

# Supplemental materials to “Reduced Rank Mixed Effects Models for Spatially Correlated Hierarchical Functional Data”

Lan Zhou, Jianhua Z. Huang, Josue G. Martinez,  
Arnab Maity, Veerabhadran Baladandayuthapani and Raymond J. Carroll

## 1 Outline

Section 2 details the computational methods for implementing the proposed methodology, including steps of the EM algorithm and techniques for avoiding storage and inverting large matrices. Section 3 provides additional simulation results. Section 4 contains residuals plots after model fitting for the real data example.

## 2 Computational methods

### 2.1 Computation of the Conditional Moments in the E-step of the EM Algorithm

In the following, the conditional expectations are calculated given the current parameter values; for simplicity of presentation, the dependence on the current parameter values is suppressed in our notation.

To calculate the conditional moments, we use some standard results for multivariate normal distributions. The covariances between  $\boldsymbol{\alpha}_{ab}$ ,  $\boldsymbol{\beta}_{ab}$  and  $\mathbf{Y}_{ab}$  are  $\text{cov}(\boldsymbol{\alpha}_{ab}, \mathbf{Y}_{ab}) = \mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^T \mathbf{B}_{ab}^T$  and  $\text{cov}(\boldsymbol{\beta}_{ab}, \mathbf{Y}_{ab}) = \mathbf{V}_{ab} \boldsymbol{\Gamma}_{\eta,ab}^T \mathbb{B}_{ab}^T$ . Since  $\text{cov}(\boldsymbol{\alpha}_{ab}, \boldsymbol{\beta}_{ab}) = 0$ , we have  $\text{cov}(\mathbf{Y}_{ab}) = \mathbf{B}_{ab} \boldsymbol{\Gamma}_{\xi} \mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^T \mathbf{B}_{ab}^T + \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \mathbf{V}_{ab} \boldsymbol{\Gamma}_{\eta,ab}^T \mathbb{B}_{ab}^T + \sigma^2 \mathbf{I}_{ab}$ , where  $\mathbf{I}_{ab}$  is the identity matrix of rank  $N_{ab}$ . The conditional distribution of  $(\boldsymbol{\alpha}_{ab}, \boldsymbol{\beta}_{ab})$  given  $\mathbf{Y}_{ab}$  is normal and is denoted as

$$\begin{pmatrix} \boldsymbol{\alpha}_{ab} \\ \boldsymbol{\beta}_{ab} \end{pmatrix} \sim N \left\{ \begin{pmatrix} \mathbf{m}_{\alpha,ab} \\ \mathbf{m}_{\beta,ab} \end{pmatrix}, \boldsymbol{\Sigma}_{ab} = \begin{pmatrix} \boldsymbol{\Sigma}_{\alpha\alpha,ab} & \boldsymbol{\Sigma}_{\alpha\beta,ab} \\ \boldsymbol{\Sigma}_{\alpha\beta,ab}^T & \boldsymbol{\Sigma}_{\beta\beta,ab} \end{pmatrix} \right\}. \quad (1)$$

The conditional means are

$$\mathbf{m}_{\alpha,ab} = E(\boldsymbol{\alpha}_{ab} | \mathbf{Y}_{ab}) = \mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^T \mathbf{B}_{ab}^T \text{cov}(\mathbf{Y}_{ab})^{-1} (\mathbf{Y}_{ab} - \mathbf{B}_{ab} \boldsymbol{\gamma}_{\mu,a}) \quad (2)$$

and

$$\mathbf{m}_{\beta,ab} = E(\boldsymbol{\beta}_{ab} | \mathbf{Y}_{ab}) = \mathbf{V}_{ab} \boldsymbol{\Gamma}_{\eta,ab}^T \mathbb{B}_{ab}^T \text{cov}(\mathbf{Y}_{ab})^{-1} (\mathbf{Y}_{ab} - \mathbf{B}_{ab} \boldsymbol{\gamma}_{\mu,a}). \quad (3)$$

The conditional covariance matrix is

$$\text{cov}\left\{\begin{pmatrix} \boldsymbol{\alpha}_{ab} \\ \boldsymbol{\beta}_{ab} \end{pmatrix} \middle| \mathbf{Y}_{ab}\right\} = \begin{pmatrix} \mathbf{D}_{\alpha,a} & 0 \\ 0 & \mathbf{V}_{ab} \end{pmatrix} - \begin{pmatrix} \mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^{\text{T}} \mathbf{B}_{ab}^{\text{T}} \\ \mathbf{V}_{ab} \boldsymbol{\Gamma}_{\eta,ab}^{\text{T}} \mathbb{B}_{ab}^{\text{T}} \end{pmatrix} \{\text{cov}(\mathbf{Y}_{ab})\}^{-1} (\mathbf{B}_{ab} \boldsymbol{\Gamma}_{\xi} \mathbf{D}_{\alpha,a}, \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \mathbf{V}_{ab}).$$

Therefore, we have

$$\boldsymbol{\Sigma}_{\alpha\alpha,ab} = \mathbf{D}_{\alpha,a} - \mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^{\text{T}} \mathbf{B}_{ab}^{\text{T}} \{\text{cov}(\mathbf{Y}_{ab})\}^{-1} \mathbf{B}_{ab} \boldsymbol{\Gamma}_{\xi} \mathbf{D}_{\alpha,a}, \quad (4)$$

$$\boldsymbol{\Sigma}_{\beta\beta,ab} = \mathbf{V}_{ab} - \mathbf{V}_{ab} \boldsymbol{\Gamma}_{\eta,ab}^{\text{T}} \mathbb{B}_{ab}^{\text{T}} \{\text{cov}(\mathbf{Y}_{ab})\}^{-1} \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \mathbf{V}_{ab}, \quad (5)$$

and

$$\boldsymbol{\Sigma}_{\alpha\beta,ab} = -\mathbf{D}_{\alpha,a} \boldsymbol{\Gamma}_{\xi}^{\text{T}} \mathbf{B}_{ab}^{\text{T}} \{\text{cov}(\mathbf{Y}_{ab})\}^{-1} \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \mathbf{V}_{ab}. \quad (6)$$

