

Supplementary Materials for “Robust and Gaussian Spatial Functional Regression Models for Analysis of Event-Related Potentials” by Zhu et al.

Hongxiao Zhu^{1,*} Francesco Versace² Paul M. Cinciripini²
Philip Rausch³ Jeffrey S. Morris⁴

¹Department of Statistics, Virginia Tech, Blacksburg, VA, USA; *hongxiao@vt.edu*

²Department of Behavioral Science and ⁴Department of Biostatistics, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

³Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany

This supplementary document consists of the following extra materials: Markov chain Monte Carlo (MCMC) algorithms (Appendices A-B), details about CNEG priors (Appendix C), posterior analysis (Appendices D-E), parameter setups (Appendix F), extra results on simulation (Appendices G-H), ERP data analysis (Appendices I-K), sensitivity analysis (Appendix L), as well as results on fitting a global model to the full data set (Appendix M).

1. APPENDIX A: MCMC ALGORITHM FOR THE GFMMC MODEL

Assuming the block-diagonal correlation structure for \mathbf{E}_{jk}^* described in Section 2.2.1 and the CNEG prior for the fixed effect \mathbf{B}^* stated in Section 2.2.2 as well as Appendix C, we design the following MCMC algorithm for posterior inference of the Gfmmc model:

Step 0. Initialize \mathbf{B}^* , $\{q_{jk}^*\}$, $\{s_{jk}^*\}$, $\{\psi_{ajk}\}$, $\{\gamma_{ajk}\}$, $\boldsymbol{\rho}$ based on automatic MLE estimation and set up prior parameters. Here $\boldsymbol{\rho} = (\alpha, \nu)$ or $\boldsymbol{\rho} = \{\alpha_{jk}, \nu_{jk}\}$.

Step 1. Let $\tilde{\mathbf{X}} = \mathbf{X}\boldsymbol{\Gamma}$. Update $\mathbf{b}_{jk}^* \sim N(\boldsymbol{\mu}_{jk}, \mathbf{V}_{jk})$, where $\mathbf{V}_{jk} = (\tilde{\mathbf{X}}^T \boldsymbol{\Sigma}_{jk}^{-1} \tilde{\mathbf{X}} + \boldsymbol{\Psi}_{jk}^{-1})^{-1}$, $\boldsymbol{\mu}_{jk} =$

$\mathbf{V}_{jk} \tilde{\mathbf{X}}^T \Sigma_{jk}^{-1} \mathbf{d}_{jk}$, $\Sigma_{jk} = q_{jk}^* \mathbf{Z} \mathbf{Z}^T + s_{jk}^* \mathbf{R}_{jk}$ and $\Psi_{jk} = \text{diag}\{\psi_{a_{jk}}, a = 1, \dots, p\}$. Here \mathbf{R}_{jk} depends on $\boldsymbol{\rho}$.

Step 2. Update $\{q_{jk}^*\}$ and $\{s_{jk}^*\}$ using Metropolis-Hastings. First propose new values of q_{jk}^*, s_{jk}^* using log transform, e.g., $\log(\tilde{q}_{jk}^*) = \log(q_{jk}^*) + \delta\epsilon$, for $\epsilon \sim N(0, 1)$ and δ is the step size. Second calculate the proposal ratio of the random walk proposal:

$$\frac{f(q_{jk}^* | \tilde{q}_{jk}^*)}{f(\tilde{q}_{jk}^* | q_{jk}^*)} = \frac{\tilde{q}_{jk}^*}{q_{jk}^*}.$$

The new parameters $\{\tilde{q}_{jk}^*, \tilde{s}_{jk}^*\}$ are accepted with probability $\min\{o_{jk}, 1\}$, where

$$o_{jk} = \frac{f(\mathbf{d}_{jk} | \tilde{q}_{jk}^*, \tilde{s}_{jk}^*, \cdot) \pi(\tilde{q}_{jk}^*) \pi(\tilde{s}_{jk}^*) f(q_{jk}^* | \tilde{q}_{jk}^*) f(s_{jk}^* | \tilde{s}_{jk}^*)}{f(\mathbf{d}_{jk} | q_{jk}^*, s_{jk}^*, \cdot) \pi(q_{jk}^*) \pi(s_{jk}^*) f(\tilde{q}_{jk}^* | q_{jk}^*) f(\tilde{s}_{jk}^* | s_{jk}^*)}.$$

Here we have assumed $q_{jk}^* \sim \text{Inv-Gamma}(a_{jk}, b_{jk})$, therefore

$$\log \frac{\pi(\tilde{q}_{jk}^*)}{\pi(q_{jk}^*)} = (a_{jk} + 1)(\log(q_{jk}^*) - \log(\tilde{q}_{jk}^*)) + b_{jk}(1/q_{jk}^* - 1/\tilde{q}_{jk}^*).$$

Similar formula holds for the ratio $\pi(\tilde{s}_{jk}^*)/\pi(s_{jk}^*)$. The log likelihood ratio takes the form

$$\log \frac{f(\mathbf{d}_{jk} | \tilde{q}_{jk}^*, \tilde{s}_{jk}^*, \cdot)}{f(\mathbf{d}_{jk} | q_{jk}^*, s_{jk}^*, \cdot)} = \frac{|\tilde{\Sigma}_{jk}|^{-1/2} \exp\{-1/2(\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)^T \tilde{\Sigma}_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)\}}{|\Sigma_{jk}|^{-1/2} \exp\{-1/2(\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)^T \Sigma_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)\}},$$

where $\Sigma_{jk} = q_{jk}^* \mathbf{Z} \mathbf{Z}^T + s_{jk}^* \mathbf{R}_{jk}$, and \mathbf{R}_{jk} .

Step 3. Update $\boldsymbol{\rho}$ using Metropolis-Hastings. The formulae are the same as the step 5 of Rfmcmc model (in Section 2 of this document), except replacing the formula of Σ_{jk} by $\Sigma_{jk} = q_{jk}^* \mathbf{Z} \mathbf{Z}^T + s_{jk}^* \mathbf{R}_{jk}$.

Step 4. Update \mathbf{U}_{jk} , the (j, k) th column of \mathbf{U} , from $(\mathbf{U}_{jk}^* | \cdot) \sim N(\boldsymbol{\mu}_{\mathbf{U}_{jk}}, \mathbf{V}_{\mathbf{U}_{jk}})$, where

$$\begin{aligned} \mathbf{V}_{\mathbf{U}_{jk}} &= (\mathbf{Z}^T (s_{jk}^* \mathbf{R}_{jk})^{-1} \mathbf{Z} + (\mathbf{P} q_{jk}^*)^{-1})^{-1}, \\ \boldsymbol{\mu}_{\mathbf{U}_{jk}} &= \mathbf{V}_{\mathbf{U}_{jk}} \mathbf{Z}^T (s_{jk}^* \mathbf{R}_{jk})^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*). \end{aligned}$$

This step is optional.

Repeat Steps 1-4 until reaching a pre-specified maximum number of iterations.

Note that the \mathbf{R}_{jk} here in the Gfmcmc model is the correlation matrix, whereas the notation \mathbf{R}_{jk} in Rfmcmc (in Section 2 in this document) contains the scaling parameters $\{\lambda_{l_{jk}}\}$.

2. APPENDIX B: MCMC ALGORITHM FOR THE RFMMC MODEL

We assume the CNEG prior for the fixed effect \mathbf{B}^* as stated in Section 2.2.2 as well as Appendix C. For the residual distribution, in Section 2.2.3 we have assumed that \mathbf{E}_{jk}^* follows a scale-mixture-of-normal setup with a block-diagonal correlation structure, i.e.,

$$\begin{aligned}\mathbf{E}_{jk}^* &\sim N(0, \boldsymbol{\Sigma}_{jk}), \quad \boldsymbol{\Sigma}_{jk} = \text{diag}\{\lambda_{ljk}; l = 1, \dots, \mathcal{L}\} \otimes \mathcal{R}_{jk}, \\ \lambda_{ljk} &\sim \text{Exp}((\nu_{jk}^E)^2/2), \quad (\nu_{jk}^E)^2 \sim \text{Gamma}(a^E, b^E),\end{aligned}$$

where \mathcal{R}_{jk} is the within-block correlation matrix and $\boldsymbol{\lambda}_{jk} = \{\lambda_{1jk}, \dots, \lambda_{\mathcal{L}jk}\}$ contains independent scaling parameters. The joint conditional posterior density of $\boldsymbol{\lambda}_{jk}$ and $\boldsymbol{\rho}_{jk}$ is

$$\begin{aligned}f(\boldsymbol{\lambda}_{jk}, \boldsymbol{\rho}_{jk}|\cdot) &\propto \prod_{l=1}^{\mathcal{L}} f(\mathbf{d}_{ljk}|\mathbf{X}_l \mathbf{B}_{jk}^* + \mathbf{Z}_l \mathbf{U}_{jk}^*, \lambda_{ljk}, \mathcal{R}_{jk}) f(\lambda_{ljk}) f(\boldsymbol{\rho}_{jk}) \\ &\propto \prod_{l=1}^{\mathcal{L}} |\lambda_{ljk} \mathcal{R}_{jk}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \tilde{\mathbf{d}}_{ljk}^T \left(\lambda_{ljk}^{-1} \mathcal{R}_{jk}^{-1}\right) \tilde{\mathbf{d}}_{ljk}\right\} \exp\left\{-\frac{(\nu_{jk}^E)^2}{2} \lambda_{ljk}\right\} f(\boldsymbol{\rho}_{jk}),\end{aligned}\quad (1)$$

where \mathbf{d}_{ljk} , \mathbf{X}_l , \mathbf{Z}_l denote the l th block of \mathbf{d}_{jk} , \mathbf{X} and \mathbf{Z} respectively, and $\tilde{\mathbf{d}}_{ljk} = \mathbf{d}_{ljk} - \mathbf{X}_l \mathbf{B}_{jk}^* - \mathbf{Z}_l \mathbf{U}_{jk}^*$. Based on the above joint density, we find that the conditional distribution of each λ_{ljk} is a generalized-inverse-Gaussian (GIG) distribution (Jørgensen 1982).

Based on these setups, we design the following MCMC algorithm for posterior inference of the Rfmmc model:

Step 0. Initialize $\{\nu_{jk}^E\}$, $\{\nu_{jk}^U\}$ and $\{\lambda_{ljk}\}$, $\{\phi_{mjk}\}$, $\{\psi_{cjk}\}$, $\{r_{cjk}\}$, $\boldsymbol{\rho}$, \mathbf{B}^* based on automatic MLE estimation and set up prior parameters. Here $\boldsymbol{\rho} = (\alpha, \nu)$ or $\boldsymbol{\rho} = \{\alpha_{jk}, \nu_{jk}\}$.

Step 1. Let $\tilde{\mathbf{X}} = \mathbf{X}\boldsymbol{\Gamma}$. Update $\mathbf{b}_{jk}^* \sim N(\boldsymbol{\mu}_{jk}, \mathbf{V}_{jk})$, where $\mathbf{V}_{jk} = (\tilde{\mathbf{X}}^T \boldsymbol{\Sigma}_{jk}^{-1} \tilde{\mathbf{X}} + \boldsymbol{\Psi}_{jk}^{-1})^{-1}$, $\boldsymbol{\mu}_{jk} = \mathbf{V}_{jk} \tilde{\mathbf{X}}^T \boldsymbol{\Sigma}_{jk}^{-1} \mathbf{d}_{jk}$, $\boldsymbol{\Sigma}_{jk} = \mathbf{Z} \boldsymbol{\Phi}_{jk} \mathbf{Z}^T + \mathbf{R}_{jk}$ and $\boldsymbol{\Psi}_{jk} = \text{diag}\{\psi_{cjk}, c = 1, \dots, p\}$.

Step 2. Update \mathbf{U}_{jk}^* , the (j, k) th column of \mathbf{U}^* , from $(\mathbf{U}_{jk}^*|\cdot) \sim N(\boldsymbol{\mu}_{\mathbf{U}_{jk}}, \mathbf{V}_{\mathbf{U}_{jk}})$, where $\mathbf{V}_{\mathbf{U}_{jk}} = (\mathbf{Z}^T \mathbf{R}_{jk}^{-1} \mathbf{Z} + \boldsymbol{\Phi}_{jk}^{-1})^{-1}$, $\boldsymbol{\mu}_{\mathbf{U}_{jk}} = \mathbf{V}_{\mathbf{U}_{jk}} \mathbf{Z}^T \mathbf{R}_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}} \mathbf{b}_{jk}^*)$.

