Supplemental material for A Spatio-Temporal Nonparametric Bayesian Variable Selection Model of fMRI Data for Clustering Correlated Time Courses

Linlin Zhang[a], Michele Guindani[b], Francesco Versace[c], Marina Vannucci[a,*]

[a] Department of Statistics, Rice University, Houston, USA
[b] Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, USA
[c] Department of Behavioral Science, The University of Texas MD Anderson Cancer Center, Houston, USA

Likelihood function and posterior distribution

For clarity purposes, in this section, we review the main components of our model, which we use in the MCMC steps detailed in the next section.

Under model

$$Y_v^* = X_v^* \beta_v + \varepsilon_v^*, \quad \varepsilon_v^* \sim N_T(0, \Sigma_v^*), \tag{1}$$

the likelihood function is

$$f(Y^*|\Theta) \propto \prod_{v=1}^{V} |\psi_v \Sigma_{\alpha_v}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(Y_v^* - X_v^* \beta_v)^T (\psi_v \Sigma_{\alpha_v})^{-1} (Y_v^* - X_v^* \beta_v)\right]$$

where $\Theta = (\beta, \gamma, \lambda, \psi, \alpha)$, Σ_{α_v} is a diagonal matrix with each element set as $(2^{\alpha_v})^{-m}$.

Let $\phi = (\psi, \alpha)$. The full posterior distribution function is obtained via Bayes theorem as

$$\pi(\Theta|Y^*) \propto f(Y^*|\Theta)\pi(\beta|\gamma)\pi(\gamma)\pi(\lambda)\pi(\phi),$$

where

$$\begin{aligned}
&\pi(\beta_v|\gamma_v = 1) \propto \tau^{-\frac{1}{2}} \exp\left(-\frac{\beta_v^2}{2\tau}\right), \\
&\pi(\beta_v|\gamma_v = 0) = \delta_0, \\
&\pi(\gamma_v|\gamma_{-v}) \propto \exp(\gamma_v(d + e \sum_{k \in N_v} \gamma_k)), \\
&\pi(\lambda_v) = \frac{1}{u_2 - u_1} I_{(u_1, u_2)}(\lambda_v), \\
&\pi(\phi_v|\phi_{-v}) = \frac{\eta}{\eta + V - 1} G_0 + \frac{1}{\eta + V - 1} \sum_{j=1}^{m} n_j \delta_{\phi_j^*}.
\end{aligned} \tag{2}$$

In equation (2), $\phi_1^*, \ldots, \phi_m^*$ are the unique values of ϕ_i's for $i \neq v$, and n_j are the frequencies of ϕ_j^* in the vector $(\phi_1, \ldots, \phi_{v-1}, \phi_{v+1}, \ldots, \phi_V)$.

*Corresponding author. Department of Statistics, Rice University, 6100 Main Street, Houston, TX 77005, USA.
Email addresses: lzl7@rice.edu (Linlin Zhang), mguindani@mdanderson.org (Michele Guindani), fversace@mdanderson.org (Francesco Versace), marina@rice.edu (Marina Vannucci)

MCMC algorithm

Update step for (β, γ)

The full conditional distribution of β_v is

$$\begin{aligned} \beta_v | Y_v^*, \psi_v, \alpha_v, \gamma_v = 1 &\sim N(\mu_v, \sigma_v^2), \\ \beta_v | Y_v^*, \psi_v, \alpha_v, \gamma_v = 0 &\sim \delta_0, \end{aligned}$$

where

$$\mu_v = \frac{Y_v^{*T}(\psi_v \Sigma_{\alpha_v})^{-1} X_v^* \tau_v}{X_v^{*T}(\psi_v \Sigma_{\alpha_v})^{-1} X_v^* \tau_v + 1}, \quad (3)$$

$$\sigma_v^2 = \frac{\tau_v}{X_v^{*T}(\psi_v \Sigma_{\alpha_v})^{-1} X_v^* \tau_v + 1}. \quad (4)$$

We perform *Add, Delete, Swap* steps as in (**?**) to jointly update (β, γ):