The desired predictions required by the EM algorithm are

$$\begin{aligned} \widehat{\boldsymbol{\alpha}}_{ab} &= E(\boldsymbol{\alpha}_{ab} | \mathbf{Y}_{ab}) = \mathbf{m}_{\alpha,ab}, & \widehat{\boldsymbol{\beta}}_{ab} &= E(\boldsymbol{\beta}_{ab} | \mathbf{Y}_{ab}) = \mathbf{m}_{\beta,ab}, \\ \widehat{\boldsymbol{\alpha}_{ab} \boldsymbol{\alpha}_{ab}^{\text{T}}} &= E(\boldsymbol{\alpha}_{ab} \boldsymbol{\alpha}_{ab}^{\text{T}} | \mathbf{Y}_{ab}) = \widehat{\boldsymbol{\alpha}}_{ab} \widehat{\boldsymbol{\alpha}}_{ab}^{\text{T}} + \boldsymbol{\Sigma}_{\alpha\alpha,ab} \\ \widehat{\boldsymbol{\beta}_{ab} \boldsymbol{\beta}_{ab}^{\text{T}}} &= E(\boldsymbol{\beta}_{ab} \boldsymbol{\beta}_{ab}^{\text{T}} | \mathbf{Y}_{ab}) = \widehat{\boldsymbol{\beta}}_{ab} \widehat{\boldsymbol{\beta}}_{ab}^{\text{T}} + \boldsymbol{\Sigma}_{\beta\beta,ab}, \\ \widehat{\boldsymbol{\alpha}_{ab} \boldsymbol{\beta}_{ab}^{\text{T}}} &= E(\boldsymbol{\alpha}_{ab} \boldsymbol{\beta}_{ab}^{\text{T}} | \mathbf{Y}_{ab}) = \widehat{\boldsymbol{\alpha}}_{ab} \widehat{\boldsymbol{\beta}}_{ab}^{\text{T}} + \boldsymbol{\Sigma}_{\alpha\beta,ab}. \end{aligned} \quad (7)$$

The first two equalities also give expressions of the best linear unbiased predictors (BLUP) of the random effects  $\boldsymbol{\alpha}_{ab}$  and  $\boldsymbol{\beta}_{ab}$ .

**Computation of  $\mathbf{V}_{ab}$ .** It is convenient to re-group the elements of  $\boldsymbol{\beta}_{ab}$  according to the principal components. Denote  $\boldsymbol{\beta}_{ab,j} = (\beta_{ab1j}, \dots, \beta_{abC_{ab}j})^{\text{T}}$ ,  $j = 1, \dots, K_{\eta}$ , and  $\widetilde{\boldsymbol{\beta}}_{ab} = (\boldsymbol{\beta}_{ab,1}^{\text{T}}, \dots, \boldsymbol{\beta}_{ab,K_{\eta}}^{\text{T}})^{\text{T}}$ . Let  $\mathbf{P}_{abj}$  be a  $C_{ab}$  by  $C_{ab}$  matrix with elements  $P_{abj,cc'} = \rho(|x_{abc} - x_{abc'}|; \theta_{aj})$ ,  $c, c' = 1, \dots, C_{ab}$ . Note that  $\mathbf{P}_{abj} = \mathbf{P}_{abj}^{\text{T}}$ . Then  $\boldsymbol{\Sigma}_{abj} = \text{cov}(\boldsymbol{\beta}_{ab,j}) = \sigma_{\beta,aj}^2 \mathbf{P}_{abj}$ . Since  $\text{cov}(\boldsymbol{\beta}_{ab,j}, \boldsymbol{\beta}_{ab,j'}) = 0$  for  $j \neq j'$ , the covariance matrix of  $\widetilde{\boldsymbol{\beta}}_{ab}$  is block diagonal  $\widetilde{\mathbf{V}}_{ab} = \text{cov}(\widetilde{\boldsymbol{\beta}}_{ab}) = \text{diag}(\boldsymbol{\Sigma}_{ab1}, \dots, \boldsymbol{\Sigma}_{abK_{\eta}})$ . Note that  $\boldsymbol{\beta}_{ab}$  is just a reordering of  $\widetilde{\boldsymbol{\beta}}_{ab}$ , that is,  $\boldsymbol{\beta}_{ab} = \mathbf{O} \widetilde{\boldsymbol{\beta}}_{ab}$  for a permutation matrix  $\mathbf{O}$ . It follows the covariance matrix of  $\boldsymbol{\beta}_{ab}$  is  $\mathbf{V}_{ab} = \mathbf{O} \widetilde{\mathbf{V}}_{ab} \mathbf{O}^{\text{T}}$ .