Step 3. Update the scaling parameters $\{\lambda_{ljk}\}$, $\{\phi_{mjk}\}$, $\{\psi_{cjk}\}$, $\{r_{cjk}\}$. In particular, for $l = 1, \dots, \mathcal{L}$, $(\lambda_{ljk}|\cdot) \sim \text{GIG}(\mathbf{A}, \mathbf{B}, q)$, where $\mathbf{A} = (\nu_{jk}^E)^2$, $\mathbf{B} = (\mathbf{d}_{jk} - \tilde{\mathbf{X}}_l \mathbf{b}_{jk}^* - \mathbf{Z}_l \mathbf{U}_{jk}^*)^T \mathcal{R}_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}_l \mathbf{b}_{jk}^* - \mathbf{Z}_l \mathbf{U}_{jk}^*)$, $q = 1 - S/2$, where S is the size of \mathcal{R}_{jk} . Here $\tilde{\mathbf{X}}_l, \mathbf{Z}_l$ are the l th row-block of the matrices $\tilde{\mathbf{X}}, \mathbf{Z}$ respectively. Furthermore, $(\phi_{mjk}^{-1}|\cdot) \sim \text{Inv-Gauss}(\sqrt{(\nu_{jk}^E)^2 / (U_{mjk}^*)^2}, (\nu_{jk}^E)^2)$, $(\psi_{cjk}^{-1}|\cdot) \sim \text{Inv-Gauss}(\sqrt{2r_{cjk} / (b_{cjk}^*)^2}, 2r_{cjk})$ and $(r_{cjk}|\cdot) \sim \text{Gamma}(a_{cjk}^B + 1, \psi_{cjk} + b_{cjk}^B)$.

Step 4. Update the population scale parameters $\{\nu_{jk}^E\}$, $\{\nu_{jk}^U\}$ from

$$(\nu_{jk}^E)^2 | \{\lambda_{ljk}\}_l \sim \text{Gamma}(\mathcal{L} + a^E, 1/2 \sum_{l=1}^{\mathcal{L}} \lambda_{ljk} + b^E),$$

and $(\nu_{jk}^U)^2 \{\phi_{mjk}\}_m \sim \text{Gamma}(M + a^U, 1/2 \sum_{m=1}^M \phi_{mjk} + b^U)$.

Step 5. Update the parameters in $\boldsymbol{\rho}$. Denote by θ one component of $\boldsymbol{\rho}$ (could be α_{jk}, ν_{jk} or α, ν depending on whether using Rfmmc $_{\boldsymbol{\rho}_{jk}}$ or Gfmmc $_{\boldsymbol{\rho}}$). Suppose that we have assumed the prior $\theta \sim \text{Unif}(0, C)$, then $\theta/C \in [0, 1]$. We take the logit transform $\zeta = \log\{(\theta/C)/(1 - \theta/C)\}$ to transfer θ to ζ on the real line, and propose a new value $\tilde{\zeta} = \zeta + \delta\epsilon$, where $\epsilon \sim N(0, 1)$ and δ is the step size. Then using change of variable techniques, we find that

$$\begin{aligned} f(\tilde{\theta}|\theta) &= f(\tilde{\zeta}|\theta) \left| \frac{d\tilde{\zeta}}{d\tilde{\theta}} \right| = (2\pi\delta^2)^{-1/2} \exp \left\{ - \left(\text{logit}\left(\frac{\tilde{\theta}}{C}\right) - \text{logit}\left(\frac{\theta}{C}\right) \right)^2 / (2\delta^2) \right\} \left| \frac{C}{\tilde{\theta}(C - \tilde{\theta})} \right|, \\ f(\theta|\tilde{\theta}) &= (2\pi\delta^2)^{-1/2} \exp \left\{ - \left(\text{logit}(\theta/C) - \text{logit}(\tilde{\theta}/C) \right)^2 / (2\delta^2) \right\} \left| \frac{C}{\theta(C - \theta)} \right|. \end{aligned}$$

Therefore the log proposal ratio is

$$\log \frac{f(\theta|\tilde{\theta})}{f(\tilde{\theta}|\theta)} = \log \frac{\tilde{\theta}}{\theta} + \log \frac{1 - \tilde{\theta}/C}{1 - \theta/C}.$$

We now describe the details of updating $\boldsymbol{\rho}$ in two cases. **Case I:** when $\{\alpha_{jk}, \nu_{jk}\}$ depend on (j, k) . Take (α_{jk}, ν_{jk}) as an example, the proposed value $(\tilde{\alpha}_{jk}, \tilde{\nu}_{jk})$ is accepted with probability $\min\{o_{jk}, 1\}$, with

$$\log(o_{jk}) = \log \frac{f(\mathbf{d}_{jk}|\tilde{\alpha}_{jk}, \tilde{\nu}_{jk}, \cdot) f(\tilde{\alpha}_{jk}) f(\tilde{\nu}_{jk}) \pi(\alpha_{jk}|\tilde{\alpha}_{jk}) \pi(\nu_{jk}|\tilde{\nu}_{jk})}{f(\mathbf{d}_{jk}|\alpha_{jk}, \nu_{jk}, \cdot) f(\alpha_{jk}) f(\nu_{jk}) \pi(\tilde{\alpha}_{jk}|\alpha_{jk}) \pi(\tilde{\nu}_{jk}|\nu_{jk})}$$

where the prior ratios $f(\tilde{\alpha}_{jk})/f(\alpha_{jk})$ and $f(\tilde{\nu}_{jk})/f(\nu_{jk})$ are constant due to the using of uniform priors. The log likelihood ratio is

$$\log \frac{f(\mathbf{d}_{jk}|\tilde{\alpha}_{jk}, \tilde{\nu}_{jk}, \cdot)}{f(\mathbf{d}_{jk}|\alpha_{jk}, \nu_{jk}, \cdot)} = \frac{|\tilde{\boldsymbol{\Sigma}}_{jk}|^{-1/2} \exp\{-1/2(\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)^T \tilde{\boldsymbol{\Sigma}}_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)\}}{|\boldsymbol{\Sigma}_{jk}|^{-1/2} \exp\{-1/2(\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)^T \boldsymbol{\Sigma}_{jk}^{-1} (\mathbf{d}_{jk} - \tilde{\mathbf{X}}\mathbf{b}_{jk}^*)\}},$$

where $\boldsymbol{\Sigma}_{jk} = \mathbf{Z}\boldsymbol{\Phi}_{jk}\mathbf{Z}^T + \mathbf{R}_{jk}$, and \mathbf{R}_{jk} depends on (α_{jk}, ν_{jk}) . **Case II:** when $\boldsymbol{\rho}$ does not depend on (j, k) , then the only parameters we consider is $\boldsymbol{\rho} = (\alpha, \nu)$. The proposed value $(\tilde{\alpha}, \tilde{\nu})$ is accepted with probability $\min\{o, 1\}$, with

$$\log(o) = \log \frac{\prod_{j,k} f(\mathbf{d}_{jk}|\tilde{\alpha}, \tilde{\nu}, \cdot) f(\tilde{\alpha}) f(\tilde{\nu}) \pi(\alpha|\tilde{\alpha}) \pi(\nu|\tilde{\nu})}{\prod_{j,k} f(\mathbf{d}_{jk}|\alpha, \nu, \cdot) f(\alpha) f(\nu) \pi(\tilde{\alpha}|\alpha) \pi(\tilde{\nu}|\nu)}.$$

Repeat Steps 1-5 until reaching a pre-specified maximum number of iterations.

3. APPENDIX C: CORRELATED NORMAL-EXPONENTIAL-GAMMA PRIORS FOR \mathbf{B}^*

In Section 2.2.2, we adopted the correlated Normal-Exponential-Gamma (CNEG) prior for the wavelet domain fixed effect matrix \mathbf{B}^* . Here, we describe more technical details. For our ERP data, the rows of \mathbf{B}^* contain A blocks, and each block corresponds to one stimulus effect at the S electrodes. Let \mathbf{B}_{jk}^* denote the (j, k) th column of \mathbf{B}^* . Borrowing the structural shrinkage idea of Griffin and Brown (2012), we assume that $\mathbf{B}_{jk}^* = \mathbf{\Gamma} \mathbf{b}_{jk}^*$, where $\mathbf{\Gamma}$ is a lower triangular matrix obtained from the Cholesky decomposition of a prior correlation matrix \mathbf{R}_B , i.e., $\mathbf{R}_B = \mathbf{\Gamma} \mathbf{\Gamma}^T$. For example, one may specify \mathbf{R}_B to be a block-diagonal matrix with A diagonal blocks of size S . We assume that entries of \mathbf{b}_{jk}^* are a priori independent. Specifically, denoting the c th component of \mathbf{b}_{jk}^* by b_{cjk}^* , we assume a scale-mixture-of-normal prior for b_{cjk}^* :

$$b_{cjk}^* \sim N(0, \psi_{cjk}), \psi_{cjk} \sim \text{Exp}(r_{cjk}), r_{cjk} \sim \text{Gamma}(a_{jk}^B, b_{jk}^B), a_{jk}^B > 1. \quad (2)$$

From the above prior, one can analytically integrate out the intermediate parameters ψ_{cjk} and r_{cjk} to obtain a normal-exponential-gamma (NEG) distribution for b_{cjk}^* . This distribution has zero mean and variance $\text{var}(b_{cjk}^*) = b_{jk}^B / (a_{jk}^B - 1)$ when $a_{jk}^B > 1$, and can be used as a general shrinkage prior for adaptive variable selection (Griffin and Brown 2012). Following Griffin and Brown (2012), we call the resulting prior for \mathbf{B}_{jk}^* the correlated normal-exponential-gamma (CNEG) prior, denoted by $\mathbf{B}_{jk}^* \sim \text{CNEG}(\mathbf{\Gamma}, a_{jk}^B, b_{jk}^B)$. We set the CNEG prior to all columns of \mathbf{B}^* independently. This prior induces sparsity in the wavelet space, which is known to lead to adaptive regularization in t , and by capturing correlation across s , induces spatial smoothness across s in $B_{as}(t)$ in model (1). The specification of the prior parameter $\mathbf{\Gamma}$ can be achieved through assuming a parametric correlation structure for \mathbf{R}_B and updating the parameter in \mathbf{R}_B via a Metropolis-Hasting sampler. In our implementation, we specified \mathbf{R}_B using a preliminary estimate of $\{\widehat{\mathbf{B}}_{jk}^*\}$ and found that the posterior estimations are not sensitive to a range of parameters in \mathbf{R}_B . Details on the specification of \mathbf{R}_B and the sensitivity analysis are available in Section 6 (Appendix F) and Section 12 (Appendix L) respectively.

We now explain why the CNE prior, i.e., assuming that

$$b_{cjk}^* \sim N(0, \psi_{cjk}), \psi_{cjk} \sim \text{Exp}(\nu_{jk}^2/2)$$

does not induce the prespecified correlation structure \mathbf{R}_B for \mathbf{B}_{jk}^* , whereas the CNEG prior does. For simplicity, we use a slightly different notation. Let $\boldsymbol{\beta}$ be one column of \mathbf{B}^* . Then a CNE prior assumes that $\boldsymbol{\beta} = \mathbf{\Gamma} \mathbf{b}$, independently across all columns. The $\mathbf{\Gamma}$ matrix is known and satisfies $\mathbf{\Gamma} \mathbf{\Gamma}^T = \mathbf{R}_B$. Denote $\mathbf{b} = (b_1, \dots, b_p)^T$. If we assume that $b_i \sim N(0, \psi_i)$, $\psi_i \sim \text{Exp}(\nu_i^2/2)$ independently for all i , then after integrating out ψ_i , we get $b_i \sim DE(0, 1/\nu_i)$. Therefore we have $\text{var}(b_i) = 2/\nu_i^2$. Notice that it is important ν_i to

depend on i because we wish to have different population scale parameters to take effect on different b_i 's. Denote $\mathbf{V} = \text{diag}(2/\nu_1^2, \dots, 2/\nu_p^2)$. The the covariance now becomes

$$\text{cov}(\boldsymbol{\beta}) = \mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^T. \quad (3)$$

Denote $(\mathbf{\Gamma})_{i,j} = \gamma_{i,j}$, and denote the correlation between β_i and β_j by $\rho_{i,j}$, then the above formula implies that

$$\rho_{ij} = \frac{\sum_{k=1}^p \gamma_{ik}\gamma_{jk}(\frac{2}{\nu_k^2})}{\sqrt{\sum_{k=1}^p \gamma_{ik}^2(\frac{2}{\nu_k^2}) \sum_{k=1}^p \gamma_{jk}^2(\frac{2}{\nu_k^2})}}.$$

The above formula implies that all sums in $\rho_{i,j}$ are weighted by the marginal variance of \mathbf{b} , and they will not cancel out unless $\nu_1 \equiv \nu_2 \equiv \dots \equiv \nu_p$. During MCMC, if the $\{\nu_i\}$ parameters will be updated in each iteration, the correlation ρ_{ij} will vary across different iterations. One critical fact is that the correlation formula $\mathbf{\Gamma}\mathbf{V}\mathbf{\Gamma}^T \neq \mathbf{V}^{1/2}\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{V}^{1/2}$. While the formula on the right hand side induces a constant correlation

$$\rho_{ij} = \frac{\sum_{k=1}^I \gamma_{ik}\gamma_{jk}}{\sqrt{\sum_{k=1}^I \gamma_{ik}^2 \sum_{k=1}^I \gamma_{jk}^2}},$$

the formula on the left hand side does not. This fact might be the main reason that in Griffin and Brown (2012), they have to design the $\mathbf{\Gamma}$ matrix carefully to achieve certain correlated structured shrinkage (e.g., fused lasso).