1) Randomly choose among the three moves below.
   i) *Add*: set $\gamma_v^* = 1$ and sample β_v^* from a $N(\mu_v, \sigma_v^2)$ proposal. Here μ_v and σ_v^2 are as shown in Equations (3) and (4). Position v is randomly chosen from the set of $v's$ where $\gamma_v = 0$ at the previous iteration.
   ii) *Delete*: set $\gamma_v^* = 0, \beta_v^* = 0$. This results in voxel v being excluded in the current iteration. Position v is randomly chosen from among those included in the model at the previous iteration.
   iii) *Swap*: perform both an *Add* and *Delete* move.
   The proposed value (γ^*, β^*) is accepted with probability

$$\begin{aligned} \alpha_{\beta,\gamma} &= \min\left\{1, \frac{\pi(\gamma^*, \beta^* | Y^*, X^*, \psi, \alpha) q(\gamma, \beta | \gamma^*, \beta^*)}{\pi(\gamma, \beta | Y^*, X^*, \psi, \alpha) q(\gamma^*, \beta^* | \gamma, \beta)}\right\} \\ &= \min\left\{1, \frac{f(Y^* | \beta^*, \gamma^*, ...) \pi(\beta^* | \gamma^*) \pi(\gamma^*)}{f(Y^* | \beta, \gamma, ...) \pi(\beta | \gamma) \pi(\gamma)}\right\}. \end{aligned}$$

2) Repeat step (1) m times.
3) Sampling from $N(\mu_v, \sigma_v^2)$ for β_v's such that $\gamma_v = 1$.

When (β^*, γ^*) is obtained via the Add or Delete move, then

$$\frac{\pi(\gamma^*)}{\pi(\gamma)} = \frac{\pi(\gamma_v^* | \gamma_k, k \in N_v)}{\pi(\gamma_v | \gamma_k, k \in N_v)},$$

where N_v is the neighbor of voxel v which is updated in the Add or Delete move.

When (β^*, γ^*) is obtained via the Swap move, if, say, voxels j and l are the ones to be updated, then

$$\frac{\pi(\gamma^*)}{\pi(\gamma)} = \frac{\pi(\gamma_j^* | \gamma_k, k \in N_j) \pi(\gamma_l^* | \gamma_k, k \in N_{l(-j)})}{\pi(\gamma_j | \gamma_k, k \in N_j) \pi(\gamma_l | \gamma_k, k \in N_{l(-j)})},$$

where N_j is the neighbor of voxel j and $N_{l(-j)}$ is the neighbor of voxel l excluding voxel j.

Update step for λ

The full conditional distribution of λ_v is

$$\lambda_v | Y_v^*, \beta_v, \psi_v, \alpha_v \propto \\ \exp\left[-\frac{1}{2}(Y_v^* - X_v^* \beta_v)^T (\psi_v \Sigma_{\alpha_v})^{-1} (Y_v^* - X_v^* \beta_v)\right] I_{(u_1, u_2)}(\lambda_v).$$

We propose $\lambda_v^* \sim U(\lambda_v - h_v, \lambda_v + h_v)$, and the proposed value is accepted with the acceptance probability

$$a_\lambda = \min\left\{1, \frac{\pi(\lambda_v^* | Y_v^*, \beta_v, \psi_v, \alpha_v) q(\lambda_v | \lambda_v^*)}{\pi(\lambda_v | Y_v^*, \beta_v, \psi_v, \alpha_v) q(\lambda_v^* | \lambda_v)}\right\}$$

Update step for (ψ, α)

Let $\phi_v = (\psi_v, \alpha_v)$. The full conditional distribution of ϕ_v is

$$\pi(\phi_v|Y^*, \phi_{-v}, \beta, \lambda, \alpha) = \begin{cases} \phi_j^* & \text{w.p. } b\, n_j\, f(Y_v^*|\lambda_v, \beta_v, \alpha_v, \phi_v = \phi_j^*) \\ h(\phi_v|Y_v^*, \beta_v, \lambda_v, \alpha_v) & \text{w.p. } b\, \eta q_0 \end{cases}$$

where

$$b = \frac{1}{\eta q_0 + \sum_{j=1}^{m} n_j f(Y_v|\lambda_v, \beta_v, \phi_v = \phi_j^*)},$$

$$q_0 = \int G_0(\psi_v, \alpha_v) f(Y_v^*|\psi_v, \alpha_v, \beta_v, \lambda_v) d\psi_v d\alpha_v$$

$$= \int (2\pi)^{-\frac{T}{2}} \times |\Sigma_{\alpha_v}|^{-\frac{1}{2}} \alpha_v^{a_1 - 1}(1-\alpha_v)^{b_1 - 1} \times$$