**Circumventing inversion of  $\text{cov}(\mathbf{Y}_{ab})$ .** We suppress the subscripts of  $\mathbf{D}_{\alpha,a}$  and  $\mathbf{V}_{ab}$  in the following discussion. Denote  $\mathbf{E} = \mathbf{B}_{ab} \boldsymbol{\Gamma}_{\xi}$ ,  $\mathbf{F} = \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab}$ , and  $\mathbf{S} = \text{cov}(\mathbf{Y}_{ab}) = \mathbf{E} \mathbf{E}^{\text{T}} + \mathbf{F} \mathbf{F}^{\text{T}} + \sigma^2 \mathbf{I}$ . To calculate the conditional moments given in (2)–(6), we need to compute  $\mathbf{D} \mathbf{E}^{\text{T}} \mathbf{S}^{-1}$  in (2),  $\mathbf{V} \mathbf{F}^{\text{T}} \mathbf{S}^{-1}$  in (3),  $\mathbf{D} - \mathbf{D} \mathbf{E}^{\text{T}} \mathbf{S}^{-1} \mathbf{E} \mathbf{D}$  in (4),  $\mathbf{V} - \mathbf{V} \mathbf{F}^{\text{T}} \mathbf{S}^{-1} \mathbf{F} \mathbf{V}$  in (6), and  $-\mathbf{D} \mathbf{E}^{\text{T}} \mathbf{S}^{-1} \mathbf{F} \mathbf{V}$  in (5). Denote  $\mathbf{J} = (\mathbf{E} \mathbf{E}^{\text{T}} + \sigma^2 \mathbf{I})^{-1}$  and  $\mathbf{K} = (\mathbf{F} \mathbf{F}^{\text{T}} + \sigma^2 \mathbf{I})^{-1}$ . Repeatedly using the identities

$$(\mathbf{A}^{-1} + \mathbf{C}^{\text{T}} \mathbf{B}^{-1} \mathbf{C})^{-1} = \mathbf{A} - \mathbf{A} \mathbf{C}^{\text{T}} (\mathbf{C} \mathbf{A} \mathbf{C}^{\text{T}} + \mathbf{B})^{-1} \mathbf{C} \mathbf{A}, \quad (8)$$

$$(\mathbf{A}^{-1} + \mathbf{C}^{\text{T}} \mathbf{B}^{-1} \mathbf{C})^{-1} \mathbf{C}^{\text{T}} \mathbf{B}^{-1} = \mathbf{A} \mathbf{C}^{\text{T}} (\mathbf{C} \mathbf{A} \mathbf{C}^{\text{T}} + \mathbf{B})^{-1}, \quad (9)$$

we obtain

$$\mathbf{J} = \sigma^{-2}\mathbf{I} - \sigma^{-2}\mathbf{E}(\mathbf{E}^T\mathbf{E} + \sigma^2\mathbf{D}^{-1})^{-1}\mathbf{E}^T, \quad (10)$$

$$\mathbf{K} = \sigma^{-2}\mathbf{I} - \sigma^{-2}\mathbf{F}(\mathbf{F}^T\mathbf{F} + \sigma^2\mathbf{V}^{-1})^{-1}\mathbf{F}^T, \quad (11)$$

and

$$\begin{aligned} \mathbf{D}\mathbf{E}^T\mathbf{S}^{-1} &= (\mathbf{D}^{-1} + \mathbf{E}^T\mathbf{K}\mathbf{E})^{-1}\mathbf{E}^T\mathbf{K}, \\ \mathbf{V}\mathbf{F}^T\mathbf{S}^{-1} &= (\mathbf{V}^{-1} + \mathbf{F}^T\mathbf{J}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{J}, \\ \mathbf{D} - \mathbf{D}\mathbf{E}^T\mathbf{S}^{-1}\mathbf{E}\mathbf{D} &= (\mathbf{D}^{-1} + \mathbf{E}^T\mathbf{K}\mathbf{E})^{-1}, \\ \mathbf{V} - \mathbf{V}\mathbf{F}^T\mathbf{S}^{-1}\mathbf{F}\mathbf{V} &= (\mathbf{V}^{-1} + \mathbf{F}^T\mathbf{J}\mathbf{F})^{-1}, \\ -\mathbf{D}\mathbf{E}^T\mathbf{S}^{-1}\mathbf{F}\mathbf{V} &= -\mathbf{D}\mathbf{E}^T\mathbf{J}\mathbf{F}(\mathbf{V}^{-1} + \mathbf{F}^T\mathbf{J}\mathbf{F})^{-1}. \end{aligned}$$

The matrices that need to be inverted here are of the same size as  $\mathbf{D}$  or  $\mathbf{V}$ , which is  $K_\xi \times K_\xi$  or  $(C_{ab}K_\eta) \times (C_{ab}K_\eta)$ , much smaller than the size of  $\text{cov}(\mathbf{Y}_{ab})$ .

## 2.2 Updating Formula in the M-step of the EM Algorithm

In the updating formulas given below, the parameters appear on the right hand side of equations are all fixed at their values obtained from the previous iteration of the algorithm.