Due to the above reason, the CNE prior does not induce the pre-specified correlation \mathbf{R}_B . We therefore suggest to add one more layer of mixture distribution to ν_i , e.g., assume $\nu_i \sim g(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is common across index i . One example is the CNEG prior stated in (2). As long as the marginal distribution of β_i have finite variance, the \mathbf{V} matrix in (3) equals $\mathbf{V} = c_0\mathbf{I}$ for a constant c_0 , therefore $\text{cov}(\boldsymbol{\beta}) = c_0\mathbf{\Gamma}\mathbf{\Gamma}^T = c_0\mathbf{R}_B$. In the NEG expression in (2), the marginal variance of b_i is obtained using results in Griffin and Brown (2005).

4. APPENDIX D: CALCULATING POSTERIOR PREDICTIVE LIKELIHOOD FOR NEW SUBJECTS

To estimate the log posterior predictive likelihood for a validation data set, a necessary step is to calculate the posterior predictive likelihood for a new subject. The model for a new subject (indexed by superscript s) at wavelet coefficient j (note we have abbreviated the double index in wavelet coefficients to a single index) can be written as

$$\mathbf{d}_j^s = \mathbf{X}^s\mathbf{B}_j + u_j^s\mathbf{1} + \mathbf{E}_j^s,$$

where u_j^s is the random effect for the new subject. Due to the fact that u_j^s can not be estimated during the training procedure, it has to be integrated out when calculating the

posterior predictive likelihood. In Gfmm and Gfmmc models, integrating out the random effect u_j^s results in a Gaussian distribution with covariance matrix $q_j^* \mathbf{J} + s_j^* \mathbf{I}$. In the Rfmm and Rfmmc models, since both u_j^s and components in \mathbf{E}_j^s follow scaled mixture of normals, the analytical solution for the integrated density is not easy to obtain. We therefore approximate the integrated density using Monte Carlo integration. In particular, in the Rfmm case, since all components of \mathbf{E}_j^s are independent $DE(0, 1/\nu_j^E)$, we have

$$f(\mathbf{d}_j^s | \Theta_j) = \int f(\mathbf{d}_j^s | \mathbf{B}_j, u_j^s, \nu_j^E) f(u_j^s | \nu_j^U) du_j^s, \quad (4)$$

where $f(\mathbf{d}_j^s | \mathbf{B}_j, u_j^s, \nu_j^E) = \prod_{i=1}^m DE(d_{ij}^s - X_i^s \mathbf{B}_j - u_j^s, 1/\nu_j^E)$, $\mathbf{d}_j^s = (d_{1j}^s, \dots, d_{mj}^s)^T$ and Θ_j contains all parameters except u_j^s . The integration in 4 is then approximated by

$$f(\mathbf{d}_j^s | \Theta_j) \approx 1/L \sum_{l=1}^L \left(\prod_{i=1}^m DE(d_{ij}^s - X_i^s \mathbf{B}_j - u_j^{s,l}, 1/\nu_j^E) \right), \quad u_j^{s,l} \sim DE(0, 1/\nu_j^U).$$

In the Rfmmc case, components of \mathbf{E}_j^s are correlated with a block-wise correlation structure determined by ρ_j . Denote by \mathbf{R}_{ρ_j} the correlation matrix of each block, then

$$f(\mathbf{d}_j^s | \Theta_j) = \int \prod_b f(\mathbf{d}_{bj}^s | \mathbf{B}_j, u_j^s, \nu_j^E, \rho_j) f(u_j^s | \nu_j^U) du_j^s, \quad (5)$$

where b is the index for blocks and

$$f(\mathbf{d}_{bj}^s | \mathbf{B}_j, u_j^s, \nu_j^E, \rho_j) = |2\pi \mathbf{R}(\rho_j)|^{-1/2} (\nu_j^E)^{n_b} \left(\nu_j^E \sqrt{q(\mathbf{d}_{bj}^s)} \right)^{1-n_b/2} K_{1-n_b/2}(\nu_j^E \sqrt{q(\mathbf{d}_{bj}^s)}),$$

where $q(\mathbf{d}_{bj}^s) = (\mathbf{d}_{bj}^s - X_b^s \mathbf{B}_j - u_j^s \mathbf{1})^T \mathbf{R}_{\rho_j}^{-1} (\mathbf{d}_{bj}^s - X_b^s \mathbf{B}_j - u_j^s \mathbf{1})$ and $K_\tau(\cdot)$ is the modified Bessel function of the second kind and order τ . The above density function is derived from the multivariate scale mixture of Gaussian employing the density function of GIG in (6) in the main text. The likelihood $f(\mathbf{d}_j^s | \Theta_j)$ in (5) can then be approximated using Monte Carlo integration.

5. APPENDIX E: FURTHER DETAILS OF BFDR ANALYSIS

Let $C^{(g)}(s_1, s_2, t)$, $g = 1, \dots, N$ be N posterior samples of a contrast effect (or any fixed effect of interest). To flag out regions that are significantly greater than a threshold δ by controlling the Bayesian false discovery rate (FDR) to be less than α , we perform the following steps:

1. Estimate the probability discovery function $p_\delta(s_1, s_2, t) = \text{pr}\{|B(s_1, s_2, t)| > \delta \mid \text{Data}\}$ by $\hat{p}_\delta(s_1, s_2, t) = N^{-1} \sum_{g=1}^N I\{|C^{(g)}(s_1, s_2, t)| > \delta\}$ for each grid point (s_1, s_2, t) .
2. Sort $\{\hat{p}_\delta(s_1, s_2, t)\}$ in descending order to obtain $\{\hat{p}_{\delta,(l)}, l = 1, \dots, M\}$, where M is the total number of grid points.

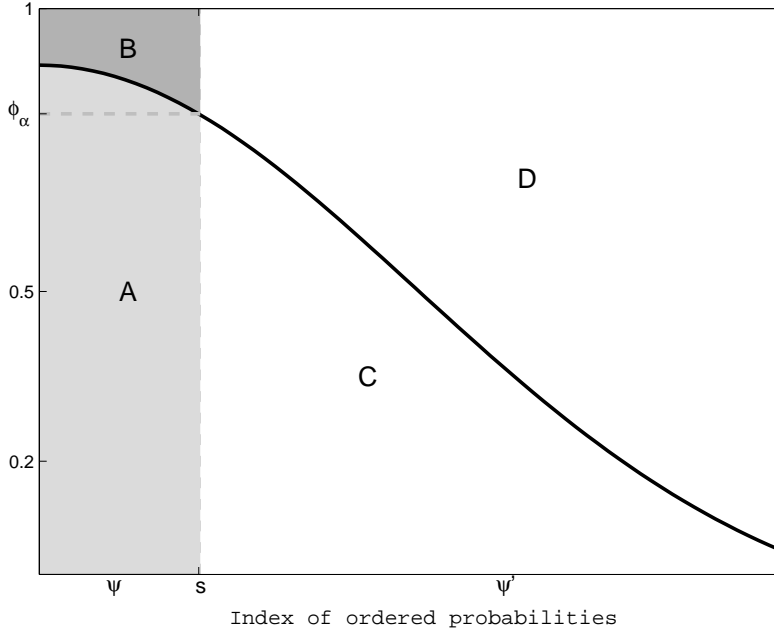


Figure 1: A diagram for Bayesian-FDR based inference.

3. Set $\phi_\alpha = \hat{p}_{\delta,(s)}$, where $s = \max\{l^* : (l^*)^{-1} \sum_{l=1}^{l^*} \{1 - \hat{p}_{\delta,(l)}\} \leq \alpha\}$.

The set of locations $\psi = \{l : \hat{p}_{\delta,(l)} > \phi_\alpha\}$ is the set of *discoveries*. The threshold ϕ_α is a cut-point on the posterior probabilities that corresponds to an expected Bayesian FDR of α , which means that on average $\geq 100(1 - \alpha)\%$ locations of the set ψ should have at least δ difference. That is, if $\mathbb{N}(\psi)$ is the cardinality of the set ψ , defined as $\mathbb{N}(\psi) = \sum_{l=1}^M I(l \in \psi)$, then $\mathbb{N}(\psi)^{-1} \sum_{(s_1, s_2, t) \in \psi} \text{pr}\{|C(s_1, s_2, t)| \leq \delta | \text{Data}\} \leq \alpha$. The above method of determining the Bayesian FDR-based threshold ϕ_α is sketched in a diagram in Figure 1, where the solid decreasing line denotes the ordered $\hat{p}_{\delta}(s_1, s_2, t)$. The areas denoted by A, B, C, D are the estimated proportions for true positives, false positives, false negatives and true negatives, respectively. The threshold ϕ_α is indeed determined by constraining $B/(A + B) \leq \alpha$. One can further calculate the corresponding false negative rate (FNR) by $C/(C + D)$, sensitivity (SEN) by $A/(A + C)$, specificity (SPEC) by $D/(B + D)$.

6. APPENDIX F: SETTING INITIAL VALUES FOR MCMC ALGORITHMS AND HYPER-PRIOR PARAMETERS

The six Bayesian FMM models compared in this paper rely on Markov Chain Monte Carlo algorithms, which requires initial values for model parameters and parameters in hyper-priors to be determined. Initial values of \mathbf{B}_{jk}^* , \mathbf{U}_{jk}^* were calculated using Henderson’s Mixed Model Equations (MME) (Searle et al. 1992, pages 275-286) independently across (j, k) . A MME assumes that:

$$\mathbf{d}_{jk} = \mathbf{X}\mathbf{B}_{jk}^* + \mathbf{Z}\mathbf{U}_{jk} + \mathbf{E}_{jk},$$

with $E\{\mathbf{U}_{jk}\} = 0, Var(\mathbf{U}_{jk}) = q_{jk}^*I_M$, $E\{\mathbf{E}_{jk}\} = 0, Var(\mathbf{E}_{jk}) = s_{jk}^*I_N$. Both \mathbf{U}_{jk} and \mathbf{E}_{jk} are assumed to be Gaussian. The algorithm follows that of Viktor Witkovsky’s algorithm (available at: <http://www.mathworks.com/matlabcentral/fileexchange/200-mixed>). The solution of MME includes $\{\hat{\mathbf{B}}_{jk}^*, \hat{\mathbf{U}}_{jk}, \hat{q}_{jk}^*, \hat{s}_{jk}^*\}$, $Var(\hat{\mathbf{B}}_{jk}^*)$, as well as $\{sd(\hat{q}_{jk}^*), sd(\hat{s}_{jk}^*)\}$, the standard deviation of the variance component estimates.

Set the prior parameter Γ for \mathbf{B}^* in `Gfmmc` and `Rfmmc`. As a prior parameter, the prior correlation matrix \mathbf{R}_B (hence Γ) of \mathbf{B}_{jk}^* can be specified using parametric forms. In our ERP analysis, we set $\mathbf{R}_B = I_A \otimes \mathbf{R}_0$, the Kronecker product of the identity matrix I_A of size A and a $S \times S$ spatial correlation matrix \mathbf{R}_0 . We determine \mathbf{R}_0 by assuming a parametric continuous-time, first-order autoregressive correlation structure, called AR(1) structure, so that $(\mathbf{R}_0)_{ss'} = \omega^{d_{ss'}}$ with $\omega \in (0, 1)$ and $d_{ss'} > 0$ (Louis 1988; Simpson et al. 2014). Here $d_{ss'}$ is the distance between electrodes s and s' on the scalp surface. Larger values of ω induce higher correlation between nearby electrodes a priori. We note that other parametric structures can be adapted in our framework without difficulty. The values of ω in AR(1) can be updated using a Metropolis-Hastings sampling step in the MCMC. In our analysis, we adopt a computationally convenient strategy by prespecifying ω . In particular, ω is estimated through a least square fit to the sample correlations of components in $\{\hat{\mathbf{B}}_{jk}^*\}$ at different pairs of the electrodes. Our simulation analysis shows that this approach provides a reasonable estimation of ω compared to the underlying truth, and our sensitivity analysis for the real data has demonstrated that the posterior results are not sensitive to different prior parameters of ω .

The rest of the parameters vary according to different models. We describe details as follows:

1. **Gfmmc:** The MME method provides initial estimates $\{\hat{q}_{jk}^*, \hat{s}_{jk}^*\}$ and their standard deviations. The initial values of $\{\tau_{ajk}\}$ and $\{\pi_{aj}\}$ are determined following the empirical Bayes method described in Section 4.4 of Morris and Carroll (2006). We have also assumed hyper-priors $\tau_{ajk} \sim \text{Inverse-Gamma}(a^\tau, b^\tau)$, $\pi_{aj} \sim \text{Beta}(a^\pi, b^\pi)$, and the parameters of these hyper-prior distributions are determined by constraining the

mode of the resulting hyper-prior distributions to be the initial values (or mean of them averaged across all indices) and the variance to be a pre-specified large value.

