SUPPLEMENTARY MATERIAL FOR THE PAPER " A SPATIO-TEMPORAL NONPARAMETRIC BAYESIAN MODEL OF MULTI-SUBJECT FMRI DATA"

BY LINLIN ZHANG[†], MICHELE GUINDANI[‡], FRANCESCO VERSACE[§], JEFFREY M. ENGELMANN[¶] AND MARINA VANNUCCI[†,*]

Rice University[†], MD Anderson Cancer Center[‡], The University of Oklahoma Health Sciences Center[§] and MD Anderson Cancer Center[¶]

MCMC algorithm. We describe the MCMC algorithm in detail here. For updating the selection parameters $\gamma_{i\nu}$ and regression coefficients $\beta_{i\nu}, i = 1, \ldots, N, \nu = 1, \ldots, V$, we first generate number of subjects n from from Poisson distribution with mean parameter $N/2$, with N the total number of subjects. If $0 < n \leq N$, then we stop and select n subjects with simple random sampling method without replacement; If $n = 0$ or $n > N$, then we resample until $0 < n \leq N$. We then update the values of $\gamma_{i\nu}$ and $\beta_{i\nu}$, $\nu = 1, \ldots, V$ for each of the selected subjects with a combination of *add-delete-swap* moves and the sampling algorithm for HDP models proposed by Teh et al. (2006). The updates on delay parameters $\lambda_{i\nu}$, innovation variance parameters $\psi_{i\nu}$, and long memory parameters $\alpha_{i\nu}$, $\nu = 1, \ldots, V$ are for all subjects. To give the details of equations involved in MCMC steps, we introduce the following notations:

- $c_{i\nu}$: index variable of the "latent cluster" for voxel ν within subject i, $\tilde{c} = (c_{i\nu} : \forall i, \nu)$, $\tilde{c}_{-i\nu} = \tilde{c} \backslash c_{i\nu}$;
- s_{ic}: index variable of the mixture component taken by cluster c within subject i, $\tilde{s} = (s_{ic} : \forall i, c)$, $s_{-ic} = \tilde{s} \backslash s_{ic}$;
- $z_{i\nu}$: equivalent to $s_{ic_{i\nu}}$, and $\tilde{z} = (z_{i\nu} : \forall i, \nu)$;
- m_{ik}: number of clusters within subject i taking component k, and $m_{\cdot k}$ is the number of clusters taking component k across subjects, $m_{\cdot\cdot}$ is the total number of clusters, $\tilde{m} = (m_{ik} : \forall i, k)$;
- n_{ick}: the number of voxels in subject i at cluster c taking mixture component k, and then $n_{ic\cdot}^{-i\nu}$ is the number of voxels in subject i whose factor is associated with $\phi_{s_{ic}}$, leaving out the data item $Y^*_{i\nu}$;
- ϕ_k: the kth mixture component;
- $\tilde{Y}^* = (Y^*_{i\nu} : \forall i, \nu)$, $\tilde{Y}^*_{ic} = (Y^*_{i\nu} : \forall \nu, c_{i\nu} = c)$, $\tilde{Y}^*_{-i\nu} = \tilde{Y}^* \backslash Y^*_{i\nu}$.

*Corresponding author. Department of Statistics, Rice University, 6100 Main Street, Houston, TX 77005, USA

***Updating* γ *and* β.** Let $S_n = \{i_1, i_2, \ldots, i_n\}$ be the set of indices of selected subjects, for subject $i \in S_n$, we randomly choose among the three moves below:

- *add* step. we randomly choose one voxel ν from the set where $\gamma_{i\nu} = 0$ at the previous iteration, and set $\gamma_{i\nu}^{\text{new}} = 1$. Then we update $\beta_{i\nu}^{\text{new}}$ from the following steps:

  Sampling c: The conditional distribution of $c_{i\nu}$ is

  $$p(c_{i\nu} = c|\tilde{c}_{-i\nu}, \tilde{s}) \propto \begin{cases} n_{ic\cdot}^{-i\nu} f_{s_{ic}}^{-Y_{i\nu}^*}(Y_{i\nu}^*), & \texttt{if c is previously used} \\ \eta_1 p(Y_{i\nu}^*|\tilde{c}_{-i\nu}, c_{i\nu} = c^{new}, \tilde{s}), & \texttt{if } c = c^{new} \end{cases},$$

  where $f_s^{-Y_{i\nu}^*}(Y_{i\nu}^*)$ is the conditional density under mixture component s given all data items except $Y_{i\nu}^*$, which is given by