1. Update the estimate of  $\sigma^2$ . Let  $N = \sum_{a=1}^A \sum_{b=1}^{B_a} N_{ab}$ . The updating formula is

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{a=1}^A \sum_{b=1}^{B_a} E(\boldsymbol{\epsilon}_{ab}^T \boldsymbol{\epsilon}_{ab} | \mathbf{Y}_{ab}),$$

where  $\boldsymbol{\epsilon}_{ab} = \mathbf{Y}_{ab} - \mathbf{B}_{ab}\boldsymbol{\gamma}_{\mu,a} - \mathbf{B}_{ab}\boldsymbol{\Gamma}_\xi\boldsymbol{\alpha}_{ab} - \mathbb{B}_{ab}\boldsymbol{\Gamma}_{\eta,ab}\boldsymbol{\beta}_{ab}$ . Using (7), we obtain that  $E(\boldsymbol{\epsilon}_{ab}^T \boldsymbol{\epsilon}_{ab} | \mathbf{Y}_{ab})$  can be expressed as

$$\begin{aligned} &(\mathbf{Y}_{ab} - \mathbf{B}_{ab}\boldsymbol{\gamma}_{\mu,a} - \mathbf{B}_{ab}\boldsymbol{\Gamma}_\xi\hat{\boldsymbol{\alpha}}_{ab} - \mathbb{B}_{ab}\boldsymbol{\Gamma}_{\eta,ab}\hat{\boldsymbol{\beta}}_{ab})^T (\mathbf{Y}_{ab} - \mathbf{B}_{ab}\boldsymbol{\gamma}_{\mu,a} - \mathbf{B}_{ab}\boldsymbol{\Gamma}_\xi\hat{\boldsymbol{\alpha}}_{ab} - \mathbb{B}_{ab}\boldsymbol{\Gamma}_{\eta,ab}\hat{\boldsymbol{\beta}}_{ab}) \\ &+ \text{trace}(\mathbf{B}_{ab}\boldsymbol{\Gamma}_\xi\boldsymbol{\Sigma}_{\alpha\alpha,ab}\boldsymbol{\Gamma}_\xi^T\mathbf{B}_{ab}^T + \mathbb{B}_{ab}\boldsymbol{\Gamma}_{\eta,ab}\boldsymbol{\Sigma}_{\beta\beta,ab}\boldsymbol{\Gamma}_{\eta,ab}^T\mathbb{B}_{ab}^T + 2\mathbf{B}_{ab}\boldsymbol{\Gamma}_\xi\boldsymbol{\Sigma}_{\alpha\beta,ab}\boldsymbol{\Gamma}_{\eta,ab}^T\mathbb{B}_{ab}^T). \end{aligned}$$

2. Update  $\mathbf{D}_{\alpha,a} = \text{diag}(\sigma_{\alpha,a_j}^2)$  and  $\mathbf{D}_{\beta,a} = \text{diag}(\sigma_{\beta,a_j}^2)$ . The updating formula for  $\mathbf{D}_{\alpha,a}$  is

$$\hat{\mathbf{D}}_{\alpha,a} = \text{diag}\left\{\frac{1}{B_a} \sum_{b=1}^{B_a} \widehat{\boldsymbol{\alpha}_{ab}\boldsymbol{\alpha}_{ab}^T}\right\}. \quad (12)$$

To update  $\mathbf{D}_{\beta,a}$ , we need to minimize the conditional expectation of

$$\sum_{b=1}^{B_a} \sum_{j=1}^{K_\eta} \log(|\mathbf{V}_{ab}|) + \boldsymbol{\beta}_{ab}^T \mathbf{V}_{ab}^{-1} \boldsymbol{\beta}_{ab} = \sum_{b=1}^{B_a} \sum_{j=1}^{K_\eta} \left\{ C_{ab} \log(\sigma_{\beta,a_j}^2) + \log|\mathbf{P}_{abj}| + \sigma_{\beta,a_j}^{-2} \boldsymbol{\beta}_{ab,j}^T \mathbf{P}_{ab,j}^{-1} \boldsymbol{\beta}_{ab,j} \right\}.$$

Thus the updating formulas are

$$\widehat{\sigma}_{\beta,a,j}^2 = \text{trace} \left( \frac{1}{\sum_{b=1}^{B_a} C_{ab}} \mathbf{P}_{ab,j}^{-1} \widehat{\boldsymbol{\beta}}_{ab,j} \widehat{\boldsymbol{\beta}}_{ab,j}^{\text{T}} \right), \quad j = 1, \dots, K_{\eta}. \quad (13)$$

If the variances of random effects do not depend on the treatment level  $a$ , we simply need to average over  $a$  in (12) and (13).

3. Update  $\boldsymbol{\gamma}_{\mu,a}$ ,  $a = 1, \dots, A$ . The updating formula are

$$\boldsymbol{\gamma}_{\mu,a} = \left\{ \sum_{b=1}^{B_a} \mathbf{B}_{ab}^{\text{T}} \mathbf{B}_{ab} + \sigma^2 \lambda_{\mu} \int \mathbf{b}''(t) \mathbf{b}''(t)^{\text{T}} \right\}^{-1} \sum_{b=1}^{B_a} \mathbf{B}_{ab}^{\text{T}} (\mathbf{Y}_{ab} - \mathbf{B}_{ab} \boldsymbol{\Gamma}_{\xi} \widehat{\boldsymbol{\alpha}}_{ab} - \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \widehat{\boldsymbol{\beta}}_{ab}).$$

4. Update the columns of  $\boldsymbol{\Gamma}_{\xi}$  sequentially. For  $j = 1, \dots, K_{\xi}$ , the updating formula are

$$\boldsymbol{\gamma}_{\xi,j} = \left\{ \sum_{a=1}^A \sum_{b=1}^{B_a} \widehat{\alpha}_{abj}^2 \mathbf{B}_{ab}^{\text{T}} \mathbf{B}_{ab} + \sigma^2 \lambda_{\xi} \int \mathbf{b}''(t) \mathbf{b}''(t)^{\text{T}} \right\}^{-1} \times \sum_{a=1}^A \sum_{b=1}^{B_a} \mathbf{B}_{ab}^{\text{T}} \left\{ (\mathbf{Y}_{ab} - \mathbf{B}_{ab} \boldsymbol{\gamma}_{\mu,a}) \widehat{\alpha}_{abj} - \mathbf{B}_{ab} \left( \sum_{l \neq j} \boldsymbol{\gamma}_{\xi,l} \widehat{\alpha}_{abl} \right) - \mathbb{B}_{ab} \boldsymbol{\Gamma}_{\eta,ab} \widehat{\boldsymbol{\beta}}_{ab} \right\}.$$