2. **Rfmm:** The Rfmm model contains “scaling parameters” $\{\lambda_{ijk}, \phi_{mjk}, \psi_{ajk}\}$ as well as the “lasso parameters” $\{\nu_{jk}^E, \nu_{jk}^U, \nu_{aj}^B\}$. By choosing $g_1(\cdot)$ to be exponential distribution, we get $E_{ijk}^* \sim DE(0, 1/\nu_{jk}^E)$, $U_{mjk}^* \sim DE(1, 1/\nu_{jk}^U)$, $B_{ajk}|\gamma_{ajk} = 1 \sim DE(0, 1/\nu_{aj}^B)$. The initial values of the lasso parameters were then determined by letting the MME estimates \widehat{s}_{jk}^* , \widehat{q}_{jk}^* , $sd(\widehat{B}_{ajk}^*)$ equal to the variances of these DE distributions. The initial values for the scaling parameters are then obtained by sampling from the conditional distributions of these parameters conditional on other initial values. The initial values for $\{\pi_{aj}\}$ and parameters of the Beta prior of π_{aj} were determined similarly as in the Gfmm case. Since we chose $g_2(\cdot)$ to be the Gamma distribution, the parameters of these hyper-priors were determined by letting the mode of the distributions equal the initial value and letting the variance to be a pre-specified large value.

3. **Gfmmc and Rfmmc:** The initial values for $\{\psi_{cjk}\}$ in the CNEG prior were determined as follows. Since in the CNEG prior, $b_{cjk}^* \sim N(0, \psi_{cjk})$, $\psi_{cjk} \sim Exp(\gamma_{cjk})$ implies that $b_{cjk}^* \sim DE(0, 1/\sqrt{2\gamma_{cjk}})$, we see that $Var(b_{cjk}^*) = 1/\gamma_{cjk}$. Therefore we set initial values of $\{\gamma_{cjk}\}$ equal $1/Var(\widehat{b}_{cjk})$, where $\{Var(\widehat{b}_{cjk})\}$ were estimated from the MME method. The initial values of $\{\psi_{cjk}\}$ were obtained by sampling from the conditional distributions of ψ_{cjk} , conditional on the initial values of b_{cjk}^* , γ_{cjk} .

The (a_{jk}^B, b_{jk}^B) parameters in the CNEG prior were determined by letting the mode of the Gamma prior (Equation (5) in the main text) to be $\bar{\gamma}_{jk}$, the averaged initial values γ_{cjk} across index c , and letting the variance to the Gamma prior to be a pre-specified value.

In both Gfmmc and Rfmmc, the correlation parameter $\boldsymbol{\rho} = (\alpha, v)$ were initialized using pre-specified appropriate values and the upper bound for the uniform hyper-prior C_α, C_v were determined to allow positive definite Matérn covariance structures. The stepsize for the proposal distribution in the Metropolis-Hastings step of updating $\boldsymbol{\rho}$ (or $\boldsymbol{\rho}_{jk}$) were tuned so that the acceptance rate is between 20% – 45%.

All other parameters in Gfmmc and Rfmmc are initialized following the similar methods in the Gfmm and Rfmm models.

7. APPENDIX G: PLOTS OF SIMULATED ERPS

Our simulation was based on a reference ERP dataset—the occipital region (R11) of the ERP data. In Figure 2, we plotted simulated ERPs (blue curves) together with reference ERPs (gray curves) for channel 69 and cigarette-related stimulus and for 10 selected individuals. The simulated datasets were denoted as $G_{\boldsymbol{\rho}_{jk}}$, $G_{\boldsymbol{\rho}}$, $DE_{\boldsymbol{\rho}_{jk}}$, and $DE_{\boldsymbol{\rho}_{jk}}$, where G

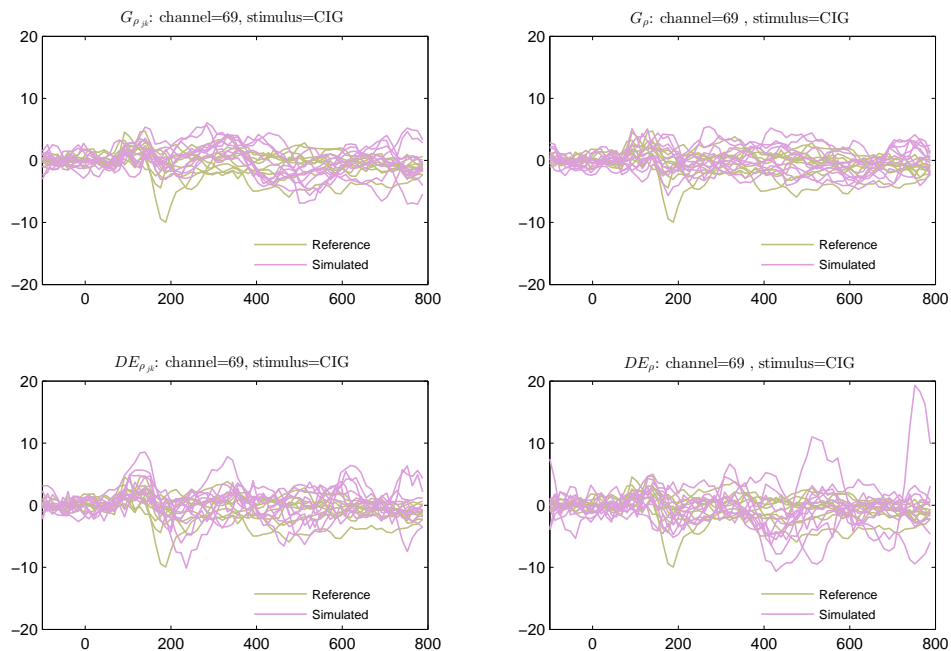


Figure 2: *Plots of simulated data.* Plots of simulated ERPs together with reference ERPs for channel 69, cigarette-related stimulus, and 10 selected individuals.

indicates data simulated with Gaussian random effects and residuals, DE indicates data simulated with DE random effects and residuals, and the subscripts ρ_{jk} or ρ indicate whether the true correlation parameters vary across index (j, k) .

8. APPENDIX H: MORE SIMULATION RESULTS

Besides FDR_{ϵ} and Sen_{ξ} defined in the main text, we also define false negative rate (FNR_{ξ}) and specificity ($SPEC_{\epsilon}$) to assess the performance of BFDR and SimBaS. The FNR_{ξ} is defined as the number of non-flagged locations with true value greater than or equal to ξ divided by the total number of non-flagged locations; the $SPEC_{\epsilon}$ is defined as the number of non-flagged locations with true value less than ϵ divided by the total number of locations with true value less than ϵ . The four summary statistics (IMSE, IPVar, IWidth, CprB₉₅ defined in the main text) for the random effects $U(t)$, the $FNR_{1.25}$ and $SPEC_{.3}$ for both the BFDR ($\delta = 0.6$) and the SimBaS approach, are averaged across all five repeated simulation runs. Results are listed in Table 1.

From Table 1, we observe that the statistics on $U(t)$ show similar patterns as that of $B(t)$ demonstrated in the main text. In particular, Gfmm and Rfmm show larger IMSE,

Table 1: Extra summary statistics of simulation study: integrated mean squared error (IMSE), integrated posterior variance (IPVar), integrated width of 95% credible interval (IWidth) and coverage probability of the 95% SCB (CPrB₉₅) of $\mathbf{U}(t)$; and the FNR_ϵ and $SPEC_\epsilon$ calculated for regions flagged using BFDR ($\delta = 0.6$) and SimBaS approaches.

Data	Model	$\mathbf{U}(t)$				BFDR ($\delta = 0.6$)		SimBaS	
		IMSE	IPVar	IWidth	CPrB ₉₅	FNR _{1,25}	SPEC _{.3}	FNR _{1,25}	SPEC _{.3}
$G_{\rho_{jk}}$	Gfmm	0.417	0.027	0.422	0.592	0.002	0.997	0.005	0.999
	Rfmm	0.466	0.027	0.422	0.567	0.003	0.992	0.006	0.995
	Gfmmcρ_{jk}	0.209	0.186	2.857	0.997	0.001	0.999	0.004	0.998
	Gfmmc ρ	0.219	0.288	4.423	0.999	0.002	0.999	0.008	0.998
	Rfmmc ρ_{jk}	0.233	0.190	2.936	0.997	0.002	0.999	0.005	0.999
	Rfmmc ρ	0.271	0.304	4.722	0.999	0.003	1.000	0.012	0.999
G_ρ	Gfmm	0.735	0.046	0.705	0.569	0.004	0.996	0.008	0.999
	Rfmm	0.793	0.046	0.715	0.560	0.004	0.987	0.008	0.993
	Gfmmc ρ_{jk}	0.336	0.338	5.201	0.998	0.002	0.998	0.006	0.998
	Gfmmcρ	0.336	0.331	5.086	0.998	0.002	0.997	0.006	0.998
	Rfmmc ρ_{jk}	0.362	0.328	5.088	0.997	0.003	0.997	0.007	0.998
	Rfmmc ρ	0.360	0.355	5.524	0.998	0.004	0.998	0.008	0.999
$DE_{\rho_{jk}}$	Gfmm	0.481	0.033	0.500	0.609	0.000	0.998	0.001	1.000
	Rfmm	0.383	0.023	0.354	0.577	0.000	0.999	0.000	0.999
	Gfmmc ρ_{jk}	0.268	0.256	3.939	0.997	0.000	0.999	0.001	1.000
	Gfmmc ρ	0.315	0.422	6.513	0.995	0.002	1.000	0.008	1.000
	Rfmmcρ_{jk}	0.147	0.135	2.084	0.999	0.000	0.999	0.000	0.999
	Rfmmc ρ	0.187	0.258	4.056	1.000	0.000	1.000	0.000	1.000
DE_ρ	Gfmm	1.457	0.088	1.348	0.577	0.001	0.992	0.003	1.000
	Rfmm	1.119	0.065	0.991	0.559	0.000	0.997	0.001	1.000
	Gfmmc ρ_{jk}	0.508	0.622	9.589	0.998	0.001	0.997	0.002	1.000
	Gfmmc ρ	0.510	0.622	9.589	0.998	0.001	0.996	0.002	1.000
	Rfmmc ρ_{jk}	0.292	0.309	4.829	0.999	0.000	0.999	0.000	0.999
	Rfmmcρ	0.288	0.323	5.078	0.999	0.000	0.999	0.000	1.000

True vs. Estimated Contrast Effect: Cigarette - Neutral

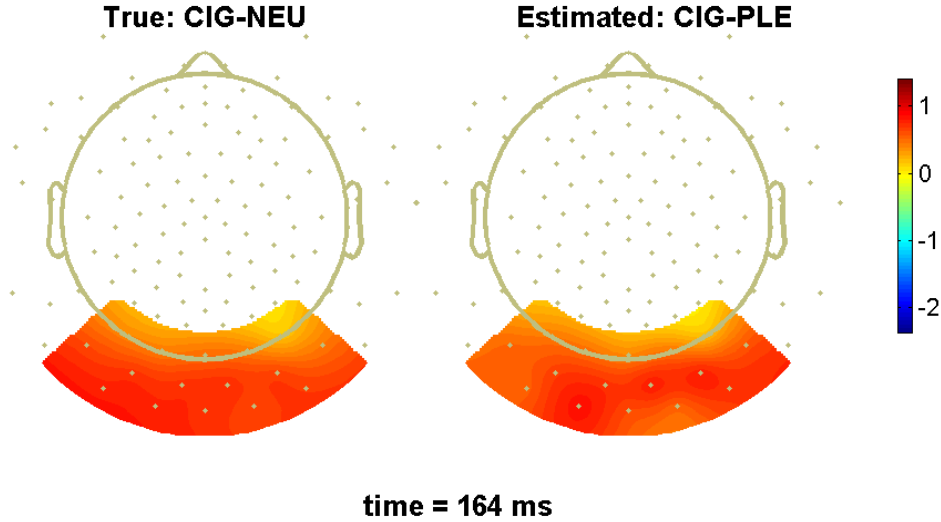


Figure 3: *Scalp plots for simulation.* True vs. estimated contrast effect at a fixed time point $t = 164$ ms. The contrast effect shown is the contrast between cigarette and neutral stimuli.

lower coverage rate, smaller IPVar and narrower IWidth than Gfmmc and Rfmmc models. These pattern implies that ignoring correlation at the lowest level of hierarchy leads to higher estimation error and underestimated posterior variance. Comparing the four FMMC models that take into account correlations, we find that for data with Gaussian tails, Rfmmc models still achieve results comparable to the Gfmmc models. For data with DE tails, the Rfmmc models perform systematically better than Gfmmc on all four statistics. In terms of BFDR and SimBaS statistics, we found that all methods tend to have low false negative rate and high specificity, which implies that most non-significant regions are correctly identified (not flagged) with high probability.

As an intuitive visualization of the ground truth and estimated effects, in Figure 3, we show scalp plots for true vs. estimated contrast effect at a fixed time point $t = 164$ ms based on one simulation run that applies the model $Gfmmc_{\rho_{jk}}$ on the simulated data $G_{\rho_{jk}}$. Additionally, we have made a movie file that shows how true vs. estimated contrast effect change over time. The file name is `simu_scalpplot_TruevsEstimated.avi`, which can be downloaded from <https://drive.google.com/open?id=1MEKcENBWA11okbvNFafC24kyeFAAQysf>.