  $$\begin{aligned} f_s^{-Y_{i\nu}^*}(Y_{i\nu}^*) &= (2\pi)^{-\frac{T}{2}} |\Sigma_{i\nu}^*|^{-\frac{1}{2}} \left[\frac{B_{xx} + 1/\tau}{B_{xx} + b_{xx} + 1/\tau}\right]^{\frac{1}{2}} \times \\ & \exp\left[-\frac{1}{2}\left(-\frac{(B_{xy} + b_{xy})^2}{B_{xx} + b_{xx} + 1/\tau} + \frac{B_{xy}^2}{B_{xx} + 1/\tau} + b_{yy}\right)\right], \end{aligned}$$

  where $b_{xx} = X_{i\nu}^{*T} \Sigma_{i\nu}^{*-1} X_{i\nu}^*$, $b_{xy} = Y_{i\nu}^{*T} \Sigma_{i\nu}^{*-1} X_{i\nu}^*$, $b_{yy} = Y_{i\nu}^{*T} \Sigma_{i\nu}^{*-1} Y_{i\nu}^*$, $B_{xx} = \sum_{i'\nu' \neq i\nu, z_{i'\nu'} = s} X_{i'\nu'}^{*T} \Sigma_{i'\nu'}^{*-1} X_{i'\nu'}^*$, $B_{xy} = \sum_{i'\nu' \neq i\nu, z_{i'\nu'} = s} Y_{i'\nu'}^{*T} \Sigma_{i'\nu'}^{*-1} X_{i'\nu'}^*$. Furthermore,

  $$p(Y_{i\nu}^*|\tilde{c}_{-i\nu}, c_{i\nu} = c^{new}, \tilde{s}) = \sum_{s=1}^{K} \frac{m_{\cdot s}}{m_{\cdot\cdot} + \eta_2} f_s^{-Y_{i\nu}^*}(Y_{i\nu}^*) + \frac{\eta_2}{m_{\cdot\cdot} + \eta_2} f_{s^{new}}^{-Y_{i\nu}^*}(Y_{i\nu}^*),$$

  where K is the number of mixture components, and

  $$f_{s^{new}}^{-Y_{i\nu}^*}(Y_{i\nu}^*) = (2\pi)^{-\frac{T}{2}} \tau^{-\frac{1}{2}} |\Sigma_{i\nu}^*|^{-\frac{1}{2}} \sqrt{\frac{1}{1/\tau + b_{xx}}} \exp\left[\frac{1}{2}\left(\frac{b_{xy}^2}{b_{xx} + 1/\tau} - b_{yy}\right)\right].$$

  If the sampled value of $c_{i\nu}$ is one of the previously used values c_0, then $s_{ic_{i\nu}} = s_{ic_0}$, $\beta_{i\nu}^{\text{new}} = \phi_{s_{ic_0}}$; if the sampled value of $c_{i\nu}$ is c^{new}, we obtain a sample of $s_{ic^{new}}$ from:

  $$p(s_{ic^{new}} = s|\tilde{c}, \tilde{s}_{-ic^{new}}) \propto \begin{cases} m_{\cdot s} f_s^{-Y_{i\nu}^*}(Y_{i\nu}^*) & \texttt{if s is previously used} \\ \eta_2 f_{s^{new}}^{-Y_{i\nu}^*}(Y_{i\nu}^*) & \texttt{if } s = s^{new} \end{cases},$$

  If the sampled value of $s_{ic^{new}}$ is one of the previously used value s_0, then $\beta_{i\nu}^{\text{new}} = \phi_{s_0}$; if the sampled value of $s_{ic^{new}}$ is s^{new}, then sample $\beta_{i\nu}^{\text{new}}$ from $\texttt{Normal}(\mu_0, \sigma_0^2)$, where

  $$\mu_0 = \frac{b_{xy}}{b_{xx} + 1/\tau}, \; \sigma_0^2 = \frac{1}{b_{xx} + 1/\tau}.$$

- *delete* step. We randomly choose one voxel ν from the set where $\gamma_{i\nu} = 1$ at the previous iteration, and set $\gamma^{\text{new}} = 0, \beta_{i\nu}^{\text{new}} = 0$.
- *swap* step. We randomly choose one voxel ν_1, and change the value of $\gamma_{i\nu_1}$ from 1 to 0, and then assign the value of 0 to $\beta_{i\nu_1}^{\text{new}}$. Simultaneously, we choose another voxel ν_2, and change the value of $\gamma_{i\nu_2}$ from 0 to 1, and then update $\beta_{i\nu_2}^{\text{new}}$ in the same way as that in *add* step.