5. Update the columns of  $\boldsymbol{\Gamma}_{\eta}$  sequentially. For  $j = 1, \dots, K_{\eta}$ , the updating formula are

$$\boldsymbol{\gamma}_{\eta,j} = \left\{ \sum_{a=1}^A \sum_{b=1}^{B_a} \sum_{c=1}^{C_{ab}} \widehat{\beta}_{abcj}^2 \mathbf{B}_{abc}^{\text{T}} \mathbf{B}_{abc} + \sigma^2 \lambda_{\eta,a} \int \mathbf{b}''(t) \mathbf{b}''(t)^{\text{T}} \right\}^{-1} \times \sum_{a=1}^A \sum_{b=1}^{B_a} \sum_{c=1}^{C_{ab}} \mathbf{B}_{abc}^{\text{T}} \left\{ (\mathbf{Y}_{abc} - \mathbf{B}_{abc} \boldsymbol{\gamma}_{\mu,a}) \widehat{\beta}_{abcj} - \mathbf{B}_{abc} \boldsymbol{\Gamma}_{\xi} \widehat{\boldsymbol{\beta}}_{abcj} \boldsymbol{\alpha}_{ab} - \mathbf{B}_{abc} \left( \sum_{l \neq j} \boldsymbol{\gamma}_{\eta,l} \widehat{\beta}_{abcl} \right) \right\}.$$

6. Orthogonalizing  $\boldsymbol{\Gamma}_{\xi}$  and  $\boldsymbol{\Gamma}_{\eta}$ . The matrix  $\boldsymbol{\Gamma}_{\xi}$  and  $\boldsymbol{\Gamma}_{\eta}$  obtained in Steps 4 and 5 need not have orthonormal columns. We orthogonalize them in this step. Compute the QR decomposition  $\widehat{\boldsymbol{\Gamma}}_{\xi} = \mathbf{Q}_{\xi} \mathbf{R}_{\xi}$  where  $\mathbf{Q}_{\xi}$  has orthonormal columns. Let  $\widehat{\boldsymbol{\Gamma}}_{\xi} \leftarrow \mathbf{Q}_{\xi}$ . Orthogonalize  $\boldsymbol{\Gamma}_{\eta}$  similarly.

7. Update the correlation parameter  $\theta_{aj}$ . Given the current estimates of other parameters, we minimize for each  $a$  and  $j$  the conditional expectation of

$$\sum_{b=1}^{B_a} \left\{ \log |\mathbf{P}_{abj}^{-1}| + \frac{1}{\sigma_{\beta,a,j}^2} \boldsymbol{\beta}_{ab,j}^{\text{T}} \mathbf{P}_{abj}^{-1} \boldsymbol{\beta}_{ab,j} \right\},$$

which is

$$\sum_{b=1}^{B_a} \left\{ \log |\mathbf{P}_{abj}^{-1}| + \frac{1}{\sigma_{\beta,a,j}^2} \text{trace} \left( \mathbf{P}_{abj}^{-1} \widehat{\boldsymbol{\beta}}_{ab,j} \widehat{\boldsymbol{\beta}}_{ab,j}^{\text{T}} \right) \right\}. \quad (14)$$

The minimizer can be found by a gradient-based method. Denote the components of  $\theta_{aj}$  as  $\theta_{ajk}$ 's. The gradient with respect to  $\theta_{aj}$  is a vector with elements

$$\sum_{b=1}^{B_a} \left\{ \text{trace} \left( \mathbf{P}_{abj}^{-1} \frac{\partial \mathbf{P}_{abj}}{\partial \theta_{ajk}} \right) - \frac{1}{\sigma_{\beta,aj}^2} \text{trace} \left( \mathbf{P}_{abj}^{-1} \frac{\partial \mathbf{P}_{abj}}{\partial \theta_{ajk}} \mathbf{P}_{abj}^{-1} \widehat{\boldsymbol{\beta}}_{ab,j} \widehat{\boldsymbol{\beta}}_{ab,j}^{\text{T}} \right) \right\}.$$

Alternatively, a direct search algorithm that does not require the calculation of the gradient, such as the downhill simplex method of Nelder and Mead (1965), can be applied to minimize (14). If the correlation parameters do not depend on  $j$ , we consider for each  $a$  a new objective function by summing (14) over  $j$ . If the correlation parameters do not depend on  $a$ , we need only sum (14) over  $a$  for each  $j$  to define the new objective function.

Note that the orthogonalization in Step 6 is used to ensure identifiability. This approach was adapted from Zhou et al. (2008) and worked well in our simulation study and real data analysis. An alternative approach is direct optimization within a restricted space (Peng and Paul, 2009) but its implementation is beyond the scope of this paper.