Table 2: More ERP data analysis results: the run-time (in hours) for 2000 MCMC iterations for all six models and 11 cortical regions.

Regions	Gfmm	Rfmm	Gfmmc ρ_{jk}	Gfmmc ρ	Rfmmc ρ_{jk}	Rfmmc ρ
R1	14.86	20.84	21.43	17.48	25.57	20.42
R2	15.91	19.03	18.48	14.00	32.24	27.76
R3	15.22	19.11	18.37	14.22	31.96	26.98
R4	19.83	20.30	22.11	18.42	34.17	28.15
R5	13.06	21.02	22.87	19.33	35.07	28.89
R6	13.68	20.37	19.68	16.25	36.43	29.54
R7	15.81	17.08	18.32	15.00	33.06	27.36
R8	18.59	17.91	21.04	17.34	36.39	29.98
R9	19.40	20.28	22.12	18.44	38.86	31.76
R10	14.52	20.54	23.10	19.07	33.80	29.25
R11	22.82	23.85	25.52	21.38	37.72	31.73

9. APPENDIX I: RUN-TIME ANALYSIS FOR ERP APPLICATION

Run-time for ERP Training Data: the run-time for training the six models, including the two FMM models (Gfmm, Rfmm) and the four FMMc models (Gfmmc ρ_{jk} , Gfmmc ρ , Rfmmc ρ_{jk} , Rfmmc ρ), for each of the 11 regions is listed in Table 2. The run-time is calculated based on 2000 MCMC iterations using a computational server with 2 Intel Xeon E5-2609 4C/4T 2.4Ghz 6.4GT/s 10mb 80w CPUs and 98.3 GB 1333 MHz Registered ECC DDR3 Memory. The results are based on the training data with 140 subjects. From Table 2, we see that all robust models take longer time than their Gaussian counterparts, and FMMc models usually take longer time to run than the FMM models. Furthermore, the separable models (Gfmmc ρ_{jk} , Rfmmc ρ_{jk}) are slower than their non-separable counterparts (Gfmmc ρ , Rfmmc ρ). Overall, if running all regions concurrently, the whole data set takes at most 39 hours for 2000 MCMC iterations.

A Run-Time Analysis: To further demonstrate how the proposed framework scales with various data setups, we performed a run-time analysis to evaluate the run-time of the six models while varying the number of sampling points per ERP, the number of electrodes, and the number of stimuli. Data used in this analysis are based on the training ERP data (containing 140 subjects) in the scalp region 11. We have plotted the run-time of all six models by varying the number of electrodes in (6, 10, 14, 18), the number of sam-

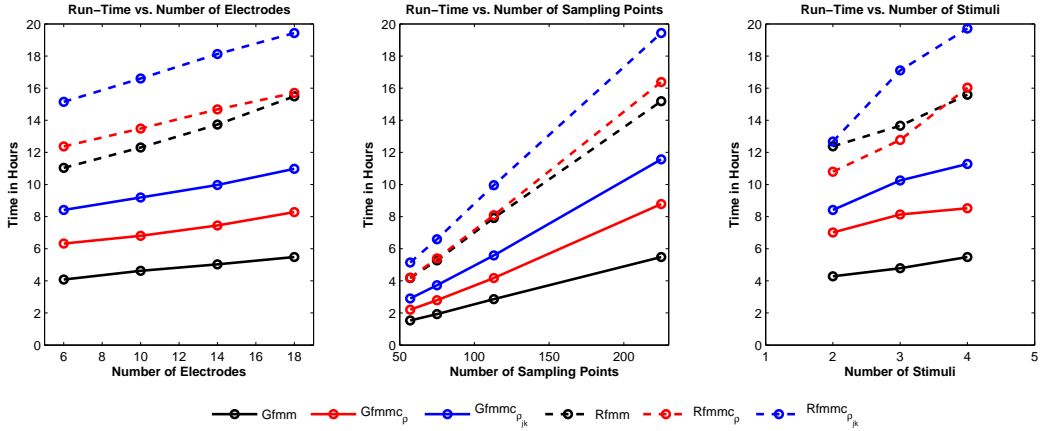


Figure 4: *Run-time analysis.* An analysis to evaluate the run time of the six models under different number of electrodes (the left figure), different number sampling points (the middle figure), and different number of stimuli (the right figure).

pling points in (57, 75, 113, 225), and the number of stimuli in (2, 3, 4). All run-time was measured in hours based on 2000 MCMC iterations. Results are demonstrated in Figure 4. Figure 4 demonstrates a rough linear increase in run-time for all six models as the data size/dimension increases. In general, the robust versions (Rfmm, Rfmmc ρ , Rfmmc ρ_{jk} ; marked with dashed lines) require longer run-time than the Gaussian versions (Gfmm, Gfmmc ρ , Gfmmc ρ_{jk} ; marked with solid lines). Furthermore, models with non-separable spatial correlations (Gfmmc ρ_{jk} , Rfmmc ρ_{jk} ; marked using blue color) require longer run-time than models with separable spatial correlations (Gfmmc ρ , Rfmmc ρ ; marked using red color), and both types of models run slower than models without spatial correlation (Gfmm, Rfmm; marked using black color).

In practice, we recommend users to estimate run-time by performing a test run using a few MCMC iterations. The overall time-cost can be approximated by multiplying the time-cost per iteration with the total number of MCMC iterations.

10. APPENDIX J: MORE ERP DATA ANALYSIS RESULTS: MOVIE FILES THAT SHOW FLAGGED REGIONS OVER TIME USING SIMBAS AND BFDR

We have made two movie files that show the flagged regions over time using either the SimBaS or the BFDR method. The name of the movie files are:

Movie 1: SimuCB_flag_bestModels.avi (results using SimBaS).

Movie 2: BFDR_flag_bestModels.avi (results using BFDR).

These files can be downloaded from the link:

<https://drive.google.com/a/vt.edu/file/d/0B-GWoYrWkA2dcmxsNTNDZWVnUGs/view?usp=sharing>

11. APPENDIX K: MORE ERP DATA ANALYSIS RESULTS: SUMMARY TABLE
AND PLOTS OF TIME INTERVALS WITH EVIDENT PATTERNS
TOGETHER WITH A DETAILED DESCRIPTION

In Table 3, we highlight six time intervals with evident patterns resulting from examining the integrated 2D-heatmaps and the .avi files. The summary plots (on the flagging regions using SimBaS and BFDR) at each of these time intervals are shown in Figures 5-10. Each plot includes the integrated 2D heatmaps of mean contrast effects (color maps) marked with flagged regions using SimBaS (row 1) and BFDR ($\delta=0.5$) (row 3), as well as scalp plots of new SimBaS values calculated at the time interval (row 3) using posterior samples averaged across the intervals, and scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions (row 4), where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval. In the integrated 2D heatmap plots, the time interval considered was marked using green vertical lines.

From Table 3, the .avi files, as well as the summary plots shown in Figures 5-10, we see how the spatial distribution of stimulus differences evolves and changes over time. Re-assuringly, no (s, t) were flagged for any time points before the image stimulus was shown ($[-100, 0]$ ms) and during the interval $[0, 100]$ ms. Between 112 ms and 160 ms, a time period known as the P1 region, we see a cigarette differential effect, whereby CIG was significantly different from NEU, PLE, and UNP in the parietal-occipital (R9-R11) region. These effects are clearly visible in both SimBaS and BFDR ($\delta=0.5$) results. From roughly 216 ms to 660 ms, we see similarities between the responses to the cigarette stimulus and the two emotional stimuli, as shown in Figures 6-9.

To be more specific, initially from 216 ms to 232 ms, we observe similar response patterns between cigarette and pleasant stimuli, with relative potentials both significantly greater than those for the neutral stimulus in the frontal right and central right regions (R4, R6), and significantly less than those for the neutral stimulus in the parietal (R9-10) region and in a part of the occipital (R11) region. Compared to the pleasant stimulus, the response to the cigarette stimulus differs more from the response to the neutral stimulus in the anterior frontal right (R2) and the occipital (R11) regions, but overall there is a high similarity between the response to the cigarette and pleasant stimuli.

Later, at 232 – 300 ms, the response to the cigarette stimulus shows more similarity with that for the pleasant stimulus than that evoked by the unpleasant stimulus. In particular, all three (cigarette, pleasant, and unpleasant) stimuli evoke relative potentials that are significantly greater than those evoked by the neutral stimulus in the anterior frontal right

Table 3: *ERP data analysis*: time intervals with pronounced flagging patterns, summarized based on both SimBaS and BFDR ($\delta = 0.5$) outputs.

Time(ms)	Effects & Primary Regions	Comments
[112, 160]	CIG-NEU,PLE,UNP(R9-11).	Cigarette differential effect. CIG is significantly different from NEU, PLE, and UNP in the parietal-occipital region.
[216, 232]	CIG,PLE-NEU(R4,6,9-11).	Similarity between cigarette and pleasant. Both CIG and PLE are different from NEU in frontal-central right, parietal, and partial occipital regions, with UNP-NEU showing different effects.
[232, 300]	CIG,PLE-NEU(R2,4,6,9-11); UNP-NEU(R1,2,4,6).	Cigarette more similar with pleasant than unpleasant. CIG,PLE,UNP are different with NEU in anterior-front-central right. CIG-NEU and PLE-NEU are different with UNP-NEU in anterior frontal left and parietal-occipital regions.
[300, 440]	CIG,PLE-NEU(R1-11); UNP-NEU(R1-7,10).	Similarity between cigarette and emotional stimuli. CIG is similar with PLE and UNP in contrast with NEU, with UNP having less flagged in the right-temporal, parietal left and occipital regions.
[440, 660]	CIG,PLE,UNP-NEU(R1,3-11).	High similarity between cigarette and emotional stimuli. CIG show similar effects with PLE and UNP, all having significant difference with neutral in all regions except anterior frontal right.
[660, 800]	CIG-NEU(R1-7);PLE,UNP-NEU(R1-11);CIG- PLE(R1-3,5,7-9,11);CIG- UNP(R2,5);PLE- UNP(R1,R5).	Difference among all stimuli. Significant differences are observed for all pairs of contrast effects.

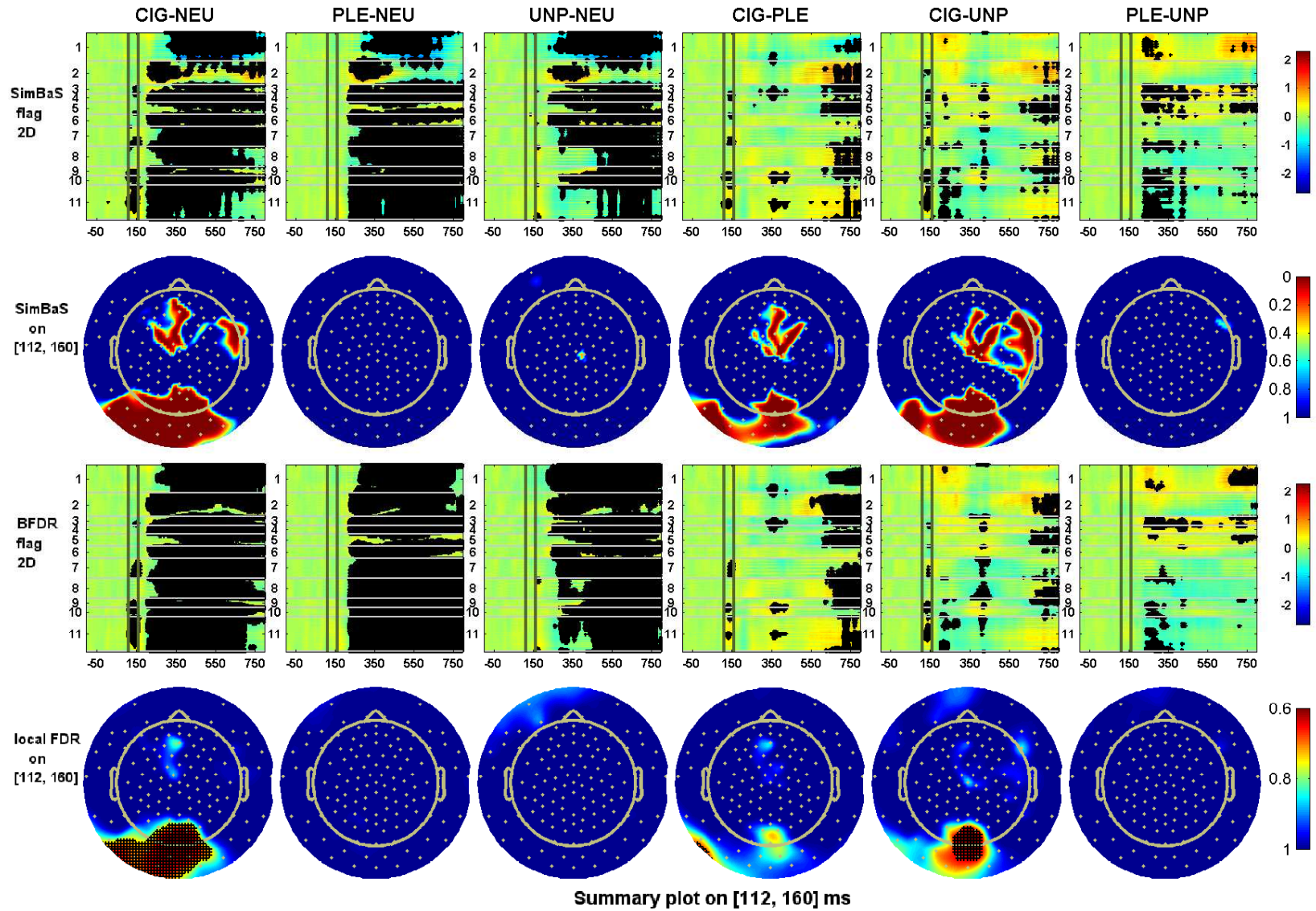


Figure 5: *Summary plots of flagging results at [112, 160] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