$$\frac{\Gamma(a_0 + \frac{T}{2})\Gamma(a_1 + b_1) b_0^{a_0}}{(b_0 + \frac{1}{2}(Y_v^* - X_v^*\beta_v)^T \Sigma_{\alpha_v}^{-1}(Y_v^* - X_v^*\beta_v))\Gamma(a_0)\Gamma(a_1)\Gamma(b_1)} d\alpha_v,$$

and

$$h(\phi_v|Y_v^*, \beta_v, \lambda_v) \propto \psi_v^{-a_0 - \frac{T}{2} - 1} \alpha_v^{a_1 - 1}(1-\alpha_v)^{b_1 - 1} |\Sigma_{\alpha_v}|^{-\frac{1}{2}} \times$$

$$\exp\left\{ \frac{-b_0 - \frac{1}{2}(Y_v^* - X_v^*\beta_v)^T \Sigma_{\alpha_v}^{-1}(Y_v^* - X_v^*\beta_v)}{\psi_v} \right\}.$$

Here ϕ_j^*'s and n_j are as defined in Equation (2).

Since q_0 cannot be computed in closed form, we use algorithm 8 proposed by **?** to update (ψ, α), which can be described as follows:

We introduce an auxiliary parameter c_v indicating which "latent cluster" is associated with ϕ_v. Suppose the state of the Markov chain consist of $c = (c_1, \dots, c_V)$, and $\phi = (\phi_c : c \in \{c_1, \dots, c_V\})$, where $\phi_c = (\psi_c, \alpha_c)$, c and ϕ can be updated as follows:

a) For $v = 1, \dots, V$: let k^- be the number of distinct c_{-v}, where c_{-v} is a set of c_j's for $j \neq v$, if $c_v \in c_{-v}$, relabel these c_{-v} with values in $\{1, \dots, k^-\}$, then draw values independently from G_0 for ϕ_{k^-+1}; if $c_v \notin c_{-v}$, relabel these c_{-v} with values in $\{1, \dots, k^-\}$, let c_v have label $k^- + 1$, then $\phi_{k^-+1} = \phi_v$. Draw a new value for c_v from $\{1, \dots, k^- + 1\}$ using the following probabilities:

$$\pi(c_v = c|Y_v^*, c_{-v}, \beta_v, \lambda_v, \phi_1, \dots, \phi_{k^-+1}) = \begin{cases} b\, \frac{n_{-v,c}}{V-1+\eta} F(Y_v^*, \phi_c) & \texttt{for } 1 \le c \le k^- \\ b\, \frac{\eta}{V-1+\eta} F(Y_v^*, \phi_c) & \texttt{for } c = k^- + 1 \end{cases}$$

where $n_{-v,c}$ is the number of c_j for $j \neq v$ that are equal to c, and b is the appropriate normalizing constant.

b) For all $c \in \{c_1, \dots, c_V\}$: sample a new value from $\phi_c|$ all Y_v^* for which $c_v = c$, that is, from the posterior distribution based on the prior G_0 and all the data points currently associated with latent cluster c. Since the posterior distribution of ϕ_c is not in a closed form, we perform Metropolis-Hasting algorithm in this part to approximate the distribution of ϕ_c. The procedure is as below: suppose the number of repeats of value ϕ_c is n_c, all the data points associated with class c is $Y_c^* = (Y_{v_1}^*, \dots, Y_{v_{n_c}}^*)^T$, β_c, λ_c are defined similarly, the posterior distribution is

$$\pi(\alpha_c, \psi_c|Y_c^*, \beta_c, \lambda_c) \propto \psi_c^{-\frac{Tn_c}{2} - a_0 - 1} \alpha_c^{a_1 - 1}(1-\alpha_c)^{b_1 - 1} |\Sigma_{\alpha_c}|^{-\frac{n_c}{2}} \times$$

$$\exp\left\{ \frac{-b_0 - \frac{1}{2}\sum_{i=1}^{n_c}(Y_{vi}^* - X_{vi}^*\beta_{vi})^T \Sigma_{\alpha_c}^{-1}(Y_{vi}^* - X_{vi}^*\beta_{vi})}{\psi_c} \right\}$$

Propose $\alpha_c^* \sim N(\alpha_c, \sigma_0^2)$, and $\psi_c^* \sim N(\psi_c, \sigma_0^2)$, then compute the acceptance rate

$$a_{\psi,\alpha} = \min\left\{1, \frac{\pi(\psi_c^*, \alpha_c^* | Y_c^*, \beta_c, \lambda_c) q(\psi_c, \alpha_c | \psi_c^*, \alpha_c^*)}{\pi(\psi_c, \alpha_c | Y_c^*, \beta_c, \lambda_c) q(\psi_c^*, \alpha_c^* | \psi_c, \alpha_c)}\right\}$$

The proposed value (ψ_c^*, α_c^*) is accepted with probability $a_{\psi,\alpha}$.

