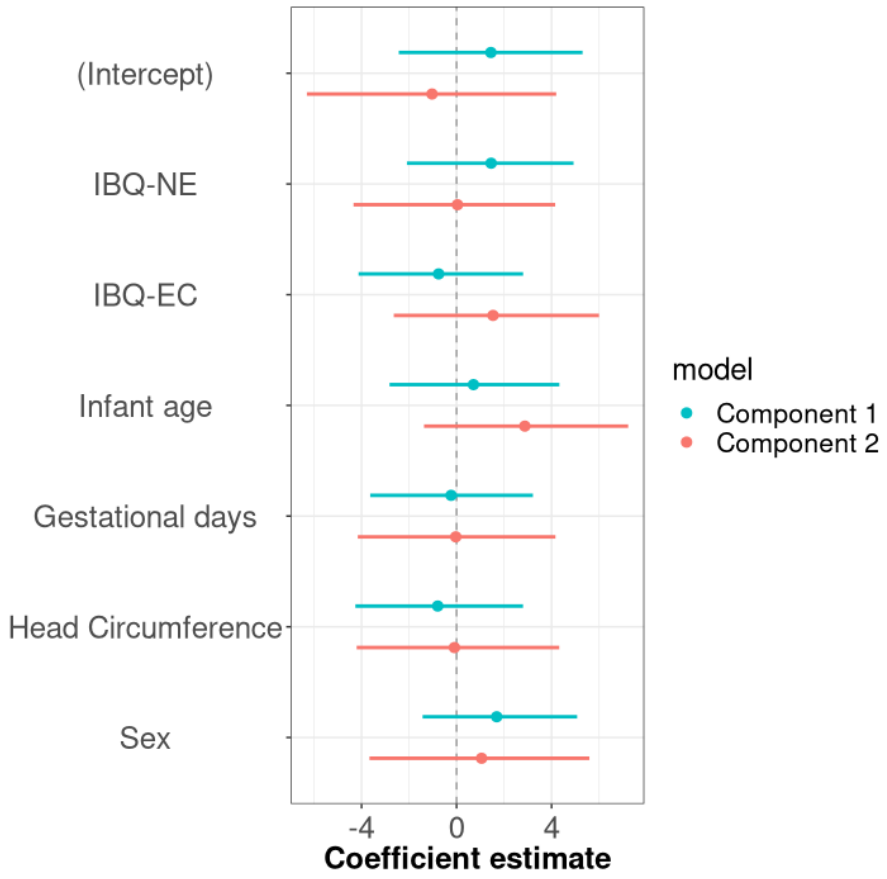

Figure 12: Logistic coefficient estimates and 95% credible intervals for each covariate of the three-component model for all twelve channels.