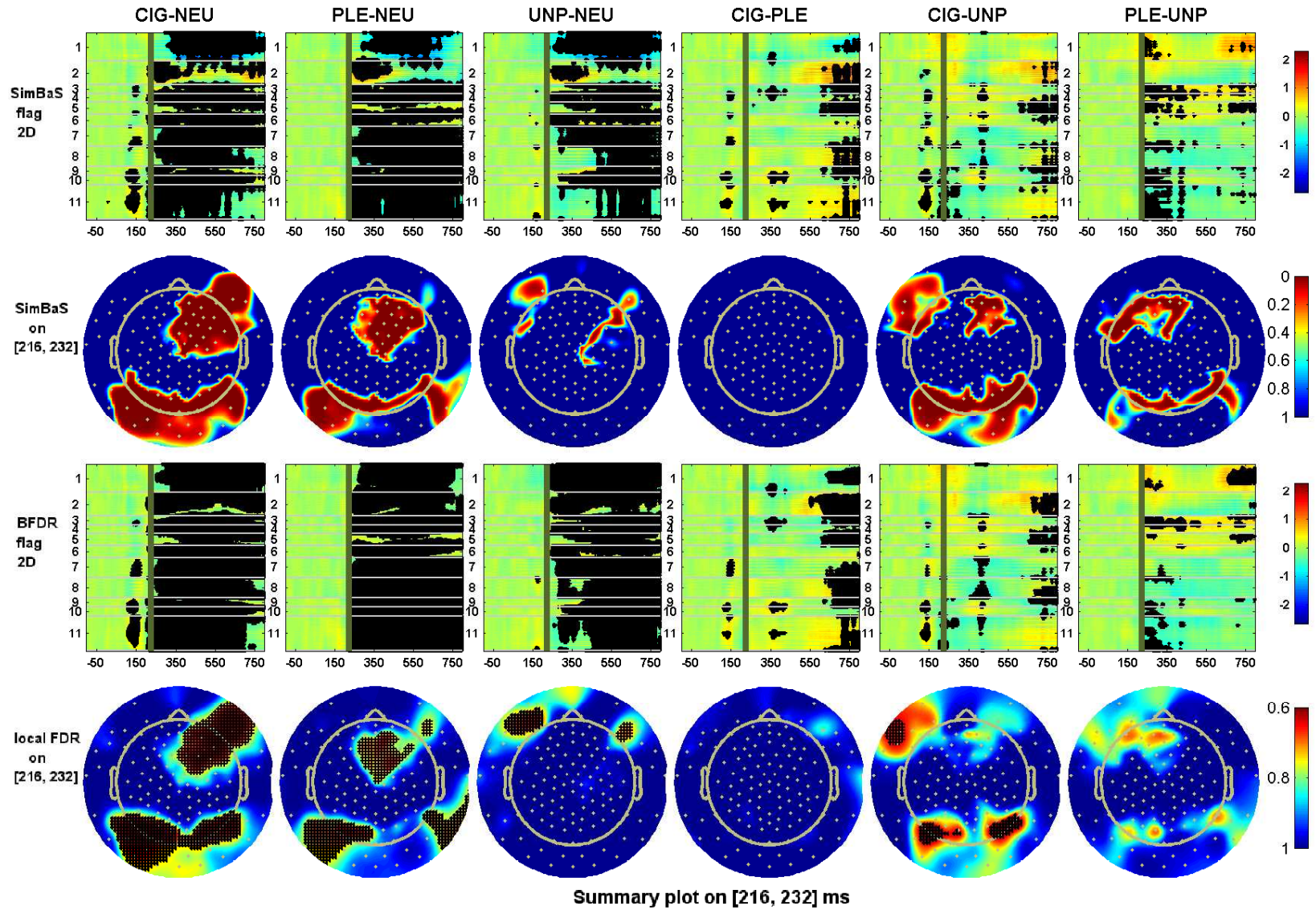


Figure 6: *Summary plots of flagging results at [216, 232] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

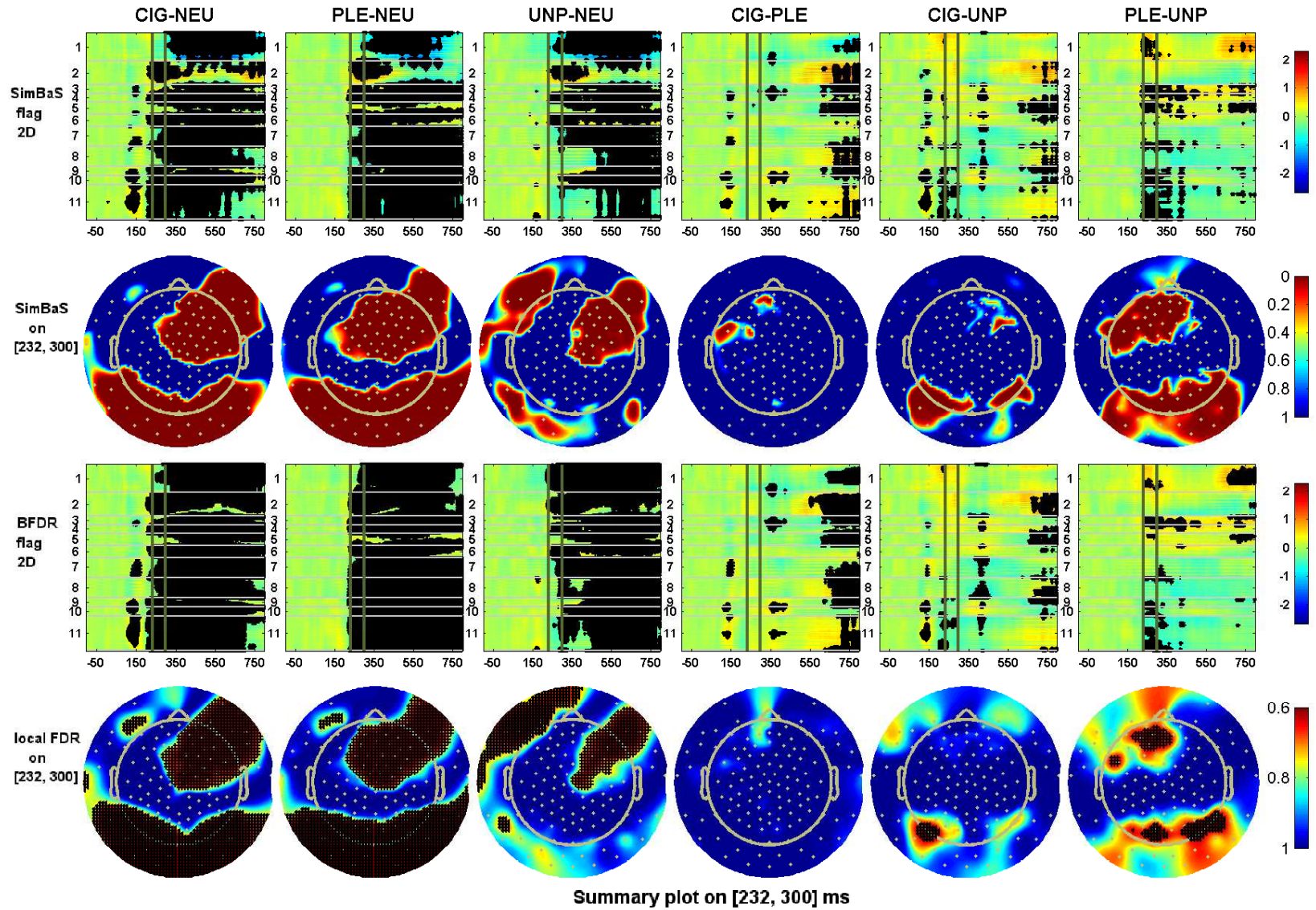


Figure 7: *Summary plots of flagging results at [232, 300] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

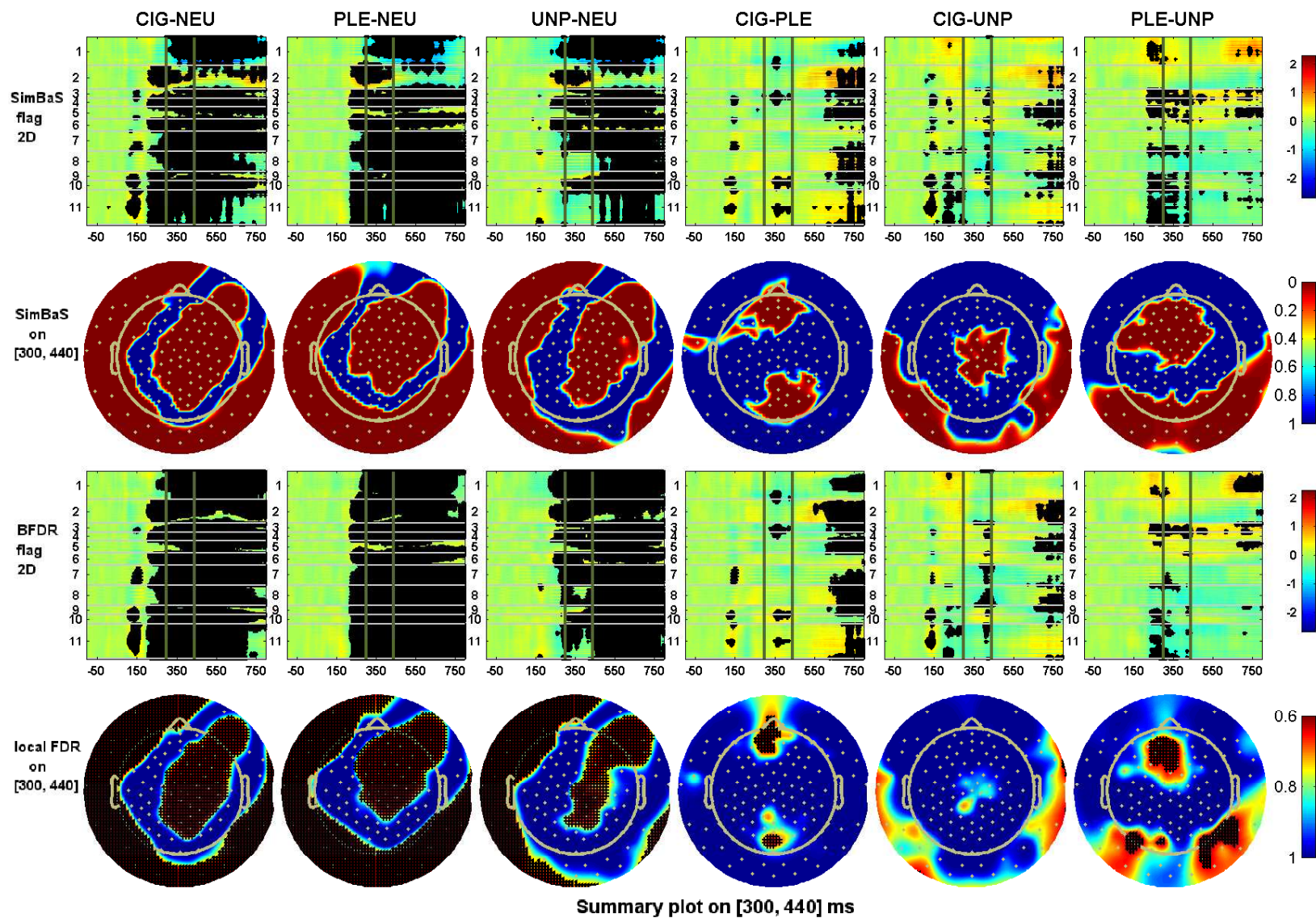


Figure 8: *Summary plots of flagging results at [300, 440] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