### 2.3 Computing the Observed Data Log Likelihood

When we do crossvalidation, we need to compute the log likelihood of observed data, which depends on the determinant and inverse of the possibly very large matrix  $\mathbf{S}$ , defined in Section 2.1. This section gives a method for computation of  $|\mathbf{S}|$  and  $\mathbf{S}^{-1}$  without constructing the matrix  $\mathbf{S}$ . Using the identity

$$|\mathbf{A} + \mathbf{BC}^{\text{T}}| = |\mathbf{A}| |\mathbf{I} + \mathbf{C}^{\text{T}} \mathbf{A}^{-1} \mathbf{B}|, \quad (15)$$

we obtain

$$|\mathbf{S}| = |\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}} + \mathbf{FVF}^{\text{T}}| = |\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}}| |\mathbf{I} + \mathbf{V}^{1/2} \mathbf{F}^{\text{T}} (\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}})^{-1} \mathbf{FV}^{1/2}|.$$

The two factors on the right hand side of the above equation can be computed as follows. Using the identity (15), we have

$$\begin{aligned} |\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}}| &= \sigma^{2N} |\mathbf{I} + \sigma^{-2} \mathbf{D}^{1/2} \mathbf{E}^{\text{T}} \mathbf{E} \mathbf{D}^{1/2}| \\ &= \sigma^{2N} |\mathbf{D}| |\mathbf{D}^{-1} + \sigma^{-2} \mathbf{E}^{\text{T}} \mathbf{E}|. \end{aligned}$$

Using  $(\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}})^{-1} = \sigma^{-2} \mathbf{I} - \sigma^{-2} \mathbf{E} (\mathbf{E}^{\text{T}} \mathbf{E} + \sigma^2 \mathbf{D}^{-1})^{-1} \mathbf{E}^{\text{T}}$ , we obtain

$$\begin{aligned} &|\mathbf{I} + \mathbf{V}^{1/2} \mathbf{F}^{\text{T}} (\sigma^2 \mathbf{I} + \mathbf{EDE}^{\text{T}})^{-1} \mathbf{FV}^{1/2}| \\ &= |\mathbf{V}| |\mathbf{V}^{-1} + \sigma^{-2} \mathbf{F}^{\text{T}} \mathbf{F} - \sigma^{-2} \mathbf{F}^{\text{T}} \mathbf{E} (\mathbf{E}^{\text{T}} \mathbf{E} + \sigma^2 \mathbf{D}^{-1})^{-1} \mathbf{E}^{\text{T}} \mathbf{F}|. \end{aligned}$$

Using the definition of  $\mathbf{J}$  in Section 2.1 and the identity (8), we have

$$\mathbf{S}^{-1} = (\mathbf{J}^{-1} + \mathbf{F}\mathbf{V}\mathbf{F}^T)^{-1} = \mathbf{J} - \mathbf{J}\mathbf{F}(\mathbf{F}^T\mathbf{J}\mathbf{F} + \mathbf{V}^{-1})^{-1}\mathbf{F}^T\mathbf{J},$$

where  $\mathbf{J}$  can be computed using (10).

## 2.4 Calculating the partial derivatives of $P_{abj}$

Consider the Matérn family of autocorrelation functions

$$\rho(d; \phi, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{2d\nu^{1/2}}{\phi} \right)^\nu K_\nu \left( \frac{2d\nu^{1/2}}{\phi} \right), \quad \phi > 0, \quad \nu > 0,$$

where  $K_\nu(\cdot)$  is the modified Bessel function of order  $\nu$  taking the form

$$K_\nu(u) = \frac{\pi}{2 \sin(\nu\pi)} \{I_{-\nu}(u) - I_\nu(u)\}$$

with

$$I_\nu(u) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m + \nu + 1)} \left( \frac{u}{2} \right)^{2m+\nu}.$$

Denote  $u = 2x\nu^{1/2}/\phi$ , then the partial derivatives of  $\rho(x; \phi, \nu)$  are

$$\frac{\partial \rho(x; \phi, \nu)}{\partial \phi} = \frac{1}{\phi \Gamma(\nu)} \left( \frac{u}{2} \right)^\nu [u \{K_{\nu-1}(u) + K_{\nu+1}(u)\} - 2\nu K_\nu(u)]$$

and

$$\begin{aligned} \frac{\partial \rho(x; \phi, \nu)}{\partial \nu} = & \left\{ -\psi(\nu) - \pi \cot(\nu\pi) + \log\left(\frac{u}{2}\right) + \frac{1}{2} \right\} \rho(x; \phi, \nu) \\ & + \frac{\pi}{\Gamma(\nu) \sin(\nu\pi)} \left( \frac{u}{2} \right)^\nu \left[ \frac{u}{2\nu} \{I_{-\nu+1}(u) - I_{\nu+1}(u)\} \right. \\ & - \left. \left\{ \log\left(\frac{u}{2}\right) + \frac{1}{2} \right\} \{I_\nu(u) + I_{-\nu}(u)\} \right. \\ & + \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m + \nu + 1)} \left( \frac{u}{2} \right)^{2m+\nu} \psi(m + \nu + 1) \\ & \left. + \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m - \nu + 1)} \left( \frac{u}{2} \right)^{2m-\nu} \psi(m - \nu + 1) \right], \end{aligned}$$

where  $\psi(z) = -C - \sum_{l=0}^{\infty} \left( \frac{1}{z+l} - \frac{1}{1+l} \right)$  with  $C = \lim_{k \rightarrow \infty} \left\{ \sum_{l=1}^k \frac{1}{l} - \ln(k) \right\} \approx .5572$ , and  $\Gamma(\cdot)$  is the Gamma function.

### 3 Additional simulation results

Two other setups from the Bayesian hierarchical model in Baladandayuthapani, et al. (2008) were also considered but the software of the Bayesian method encountered some serious numerical problems and could only run on a small proportion of simulated data sets. We ran simulation 100 times for each setup. The results of our methods are based on three unit and subunit level principal components. Our methods successfully ran on all data sets.