Event-related Design

Figure 1 shows the posterior activation map, and the posterior mean estimates for the parameters β and λ for the event-related design in the simulation described in section 3.1 of the main text. Plot (a) in Figure 2 shows the resulting clustering of the voxels for the event-related design. There are 3 clusters in plot (a). The posterior mean maps for the parameter ψ and α are shown in plots (b) and (c) of Figure 2 for event-related design. We find the good estimates to the true values of all the parameters.

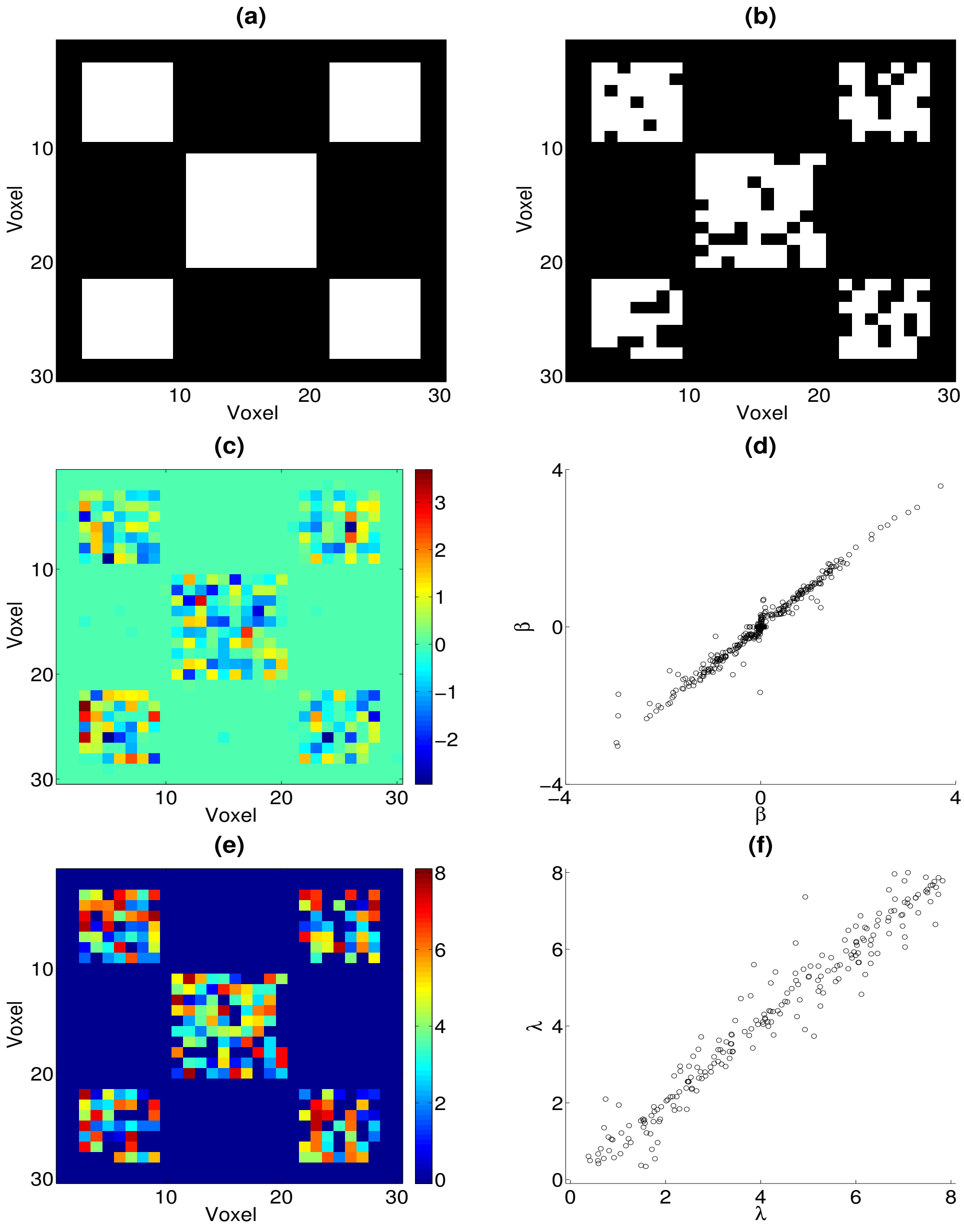

Figure 1: **Simulated data with event-related design:** (a) True map of the activation indicators γ; (b) Posterior activation map obtained by assigning value 1 to those voxels with $p(\gamma_v = 1|y) > 0.8$ and value 0 otherwise; (c) Posterior mean map of β; (d) Scatter plot of posterior mean estimates vs. true values for β; (e) Posterior mean map for λ;(f) same as (d) for λ.