The proposed move is accepted with probability

$$\min\left\{1, \frac{f(Y^*|\beta^{\text{new}}, \gamma^{\text{new}}, \lambda, \psi, \alpha)\pi(\beta^{\text{new}}|\gamma^{\text{new}})\pi(\gamma^{\text{new}})}{f(Y^*|\beta^{\text{old}}, \gamma^{\text{old}}, \lambda, \psi, \alpha)\pi(\beta^{\text{old}}|\gamma^{\text{old}})\pi(\gamma^{\text{old}})}\right\}.$$

We repeat the *add-delete-swap* steps q times. For all $k = 1, \ldots, K$, we sample a new value for ϕ_k from the posterior distribution based on the prior P_0 and all the data points associated with component k. For those data items, let $I_k = \{i_1, \ldots, i_{l_k}\}$, $V_k = \{\nu_1, \ldots, \nu_{n_k}\}$ be the indices for subjects and voxels, respectively. We resample ϕ_k from $N(\mu_k, \sigma_k^2)$, where

$$\begin{aligned}
\mu_k &= \frac{\sum_{i\in I_k}\sum_{\nu\in V_k} Y_{i\nu}^{*T}\Sigma_{i\nu}^{*-1}X_{i\nu}^*}{\sum_{i\in I_k}\sum_{\nu\in V_k} X_{i\nu}^{*T}\Sigma_{i\nu}^{*-1}X_{i\nu}^* + 1/\tau}, \\
\sigma_k^2 &= \frac{1}{\sum_{i\in I_k}\sum_{\nu\in V_k} X_{i\nu}^{*T}\Sigma_{i\nu}^{*-1}X_{i\nu}^* + 1/\tau},
\end{aligned}$$

and assign the updated value of ϕ_k to all the $\beta_{i\nu}$'s taking mixture component k.

Updating λ**.** The full conditional distribution of $\lambda_{i\nu}, i = 1, \ldots, N, \nu = 1, \ldots, V$ is

$$\begin{aligned}
&\lambda_{i\nu}|Y_{i\nu}^*, \beta_{i\nu}, \psi_{i\nu}, \alpha_{i\nu} \propto \\
&\exp\left[-\frac{1}{2}(Y_{i\nu}^* - X_{i\nu}^*\beta_{i\nu})^T\Sigma_{i\nu}^{*-1}(Y_{i\nu}^* - X_{i\nu}^*\beta_{i\nu})\right] I_{(u_1,u_2)}(\lambda_{i\nu}).
\end{aligned}$$

We propose $\lambda_{i\nu}^{\text{new}} \sim U(\lambda_{i\nu}^{\text{old}} - h, \lambda_{i\nu}^{\text{old}} + h)$, and the proposed value is accepted with the acceptance probability

$$\min\left\{1, \frac{\pi(\lambda_{i\nu}^{\text{new}}|Y_{i\nu}^*, \beta_{i\nu}, \psi_{i\nu}, \alpha_{i\nu})q(\lambda_{i\nu}^{\text{old}}|\lambda_{i\nu}^{\text{new}})}{\pi(\lambda_{i\nu}^{\text{old}}|Y_{i\nu}^*, \beta_{i\nu}, \psi_{i\nu}, \alpha_{i\nu})q(\lambda_{i\nu}^{\text{new}}|\lambda_{i\nu}^{\text{old}})}\right\}.$$

Updating ψ**.** We propose $\psi_{i\nu}^{\text{new}}$ from truncated normal distribution $N(\psi_{i\nu}^{\text{old}}, \sigma_\psi^2)$ with support $(0, \infty)$, and accept it with probability

$$\min\left\{1, \frac{\pi(\psi_{i\nu}^{\text{new}}|Y_{i\nu}^*, \beta_{i\nu}, \alpha_{i\nu})q(\psi_{i\nu}^{\text{old}}|\psi_{i\nu}^{\text{new}})}{\pi(\psi_{i\nu}^{\text{old}}|Y_{i\nu}^*, \beta_{i\nu}, \alpha_{i\nu})q(\psi_{i\nu}^{\text{new}}|\psi_{i\nu}^{\text{old}})}\right\}.$$

where

$$\pi(\psi_{i\nu}|Y^*_{i\nu},\beta_{i\nu},\alpha_{i\nu}) \propto \psi_{i\nu}^{-(a_0+T/2)-1}\exp\left(-\frac{b_0+d_{xy}}{\psi_{i\nu}}\right)$$

with $d_{xy}=\frac{1}{2}(Y^*_{i\nu}-X^*_{i\nu}\beta_{i\nu})^T(\text{diag}((2^{\alpha_{i\nu}})^m))(Y^*_{i\nu}-X^*_{i\nu}\beta_{i\nu})$.