**Setup 6.** Same as Setup 5 except that

$$\Sigma_{21}^* = \Sigma_{22}^* = \begin{pmatrix} 1.00 & 0.84 & 0.18 \\ 0.84 & 1.44 & 0.54 \\ 0.18 & 0.54 & 0.81 \end{pmatrix}$$

and

$$\Sigma_{31}^* = \Sigma_{32}^* = \begin{pmatrix} 1.440 & 0.756 & 0.240 \\ 0.756 & 0.810 & 0.450 \\ 0.240 & 0.450 & 1.00 \end{pmatrix}.$$

For this setup, the Bayesian method only ran on 10 out of 100 simulated data sets.

**Setup 7.** Same as Setup 6 except that  $\sigma_2^2 = \sigma_3^2 = 0.4$ . For this setup, the Bayesian method only ran on 23 out of 100 simulated data sets.

We used the measures defined at the beginning of Section 5 of the paper to assess/compare performance of the two methods. For our reduced rank methods, the measures were computed using all 100 data sets as well as only those data sets that the Bayesian method ran. The results are summarized in Table 1. Our reduced rank method is comparable to the Bayesian method for estimating the mean functions and predicting the random effects.

Next we report some simulation results on parameter estimation for setups following our model. Note that comparison with the Bayesian method is not available since two methods use different parameters. Table 2 shows the summary statistics of correlation parameter estimates for the first three simulation setups where the data were generated from our model. The parameter estimates are reasonably unbiased. Figure 1 presents the pointwise sample mean of the estimated treatment effects, unit level and sub-unit level PCs, along with 90% pointwise coverage intervals for simulation Setup 2. It is not surprising that sub-unit level PCs are better estimated than the unit level PCs since there are more sub-unit level data than the unit level data.

Table 1: Comparison of two methods based on 100 simulation runs. Mean (SD) of the integrated absolute errors of estimating the mean functions and predicting the unit and subunit level random effects. Numbers shown are the actual numbers multiplied by 10. “Reduced rank” refers to our method; “Bayesian” refers to the Bayesian method of Baladandayuthapani, et al. (2008). “100 sims”, “10 sims” and “23 sims” indicate the number of data sets used in the calculation.

Setup	Method	Mean	Unit	Subunit
6	Reduced rank (100 sims)	4.137 (1.425)	5.568 (1.088)	4.105 (0.481)
	Reduced rank (10 sims)	4.904 (0.963)	6.148 (1.044)	3.939 (0.305)
	Bayesian (10 sims)	4.901 (0.963)	6.255 (0.899)	4.081 (0.293)
7	Reduced rank (100 sims)	4.478 (1.515)	7.737 (1.187)	6.800 (1.064)
	Reduced rank (23 sims)	4.841 (1.342)	7.826 (0.890)	6.458 (0.845)
	Bayesian (23 sims)	4.844 (1.344)	7.824 (0.929)	6.549 (0.567)

Table 2: Sample mean (SD) of correlation parameter estimates, based on 100 simulation runs.

Setup	PC 1		PC 2	
	$\phi = 8$	$\nu = 0.1$	$\phi = 4$	$\nu = 0.3$
1	7.708 (4.298)	0.104 (0.020)		
2	9.757 (5.672)	0.103 (0.018)	4.226 (1.306)	0.308 (0.040)
3	7.630 (2.369)	0.105 (0.019)	4.135 (1.129)	0.311 (0.036)