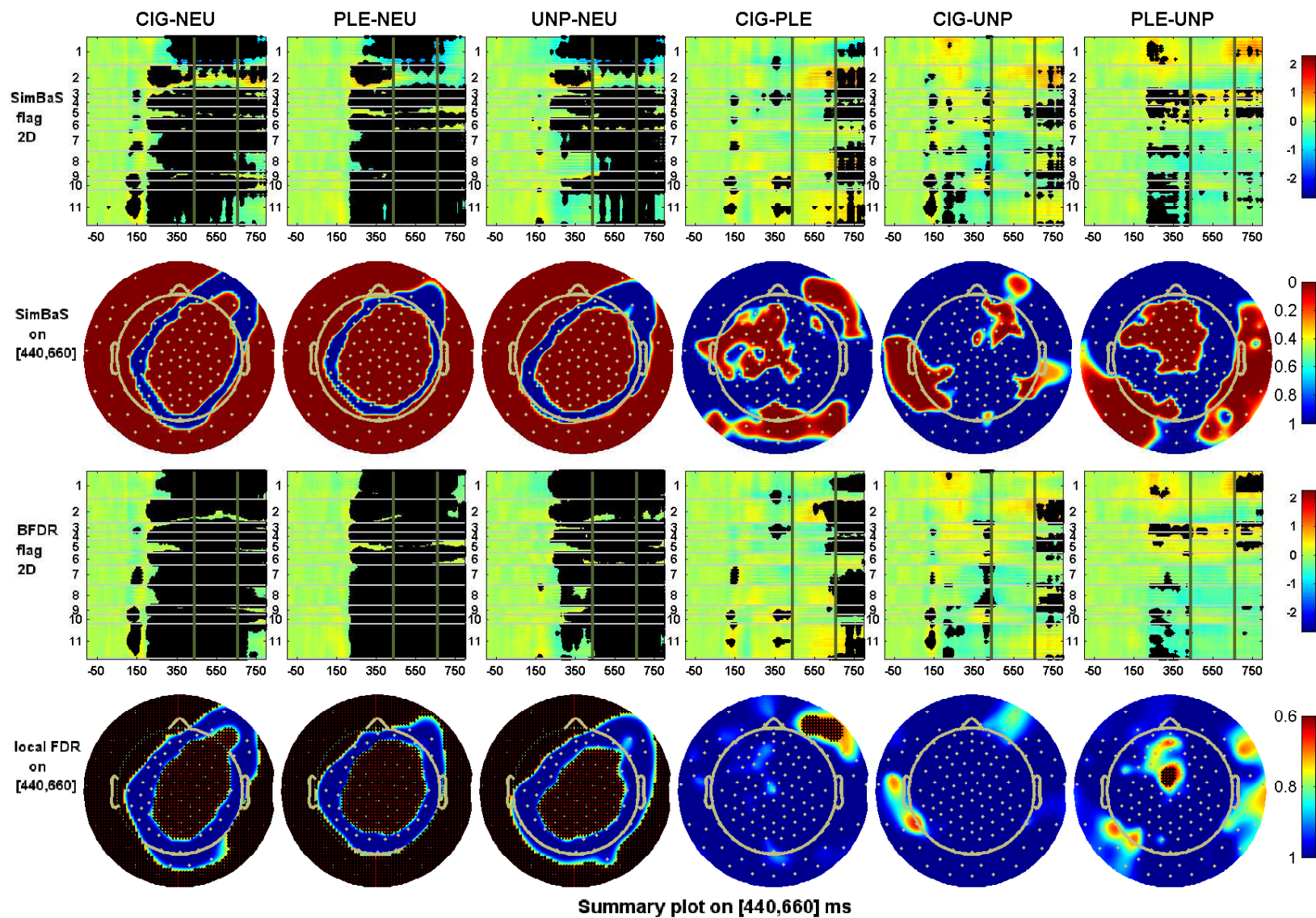


Figure 9: *Summary plots of flagging results at [440, 660] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

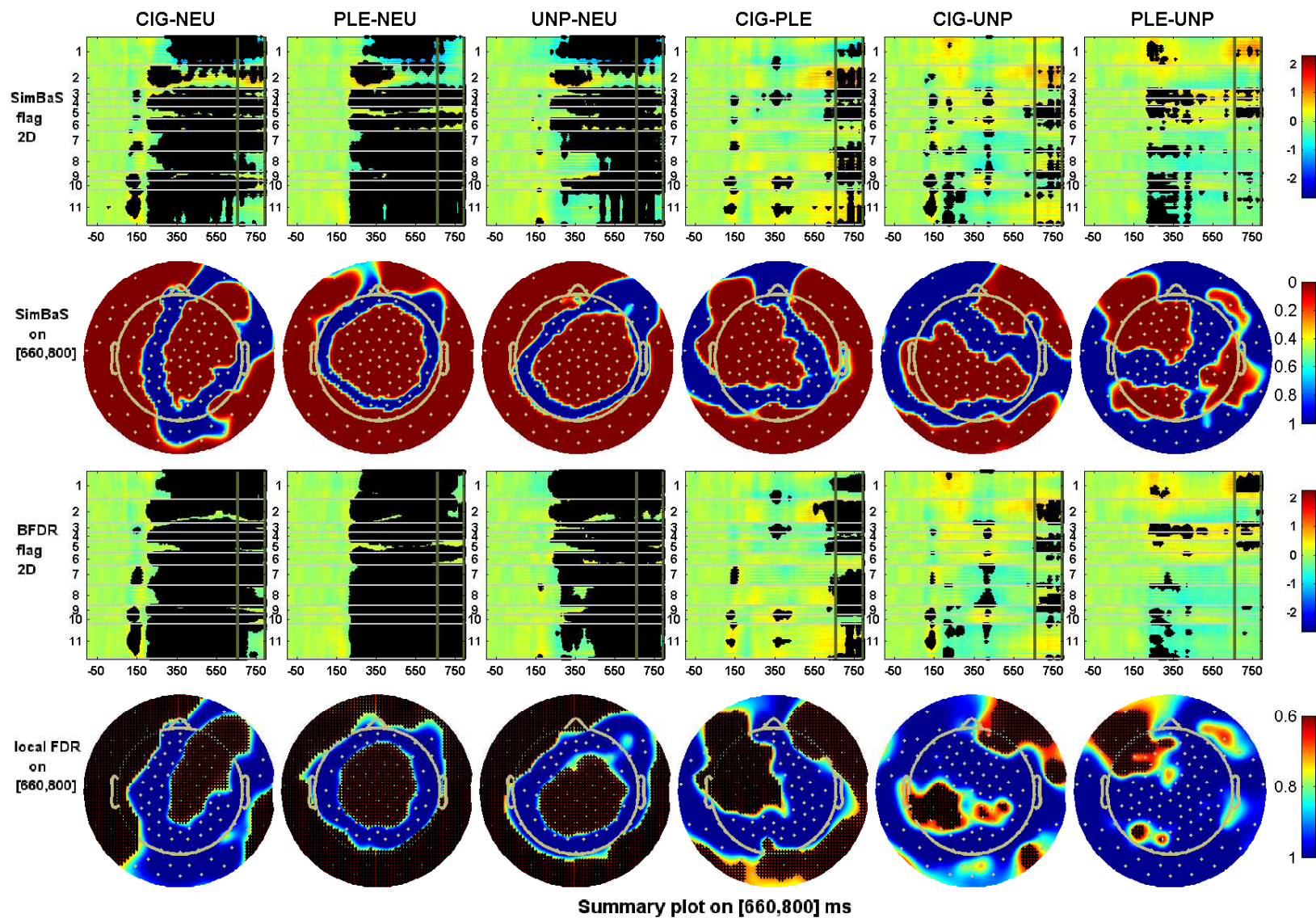


Figure 10: *Summary plots of flagging results at [660, 800] ms.* Row 1: integrated 2D heatmaps of the contrast effects (colormaps) flagged using SimBaS. Black dots mark the flagged regions. The x-axis indicates time and the y-axis indicates vectorized spatial locations of the scalp (indexed by region number). Row 2: scalp plots of new SimBaS calculated with posterior samples of contrast effects averaged at the time interval. Row 3: integrated 2D heatmaps of the contrast effects flagged using BFDR ($\delta = 0.5$). Row 4: the scalp plots of local FDR marked with BFDR ($\delta = 0.5$) flagged regions, where the local FDR and the BFDR flagging results were re-calculated using posterior samples averaged across the interval.

(R2) and the frontal-central right (R4, R6) regions, but only the response to cigarette and pleasant stimuli are significantly less than the response to the neutral stimulus in the parietal-occipital region (R9-11). Additionally, both the BFDR and the SimBaS results show that the unpleasant stimulus evokes relative potentials that are significantly less than those evoked by the neutral stimulus in the anterior frontal left (R1) region, but such a difference is only identified in a tiny area in response to the cigarette and pleasant stimuli.

During the next period (300 – 440 ms), the cigarette stimulus evokes a pattern very similar to those evoked by the pleasant and unpleasant stimuli, in contrast with those resulting from the neutral stimulus, except that fewer areas are flagged for UNP-NEU than for CIG-NEU and PLE-NEU in the temporal right (R8), parietal left (R9) and occipital (R11) regions. From 440ms to 660ms, the cigarette stimulus shows a response pattern very similar to those of both the pleasant and unpleasant stimuli, all detecting differences with the response to the neutral stimulus in the anterior frontal left (R1), and all frontal, central, temporal, and parietal-occipital regions (R3-R11).

Finally, from 660 ms-800 ms, we observe significant differences between the response to all pairs of stimuli. In particular, the cigarette stimulus evokes relative potential different from the pleasant stimulus in the anterior frontal (R1-2), frontal-central left (R3, R5), temporal right (R8), parietal left (R9), and partial occipital regions (R11). The response to the cigarette stimulus is different from that evoked by the unpleasant stimulus in the anterior frontal right (R2) and central left (R5) regions.

In summary, our analysis reveals significant neurological patterns induced by different types of visual stimuli, with an early differential effect evoked by cigarette stimulus followed by different degrees of similarities between the cigarette stimulus and the two emotional stimuli, and followed by significant differences between all pairs of stimuli.

12. APPENDIX L: SENSITIVITY ANALYSIS

The results presented in Section 3.2 in the main text rely on several modeling choices, including modeling ERP curves by 11 scalp regions, determining the prior correlation of \mathbf{B}_{jk}^* based on the initial estimate $\widehat{\mathbf{B}}_{jk}^*$, and selecting models using cross-validated computation. To study the sensitivity of the outputs to these modeling choices, we have repeated several analyses under different modeling choices. These analyses include: (1) **Jittered cortical partitions.** Re-fit $\text{Gfmmc}_{\rho_{jk}}$ to the ERP data based on a new partition—a “jittered” partition of the scalp into 11 regions so that the boundaries are different from the original partition. (2) **The $\omega = 0.3$ spatial prior.** Re-fit $\text{Gfmmc}_{\rho_{jk}}$ by setting the prior correlation for \mathbf{B}_{jk}^* using an AR(1) structure with a fixed parameter $\omega = 0.3$. (3) **The $\omega = 0.8$ spatial prior.** Repeat with the fixed parameter $\omega = 0.8$. (4) **A different cross-validation.** Re-do the model selection step using a different data split, so that the new validation set was sampled from the original training set.

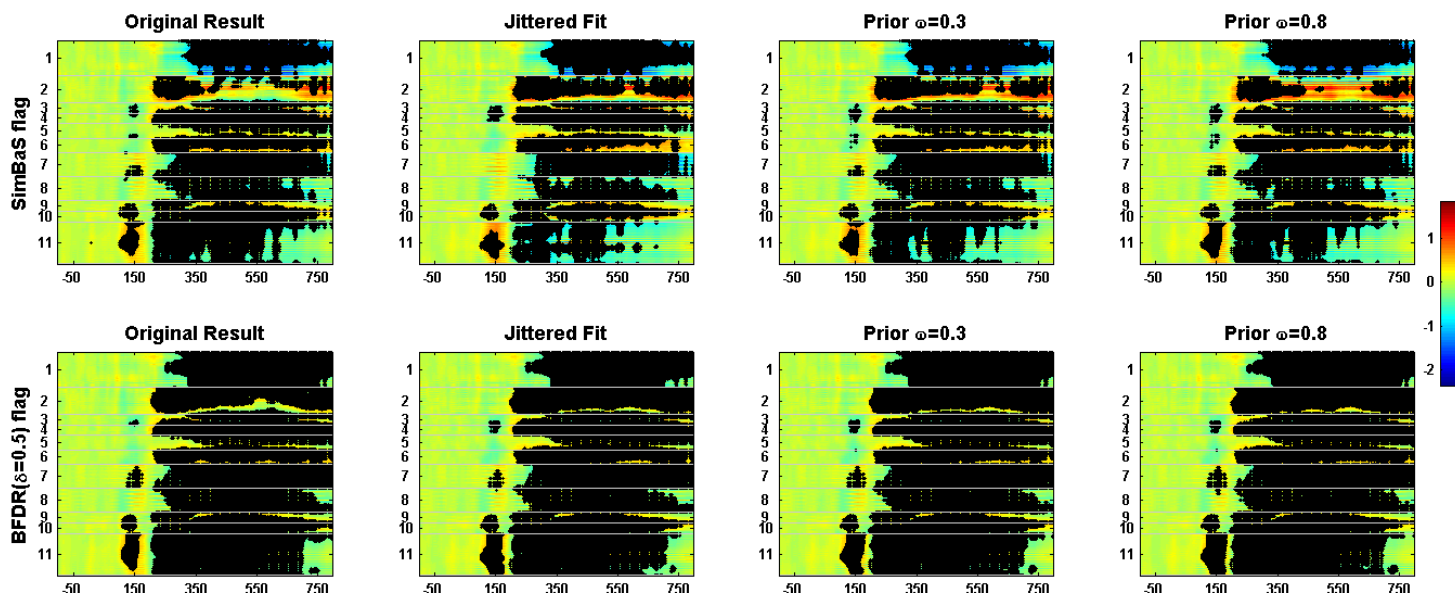


Figure 11: *Sensitivity analysis*. The 2D integrated heatmaps of one contrast effect CIG-NEU with flagged regions using both SimBaS (row 1) and BFDR $\delta = 0.5$ (row 2) approaches, compared between the original fit and three repeated fits with different modeling choices: the jittered fit, the fit with fixed prior parameter $\omega = 0.3$, and the fit with fixed prior parameter $\omega = 0.8$. All results are based on the model $Gfmmc_{\rho_{jk}}$.