Updating α**.** We propose $\alpha_{i\nu}^{\text{new}}$ from truncated normal distribution $N(\alpha_{i\nu}^{\text{old}},\sigma_\alpha^2)$ with support $(0,1)$, accept it with probability

$$\min\left\{1,\frac{\pi(\alpha_{i\nu}^{\text{new}}|Y^*_{i\nu},\beta_{i\nu},\lambda_{i\nu})q(\alpha_{i\nu}^{\text{old}}|\alpha_{i\nu}^{\text{new}})}{\pi(\alpha_{i\nu}^{\text{old}}|Y^*_{i\nu},\beta_{i\nu},\lambda_{i\nu})q(\alpha_{i\nu}^{\text{new}}|\alpha_{i\nu}^{\text{old}})}\right\},$$

where

$$\pi(\alpha_{i\nu}|Y^*_{i\nu},\beta_{i\nu},\lambda_{i\nu}) \propto \frac{|\Sigma_{\alpha_{i\nu}}|^{-1/2}\alpha_{i\nu}^{a_1-1}(1-\alpha_{i\nu})^{b_1-1}}{\left(b_0+\frac{1}{2}(Y^*_{i\nu}-X^*_{i\nu}\beta_{i\nu})^T\Sigma^{-1}_{\alpha_{i\nu}}(Y^*_{i\nu}-X^*_{i\nu}\beta_{i\nu})\right)^{a_0+T/2}}.$$

Variational Bayes Algorithm. The second inference strategy is to combine the variational inference and importance sampling algorithm. In the outer loop of the algorithm, we update $\lambda_{i\nu}$ and $\alpha_{i\nu}$, $i=1,\dots,N,\nu=1,\dots,V$ withe the importance sampling procedure; in the inner loop of the algorithm, we update $\beta_{i\nu},\gamma_{i\nu}$ and $\psi_{i\nu}$ via variational Bayes method. First we introduce the notations as follows:

- $c_{i\nu}$: index variable of the "latent cluster" to which the voxel ν belong within subject i;
- s_{ic}: index variable of the mixture component with which cluster c in subject i is associated;
- K_0: truncation on number of mixture components;
- C_0: truncation on number of clusters within subjects;
- $\xi'=(\xi'_k)_{k=1}^{K_0}$: top-level stick proportions;
- $\pi'_i=(\pi'_{ic})_{c=1}^{C_0}$: bottom-level stick proportions;
- $\phi=(\phi_k)_{k=1}^{K_0}$: mixture components or atoms;
- A_i: index set for active voxels in subject i.

The outline of the algorithm is as follows:

(Outer loop)
For m=1, …, M,

- Draw $(\alpha_{i\nu}^{(m)},\lambda_{i\nu}^{(m)})$ from importance sampling distribution $P(\alpha,\lambda)=\text{Uniform}(0,1)\times\text{Uniform}(u_1,u_2)$.

  (Inner loop)

- For $i = 1, \ldots, N, \nu \in A_i$
  - Repeat until convergence
    - Iteratively update the variational distribution of $\phi_k, \xi'_k, \pi'_{ic}$, $k = 1, \ldots, K_0, c = 1, \ldots, C_0$.
  - Update $c_{i\nu} \sim$ Multinomial$(1, \{o_{i\nu c}\}_{c=1}^{C_0})$.
  - Update $s_{ic} \sim$ Multinomial$(1, \{w_{ick}\}_{k=1}^{K_0})$.
  - Estimate $\beta_{i\nu}^{(m)}$ as $\phi_{s_{ic_{i\nu}}}$ with the generated $c_{i\nu}$ and s_{ic} in the previous steps.
- For $i = 1, \ldots, N, \nu = 1, \ldots, V$
  - Update variational distribution of $\psi_{i\nu}$.
  - Update variational distribution $q(\gamma_{i\nu}^{(m)} = 1)$.
  - Estimate $\psi_{i\nu}^{(m)}$ as the mean of its variational distribution.
  - Compute importance weights $\hat{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) = \frac{q(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})}{\tilde{p}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})}$ with $q(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})$ the variational distribution of $(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})$, and $\tilde{p}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})$ the importance sampling density of $(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})$.
  - Normalize importance weights $\hat{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}))$, denoted by $\tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)})$.

Average over parameters

- $\hat{q}(\gamma_{i\nu} = 1) = \sum_{m=1}^{M} \tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) q(\gamma_{i\nu}^{(m)} = 1)$.
- $\hat{\beta}_{i\nu} = \sum_{m=1}^{M} \tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) \beta_{i\nu}^{(m)}$.
- $\hat{\psi}_{i\nu} = \sum_{m=1}^{M} \tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) \psi_{i\nu}^{(m)}$.
- $\hat{\alpha}_{i\nu} = \sum_{m=1}^{M} \tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) \alpha_{i\nu}^{(m)}$.
- $\hat{\lambda}_{i\nu} = \sum_{m=1}^{M} \tilde{w}(\alpha_{i\nu}^{(m)}, \lambda_{i\nu}^{(m)}) \lambda_{i\nu}^{(m)}$.