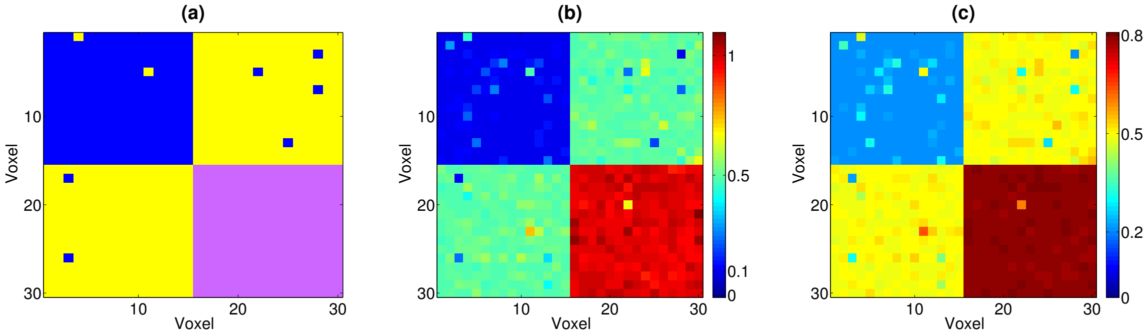

Figure 2: **Simulated data with event-related design:** (a) Posterior clustering map - different colors correspond to different clustering allocations; (b) Posterior mean map of ψ; (c) Posterior mean map of α.

A case study for fMRI data

Figure 3 shows the fit of the time series for one active voxel on each of the slices V1, V5, and PP. In the plot, the continuous black curves represent the real time series and the dashed black curves represent the fitted response.

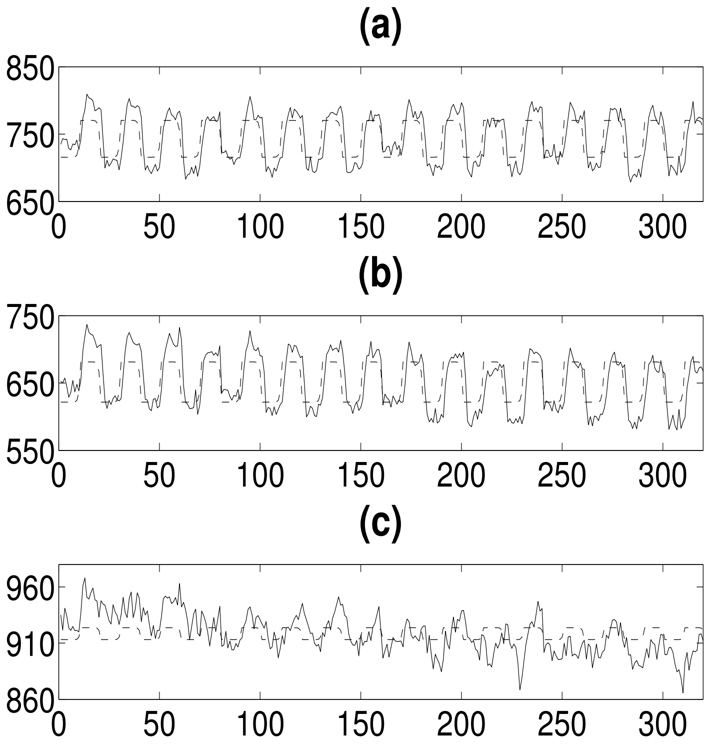

Figure 3: **Real fMRI data:** Time series fitting for one active voxel on (a) V1, (b) V5, and (c) PP. The continuous black curves represent the real time series and the dashed black curves represent the fitted response.