Table 4: *Sensitivity analysis*: LPPLs on validation data under a new cross-validation analysis. The values listed are on the scale of 10^4 . The value with the highest LPPL in each region (row) is highlighted with boldface.

Region	Gfmm	Rfmm	Gfmmc ρ_{jk}	Gfmmc ρ	Rfmmc ρ_{jk}	Rfmmc ρ
Ant Frontal L (R1)	-18.902	-95.646	-11.105	-11.388	-10.329	-12.962
Ant Frontal R (R2)	-16.982	-79.657	-11.390	-11.524	-10.158	-12.517
Frontal L (R3)	-1.635	-86.805	6.235	5.946	4.648	3.743
Frontal R (R4)	-7.389	-86.950	-1.041	-1.372	1.596	0.862
Central L (R5)	-6.807	-105.042	0.428	-0.018	12.638	10.523
Central R (R6)	1.701	-105.544	16.185	15.876	14.556	12.344
Temporal L (R7)	-0.699	-58.476	5.644	5.414	2.157	0.736
Temporal R (R8)	1.519	-58.530	7.198	7.036	2.208	0.780
Parietal L (R9)	15.441	-98.899	30.174	29.948	20.270	19.313
Parietal R (R10)	18.663	-114.126	34.933	34.706	26.556	25.071
Occipital (R11)	19.719	-178.844	54.746	54.540	46.254	44.985

Figure 11 contains summaries of the flagged regions for the contrast effect CIG-NEU using both the SimBaS (row 1) and the BFDR (row 2) approach from original model fits and sensitivity analyses (1)-(3) above. From this, we see that the three sensitivity analyses result in similar flagged regions as the original analysis using either the SimBaS or BFDR criteria. Thus, we conclude that our results are robust to the cortical partition boundaries and choice of spatial prior parameters for the \mathbf{B}_{jk}^* .

Outputs of the new cross-validation analysis described in (4) are listed in Table 4. Comparing these results with the original model selection outputs listed in Table 2 in the main text, we found that the LPPL uniformly favored the nonstationary spatial models in all cases, with the Gaussian model favored for some regions but the robust model favored for others. For cortical regions R2, R4, and R5, both training-validation splits favored the robust model, and for R3, R7, R8, and R9 both favored the Gaussian models. However, regions R6, R10, and R11 favored the robust model when the first split was used, and the Gaussian model when the new split was used, and region R1 favored the Gaussian model in the original split but the robust model in the new split. These results could be caused by the random assignment of ERPs with potential outliers into either the training and validation splits, and suggest that the decision of robust vs. Gaussian models by LPPL

may be data-dependent.

13. APPENDIX M: MODEL FITTING USING A GLOBAL MODEL

While our proposed framework is suitable to include all electrodes, we have chosen to fit separate models for each of the 11 cortical regions, and the reasons and benefits have been explained in the main text. To demonstrate the results obtained by fitting a global model to the full data set (including all 129 electrodes), in Figure 12, we plot scalp locations flagged using the BFDR ($\delta=0.5$) approach based on fitting a single model, $Gfmmc_{\rho_{jk}}$ to the full data set. Figure 12 demonstrates similar significance patterns as the results shown in the main text, with less significant regions flagged for contrast effects CIG-NEU and UNP-NEU at region 11.

To further demonstrate the benefits of allowing different models for different regions, we also calculated the model selection statistic—the log posterior predictive likelihood (LPPL) on validation data based on fitting a global model $Gfmmc_{\rho_{jk}}$, which gives 5.7273×10^5 . In contrast, the LPPL obtained using our regional model fitting strategy is 10.627×10^5 , which is 1.86 times higher (in log scale) than the global model fitting approach. This provides strong evidence in support of our modeling strategy.

As we have explained in the main text, one of the main reasons for model fitting by scalp regions is that, the spatial correlation between electrodes tends to vary across scalp regions. To demonstrate this, in Figure 13 we plot the estimated Matérn parameters (posterior means) across the brain regions, including the $\{(\hat{\alpha}_{jk}, \hat{\nu}_{jk})\}$ parameters estimated by $Gfmmc_{\rho_{jk}}$ for the first 10 wavelet components (in gray lines) and the $(\hat{\alpha}, \hat{\nu})$ parameters estimated by $Gfmmc_{\rho}$ (in red lines). Figure 13 shows that the Matérn parameters vary substantially across regions.

REFERENCES

- Griffin, J. E., and Brown, P. J. (2005), Alternative prior distributions for variable selection with very many more variables than observations,, Technical report, University of Warwick, Centre for Research in Statistical Methodology.
- Griffin, J. E., and Brown, P. J. (2012), “Structuring shrinkage: some correlated priors for regression,” Biometrika, 99(2), 481–487.
- Jørgensen, B. (1982), Statistical Properties of the Generalized Inverse Gaussian Distribution, Lecture Notes in Statistics, New York, U.S.A.: Springer-Verlag New York.
- Louis, T. A. (1988), “General methods for analysing repeated measures,” Statistics in Medicine, 7(1-2), 29–45.

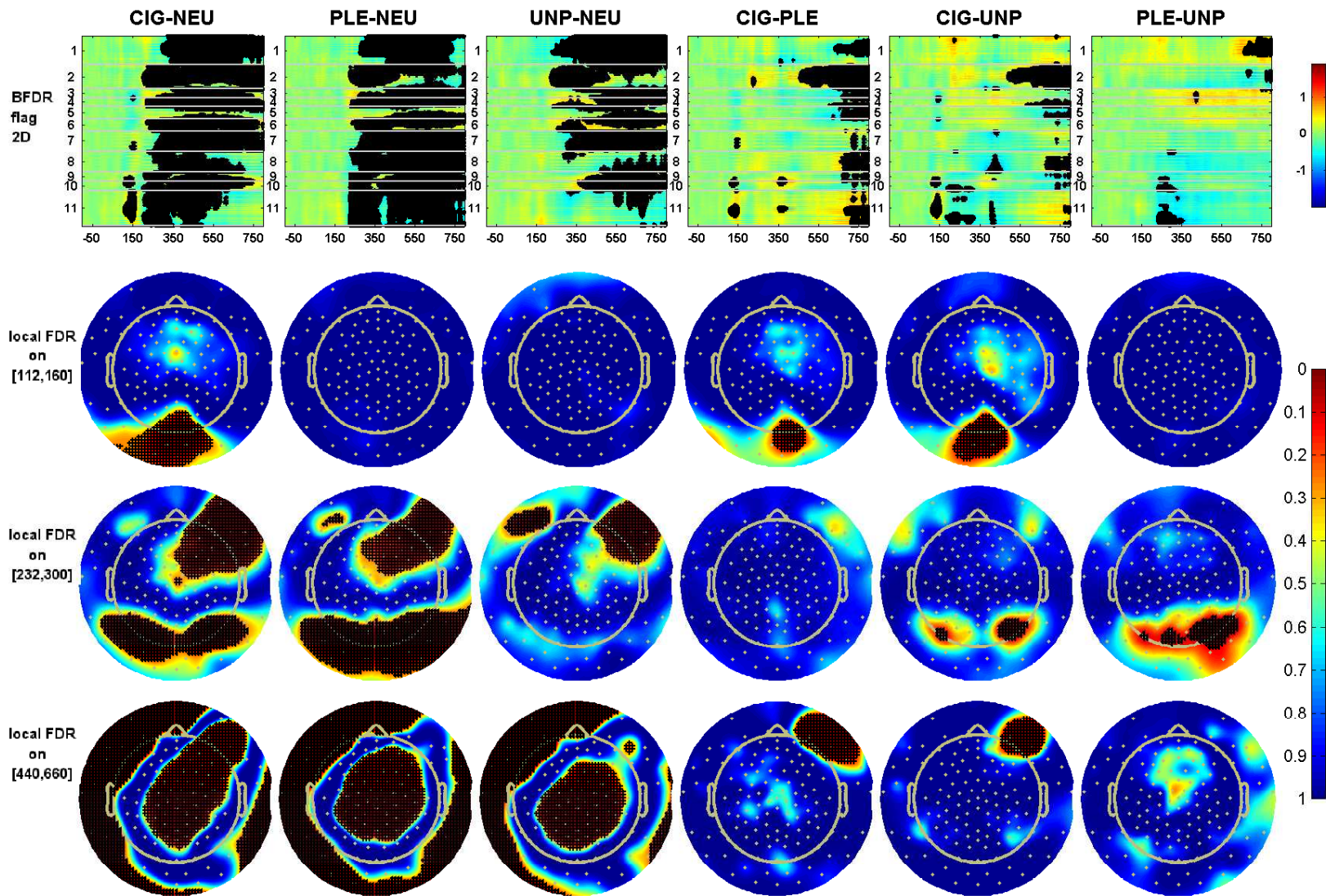


Figure 12: *Regions flagged using BFDR by fitting a global model to the full data set.* Row 1: integrated 2D heatmaps of mean contrast effects marked with BFDR ($\delta=0.5$) flagged regions (black dots)—the x-axis is time and the y-axis is vectorized spatial locations of the 2D scalp (indexed by region number). Row 2-4: scalp plots of local FDR at three time intervals ([112, 140] ms, [232, 300] ms, and [440, 660] ms), marked with BFDR flagged regions. Here the local FDR and the BFDR flagging results were re-calculated based on posterior samples averaged across the time intervals.

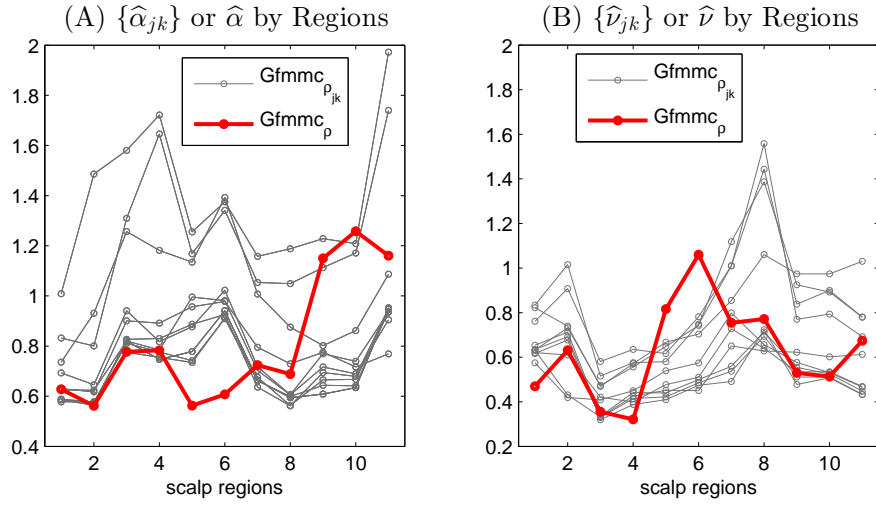


Figure 13: *Estimated Matérn parameters by regions:* the estimated $\hat{\alpha}$ parameters (A) and $\hat{\nu}$ parameters (B) across 11 regions. The x-axis denotes the scalp regions. In both plots, the gray lines denote $\{\hat{\alpha}_{jk}\}$ and $\{\hat{\nu}_{jk}\}$ for the first 10 wavelet components, estimated using the $\text{Gfmmc}_{\rho_{jk}}$ model; the red lines denote $\{\hat{\alpha}\}$ and $\{\hat{\nu}\}$ estimated using the Gfmmc_{ρ} model.

- Morris, J. S., and Carroll, R. J. (2006), “Wavelet-based functional mixed models,” J. R. Statist. Soc. B, 68, 179–199.
- Searle, S. R., Casella, G., and McCulloch, C. E., eds (1992), Variance Components, New York: John Wiley & Sons.
- Simpson, S. L., Edwards, L. J., Styner, M. A., and Muller, K. E. (2014), “Kronecker Product Linear Exponent AR(1) Correlation Structures for Multivariate Repeated Measures,” PLOS ONE, 9(2), 1–10.