By the coordinate descent updates, the optimal variational distribution of ϕ is given by

$$q(\phi) = \prod_{k=1}^{K_0} \text{Normal}(r_{1k}/r_{2k}, 1/r_{2k}),$$

where

$$\begin{aligned}
r_{1k} &= \sum_{i=1}^{N} \sum_{\nu \in A_i} Y_{i\nu}^{*T} E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^{*}] \sum_{c=1}^{C_0} q(c_{i\nu} = c) q(s_{ic} = k), \\
r_{2k} &= \sum_{i=1}^{N} \sum_{\nu \in A_i} E_q[X_{i\nu}^{*T}] E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^{*}] \sum_{c=1}^{C_0} q(c_{i\nu} = c) q(s_{ic} = k) + \frac{1}{\tau}.
\end{aligned}$$

The optimal variational distribution of ξ' is given by

$$q(\xi') = \prod_{k=1}^{K_0} \text{Beta}(u_{1k}, u_{2k}),$$

with

$$u_{1k} = \sum_{i=1}^{N}\sum_{c=1}^{C_0} q(s_{ic} = k) + 1,\ u_{2k} = \sum_{i=1}^{N}\sum_{c=1}^{C_0}\sum_{l=k+1}^{K_0} q(s_{ic} = l) + \eta_2.$$

The optimal variational distribution of π' is

$$q(\pi') = \prod_{i=1}^{N}\prod_{c=1}^{C_0} \text{Beta}(a_{ic}, b_{ic}),$$

where

$$a_{ic} = \sum_{\nu\in A_i} q(c_{i\nu} = c) + 1,\ b_{ic} = \sum_{\nu\in A_i}\sum_{l=c+1}^{C_0} q(c_{i\nu} = l) + \eta_1.$$

The optimal variational distribution of s is

$$q(s) = \prod_{i=1}^{N}\prod_{c=1}^{C_0} \text{Multinomiall}(\{w_{ick}\}_{k=1}^{K_0}),$$

where

$$\begin{aligned} w_{ick} \quad \propto \quad \exp \Bigg[& E_q(\phi_k) \sum_{\nu\in A_i} Y_{i\nu}^{*T} E_q(\Sigma_{i\nu}^{*-1}) E_q(X_{i\nu}^*) q(c_{i\nu} = c) - \\ & \frac{1}{2} E_q(\phi_k^2) \sum_{\nu\in A_i} E_q(X_{i\nu}^{*T}) E_q(\Sigma_{i\nu}^{*-1}) E_q(X_{i\nu}^*) q(c_{i\nu} = c) + \\ & E_q(\log(\xi_k')) + \sum_{l=1}^{k-1} E_q(\log(1-\xi_l')) \Bigg]. \end{aligned}$$

The optimal variational distribution of c is

$$q(c) = \prod_{i=1}^{N}\prod_{\nu\in A_i} \text{Multinomial}(\{o_{i\nu c}\}_{c=1}^{C_0}),$$

where

$$\begin{aligned} o_{i\nu c} \quad \propto \quad & \exp\Bigg[Y_{i\nu}^{*T} E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^*] \sum_{k=1}^{K_0} E_q[\phi_k] q(s_{ic}=k) - \\ & \frac{1}{2} E_q[X_{i\nu}^{*T}] E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^*] \sum_{k=1}^{K_0} E_q[\phi_k^2] q(s_{ic}=k) + \\ & E_q[\log(\pi'_{ic})] + \sum_{l=1}^{c-1} E_q[\log(1-\pi'_{il})]\Bigg]. \end{aligned}$$

The optimal variational distribution of ψ is given by

$$q(\psi) = \prod_{i=1}^{N} \prod_{\nu=1}^{V} \text{Inverse-Gamma}(g_{i\nu}, h_{i\nu}),$$

where

$$\begin{aligned} g_{i\nu} \quad &= \quad a_0 + \frac{T}{2}, \\ h_{i\nu} \quad &= \quad b_0 + \frac{1}{2}\left(Y_{i\nu}^* - E_q[X_{i\nu}^*] \sum_{c=1}^{C_0} \sum_{k=1}^{K_0} E_q[\phi_k] q(c_{i\nu}=c) q(s_{ic}=k)\right)^T \times \\ & \quad E_q[\text{diag}(2^{\alpha_{i\nu}})^m] \left(Y_{i\nu}^* - E_q[X_{i\nu}^*] \sum_{c=1}^{C_0} \sum_{k=1}^{K_0} E_q[\phi_k] q(c_{i\nu}=c) q(s_{ic}=k)\right). \end{aligned}$$