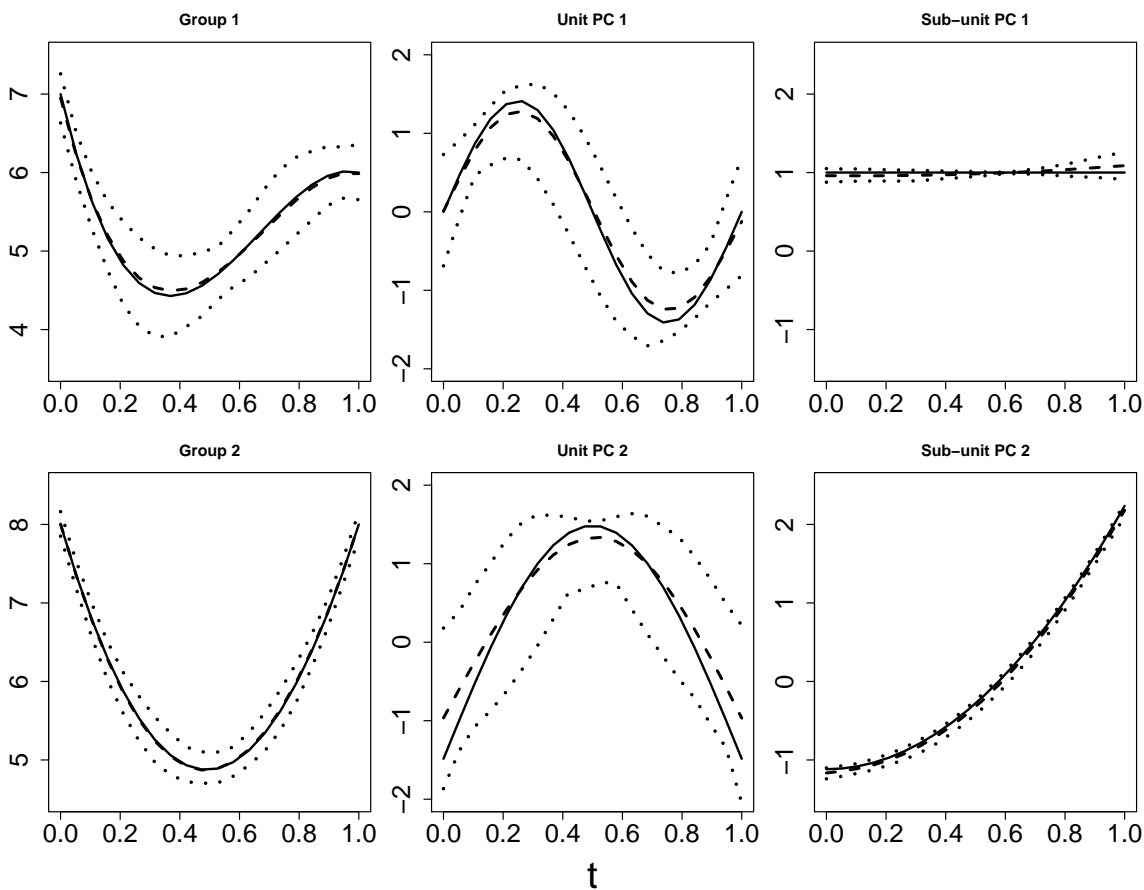
## 4 Colon Carcinogenesis Data

Figures 2 and 3 contains respectively residual plots after fitting our reduced rank model and the Bayesian model of Baladandayuthapani, et al. (2008). It is clear that our model fits the data better. After fitting the Bayesian model, there were still obvious patterns left in the residual plots for treatment groups FO+B and FO-B.

## References

- Baladandayuthapani, V., Mallick, B., Turner, N., Hong, M., Chapkin, R., Lupton, J. and Carroll, R. J. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* **64**, 64–73.
- Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–13.
- Peng, J. and Paul, D. (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, to appear.

Figure 1: Simulation Setup 2. The pointwise sample mean of the estimated treatment effects, unit level and sub-unit level PCs, along with 90% pointwise coverage intervals.



Zhou, L., Huang, J.Z. and Carroll, J.R. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika*, **95**, 601–619.

Figure 2: Residual plots after fitting the reduced rank model. Separate plots are drawn for the four diet groups. CO is Corn Oil, FO is Fish Oil, +B and -B represent with or without ( $\pm$ ) Butyrate supplement.

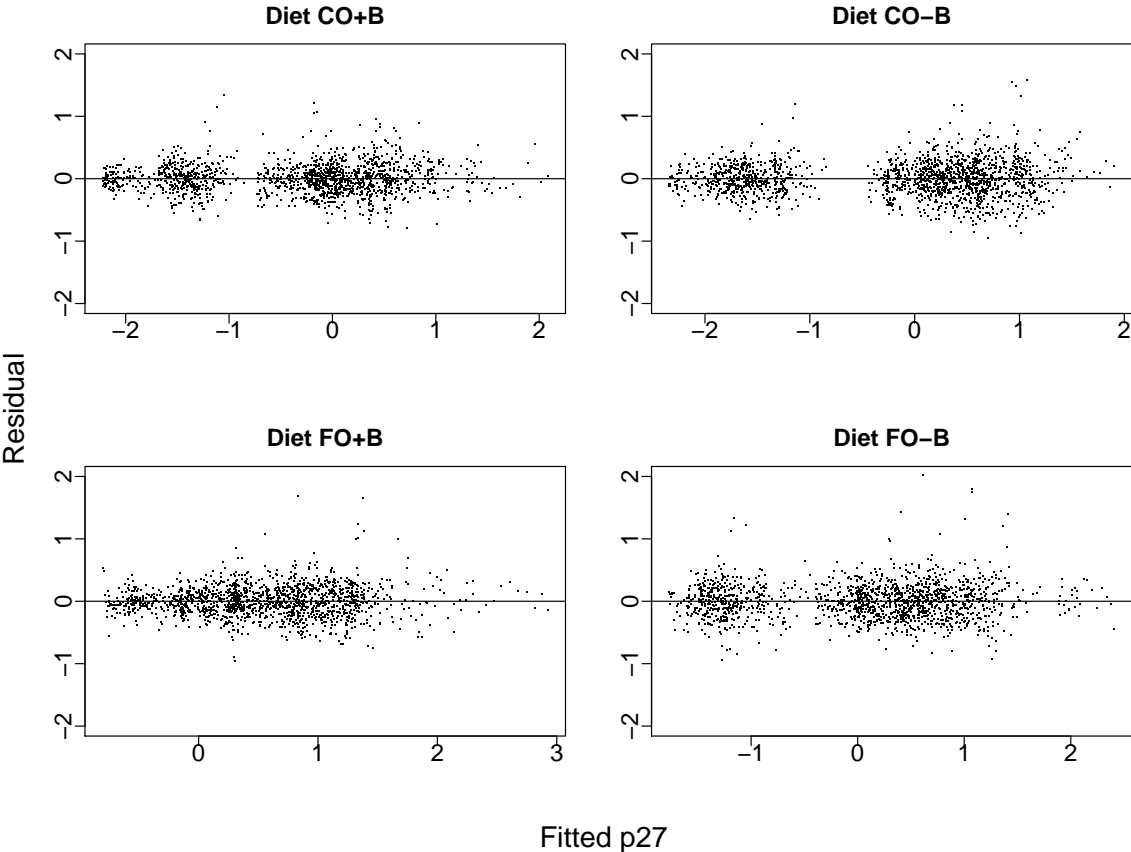


Figure 3: Residual plots after fitting the Bayesian model of Baladandayuthapani, et al. (2008). Separate plots are drawn for the four diet groups. CO is Corn Oil, FO is Fish Oil, +B and -B represent with or without Butyrate supplement.