The optimal variational distribution of γ is given by

$$q(\gamma_{i\nu}=1) = \frac{\rho_{i\nu}}{\rho_{i\nu}+1},$$

where

$$\begin{aligned} \rho_{i\nu} \quad = \quad & \exp\Bigg\{\sum_{c=1}^{C_0} \sum_{k=1}^{K_0} \Big[Y_{i\nu}^{*T} E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^*] E_q[\phi_k] q(c_{i\nu}=c) q(s_{ic}=k) - \\ & \frac{1}{2} E_q[X_{i\nu}^{*T}] E_q[\Sigma_{i\nu}^{*-1}] E_q[X_{i\nu}^*] E_q[\phi_k^2] q(c_{i\nu}=c) q(s_{ic}=k)\Big] + \\ & d + e \sum_{l \in N_{i\nu}} q(\gamma_{il}=1)\Bigg\}. \end{aligned}$$

The optimal variational distribution of (α, λ) is given by

$$q(\alpha,\lambda) \propto \exp\left\{\sum_{i=1}^{N}\sum_{\nu=1}^{V}\left[-\frac{1}{2}\log(|\text{diag}(2^{\alpha_{i\nu}})^{-m}|) + \log(\Gamma(a_0+T/2)) - \right.\right.$$
$$\left.(a_0+T/2)\log(\zeta_{i\nu}) + \sum_{i=1}^{N}\sum_{\nu=1}^{V}(a_1-1)\log(\alpha_{i\nu}) + (b_1-1)\log(\alpha_{i\nu})\right\},$$

where

$$\zeta_{i\nu} = \frac{1}{2}(Y^*_{i\nu} - X^*_{i\nu}q(\gamma_{i\nu}=1)\sum_{c=1}^{C_0}\sum_{k=1}^{K_0}E_q[\phi_k]q(c_{i\nu}=c)q(s_{ic}=k))^T\text{diag}(2^{\alpha_{i\nu}})^m \times$$
$$(Y^*_{i\nu} - X^*_{i\nu}q(\gamma_{i\nu}=1)\sum_{c=1}^{C_0}\sum_{k=1}^{K_0}E_q[\phi_k]q(c_{i\nu}=c)q(s_{ic}=k)) + b_0.$$

Our model formulation does not allow analytically tractable updates for these parameters. Therefore, we employ importance sampling algorithm to update α and λ, with importance sampling density $\tilde{p}(\alpha,\lambda) = \frac{1}{u_2-u_1}I_{(0<\alpha<1)}I_{(u_1<\lambda<u_2)}$.

Additional Simulation Study. In order to assess whether performances of our method can be affected by the size and location of the activated regions, we simulated an additional scenario with activated regions at the boundary of the slice and where the size of the activated patches are not even. In brief, we simulated data using the same setting as in Section 4.1 of the paper but with activated patterns chosen as shown in the first column of Figure 1.

We used the same hyperparameter settings as in Section 4.1. Figure 1 shows the activation maps estimated via VB (second column), for one subject for each of the true four activation patterns. Figure 2 shows scatter plots of the posterior estimates of β and λ parameters versus the true values, for the same four subjects of Figure 1. Figure 3 shows scatter plots for the ψ and α parameters. Finally, Table 1 reports results on the detection of activated voxels in terms of accuracy, False Negative Rate (FNR), False Positive Rate (FPR), Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC), averaged over 30 replicates, for each one of the 30 subjects. Overall, results are very similar to those found in Section 4.1 and suggest that the performance of our method is not affected by the size and location of the activated regions.

References.

Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical Dirichlet processes. Journal of the American Statistical Association, 101(476), 2006.

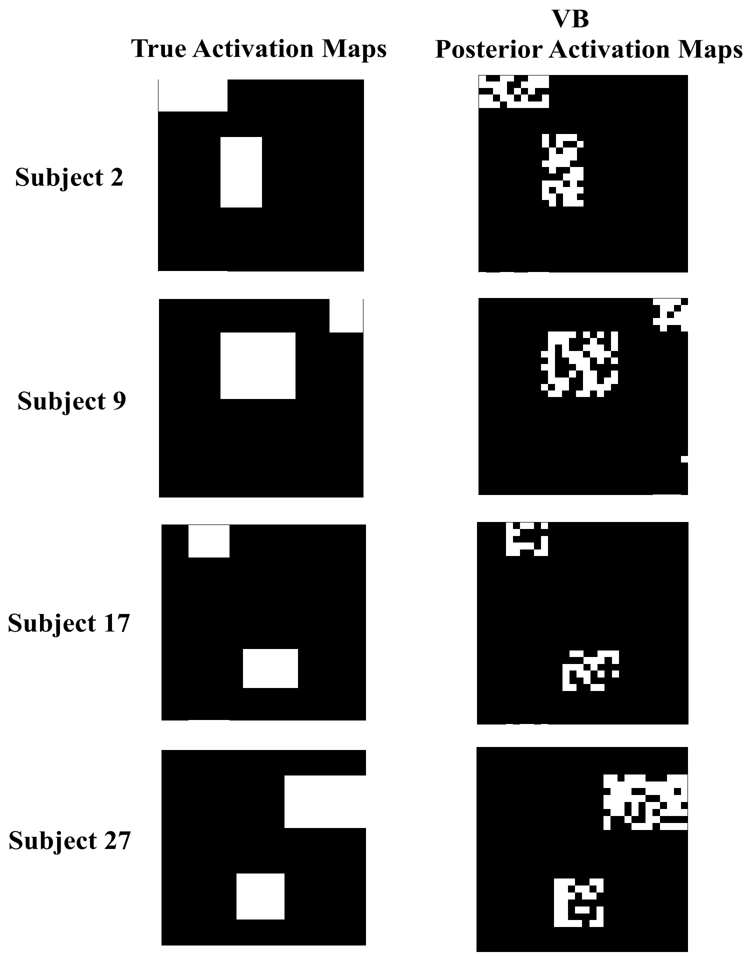

Fig 1: New simulation study: True activation maps (*1st column*), posterior estimated maps estimated via VB (*2nd column*). Results are shown for one subject for each activation pattern.

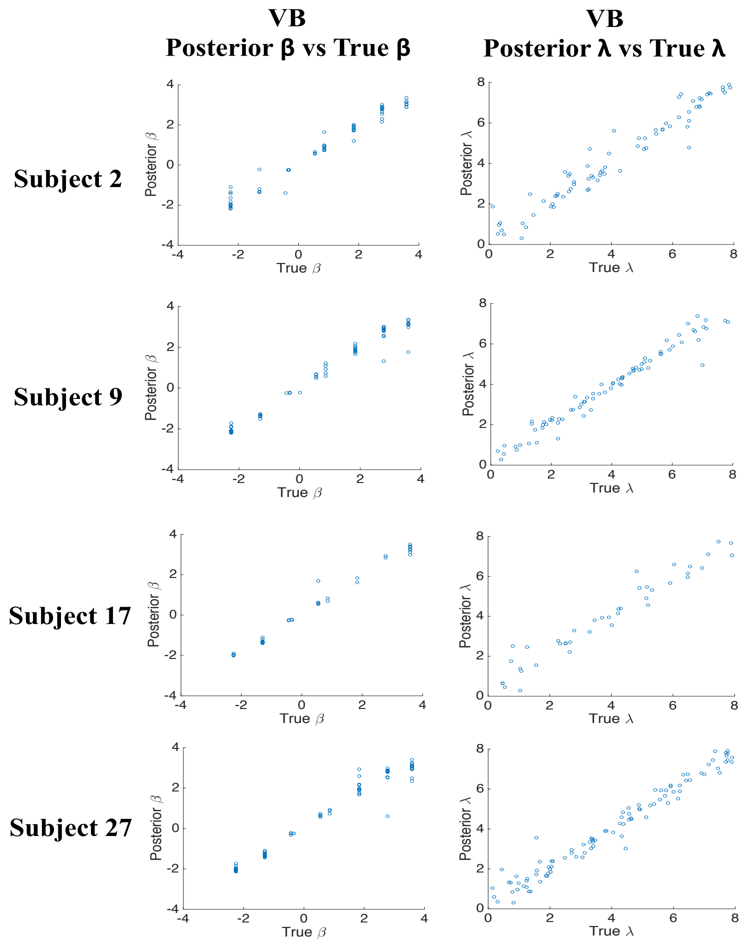

Fig 2: New simulation study: Scatter plots of posterior estimates of the β and λ parameters versus their true values. Results are shown for one subject for each activation pattern.

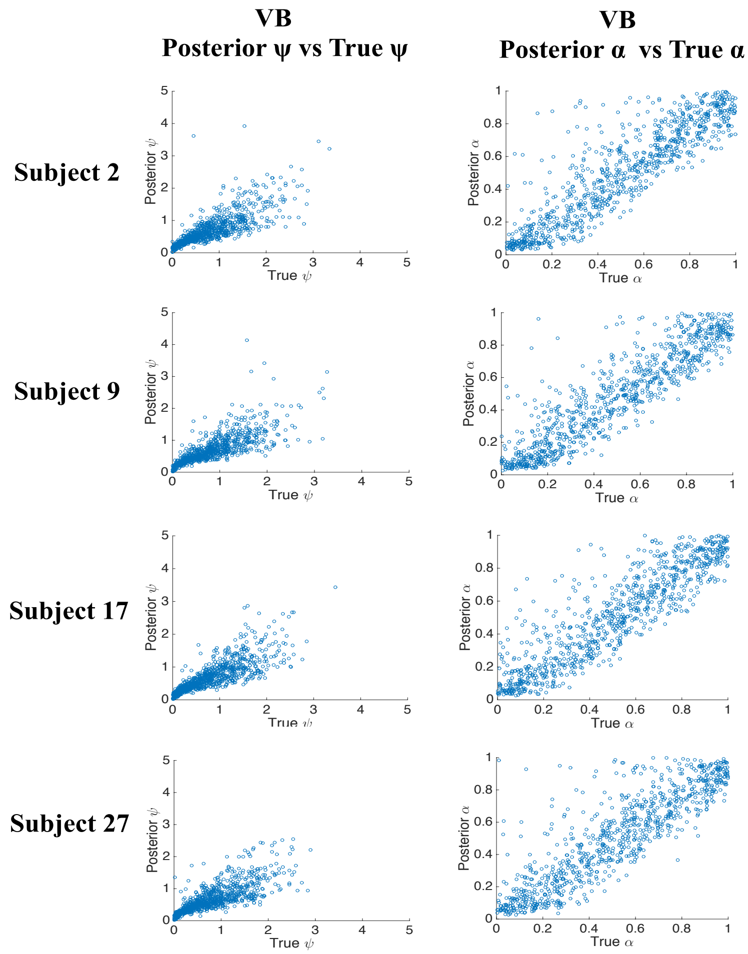

Fig 3: New simulation study: Scatter plots of the posterior estimates of the ψ and α parameters versus their true values. Results are shown for one subject for each activation pattern.

VB

Subject	1	2	3	4	5	6	7	8	9	10
Accuracy(%)	93.400	93.670	93.996	93.970	94.315	94.011	94.552	94.304	93.478	93.152
FNR(%)	50.431	48.161	45.862	45.977	43.333	45.259	41.408	37.259	42.321	45.062
FPR(%)	0.115	0.140	0.106	0.119	0.115	0.179	0.128	0.126	0.205	0.105
MCC	0.672	0.687	0.705	0.704	0.723	0.707	0.736	0.762	0.721	0.708
AUC	0.831	0.849	0.833	0.875	0.851	0.840	0.885	0.866	0.850	0.860
Subject	11	12	13	14	15	16	17	18	19	20
Accuracy(%)	93.130	93.641	93.211	93.222	93.548	96.256	95.559	95.456	96.222	95.582
FNR(%)	44.988	41.062	44.519	44.494	42.346	42.094	50.556	51.282	41.923	50.043
FPR(%)	0.144	0.235	0.131	0.122	0.118	0.105	0.065	0.110	0.158	0.089
MCC	0.707	0.730	0.712	0.712	0.727	0.738	0.681	0.672	0.734	0.683
AUC	0.862	0.839	0.864	0.852	0.860	0.866	0.827	0.831	0.864	0.835
Subject	21	22	23	24	25	26	27	28	29	30
Accuracy(%)	96.359	96.748	92.626	92.530	92.107	91.356	93.804	93.196	92.819	92.878
FNR(%)	40.684	36.111	44.828	45.632	48.276	53.058	37.333	41.333	43.563	43.195
FPR(%)	0.126	0.134	0.181	0.141	0.137	0.115	0.216	0.172	0.194	0.194
MCC	0.746	0.776	0.704	0.700	0.681	0.646	0.756	0.730	0.712	0.715
AUC	0.856	0.883	0.852	0.858	0.854	0.844	0.862	0.861	0.847	0.829

TABLE 1

New simulation study: Detection of the activated voxels in terms of accuracy, False Negative Rate (FNR), False Positive Rate (FPR), Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC) for all 30 subjects, based on Variational Bayes (VB) estimates. Results are given as averages over 30 replicated datasets.

DEPARTMENT OF STATISTICS
RICE UNIVERSITY
HOUSTON, TEXAS 77005
USA,
E-MAIL: linlin.zhangrice@gmail.com.
marina@rice.edu

DEPARTMENT OF BIOSTATISTICS
MD ANDERSON CANCER CENTER
HOUSTON, TEXAS 77030
USA,
E-MAIL: mguindani@mdanderson.org

DEPARTMENT OF FAMILY AND PREVENTIVE MEDICINE
THE UNIVERSITY OF OKLAHOMA HEALTH SCIENCES CENTER
OKLAHOMA CITY, OK 73104
USA,
E-MAIL: Francesco-Versace@ouhsc.edu

DEPARTMENT OF BEHAVIORAL SCIENCE
MD ANDERSON CANCER CENTER
HOUSTON, TEXAS 77030
USA,
E-MAIL: jmengelmann@mdanderson.org