

Supplementary materials for *Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing*

by Yufen Zhang, James S. Hodges and Sudipto Banerjee.

APPENDICES

A Derivation of the precision matrix for Eq.(11)

Proof. Let $\mathbf{T} = (\boldsymbol{\Theta}'_{CO}, \boldsymbol{\Theta}'_{CO \times CA_1}, \dots, \boldsymbol{\Theta}'_{CO \times CA_{n-1}})'$; $f(\mathbf{T}|W) \propto \exp(-\frac{1}{2}\mathbf{T}'W\mathbf{T})$ with precision matrix W . Define $\mathbf{F} = K\mathbf{T}$, where $K = (V^{(-)} \otimes \frac{1}{\sqrt{n}}\mathbf{1}_n | V^{(-)} \otimes H_{CA}^{(1)} | \dots | V^{(-)} \otimes H_{CA}^{(n-1)})$ as in section ???. Because $K'K = I_{n(N-1)}$, therefore $K'\mathbf{F} = K'K\mathbf{T} = \mathbf{T}$. Changing variables in f from \mathbf{T} to \mathbf{F} , $f(\mathbf{F}|W) \propto \exp(-\frac{1}{2}\mathbf{F}'KWK'\mathbf{F})$, so \mathbf{F} has precision matrix KWK' . With

$$\mathbf{W} = \begin{bmatrix} D^{(-)} \otimes \tau_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D^{(-)} \otimes \tau_{n-1} \end{bmatrix},$$

$K\mathbf{T}$ has precision KWK'

$$\begin{aligned} &= (V^{(-)} \otimes \frac{1}{\sqrt{n}}\mathbf{1}_n | \dots | V^{(-)} \otimes H_{CA}^{(n-1)}) \begin{bmatrix} D^{(-)} \otimes \tau_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D^{(-)} \otimes \tau_{n-1} \end{bmatrix} \begin{pmatrix} V^{(-)'} \otimes \frac{1}{\sqrt{n}}\mathbf{1}'_n \\ \vdots \\ V^{(-)'} \otimes H_{CA}^{(n-1)'} \end{pmatrix} \\ &= V^{(-)}D^{(-)}V^{(-)'} \otimes \frac{\tau_0}{n}\mathbf{1}_n\mathbf{1}'_n + V^{(-)}D^{(-)}V^{(-)'} \otimes \tau_1 H_{CA}^{(1)} H_{CA}^{(1)'} + \dots \\ &= Q \otimes H_A^{(+)} \text{diag}(\tau_j) H_A^{(+)' \end{aligned}$$

where $H_A^{(+)} = (\frac{1}{\sqrt{n}}\mathbf{1}_n | H_{CA})$ is an orthogonal matrix; τ_j is unknown and $H_A^{(+)} \text{diag}(\tau_j) H_A^{(+)'$ is a precision matrix; $Q = VD V' = V^{(-)}D^{(-)}V^{(-)'$ is a function of the neighborhood structure as defined in section ???. □

B MCMC algorithms

For the normal error case, we used the conditional distribution of η_0 as in HCSC. If $r_j = \frac{\eta_j}{\eta_0}$ and η_0 has a $\Gamma(\alpha, \beta)$ prior, it follows that

$$f(\eta_0 | \mathbf{r}, \mathbf{Y}) = \frac{\beta + \frac{1}{2}W(\mathbf{r})^\xi}{\Gamma(\xi)} \eta_0^{\xi-1} \exp(-\eta_0(\beta + W(\mathbf{r})/2)) \quad (1)$$

where $\xi = \frac{cn-M_1}{2} + \alpha$; $\mathbf{r} = \{r_j\}$; and $W(\mathbf{r}) = \mathbf{Y}'S^{-1}\mathbf{Y} - \mathbf{Y}'S^{-1}X(X'S^{-1}X)^{-1}X'S^{-1}\mathbf{Y}$, with $S^{-1} = \Gamma^{-1}/\eta_0$, which is a function of \mathbf{r} . The only change here compared to HCSC is that we normalized the columns of X_D (data cases) so each column of X_D has length 1. To do MCMC for both the normal and Poisson models, we first draw Θ 's components univariately using an M-H algorithm. For a gamma prior on τ_j , τ_j 's full conditional posterior is a gamma, so we draw τ_j using a Gibbs sampler. For normal data, an additional step is to draw η_0 by Gibbs sampler using (1). For Poisson data, a transformation of Θ was used in the chain, so the new parameter Ξ was drawn first, then Θ was calculated from Ξ ; Appendix C gives details. Since simulation steps for the normal and Poisson data are similar, we only list them for normal data. For current draws θ_i , τ_j and η_0 :

- for $i = 1, \dots, 60$ (Nn):
 - Generate $\theta_i^* \sim N(\theta_i, 4)$
 - Calculate the MH ratio as $\frac{f(\theta_i^*|\mathbf{Y},\boldsymbol{\tau})}{f(\theta_i|\mathbf{Y},\boldsymbol{\tau})}$; accept θ_i^* with probability $\min(1, \text{MH ratio})$.
- Generate $\tau_j^* \sim \Gamma(\alpha + \frac{19}{2}, \beta + \frac{1}{2}\sum_{i=1}^{19} D_{ii}^{(-)} \theta_{(i+3+19 \times j)}^2)$, where $j = 0, 1, 2$
- Sample η_0^* from the gamma distribution in (1).

All computations were carried out on a dual-core 2.20GHz AMD Athlon processor with 1.93GB of RAM physical address extension. MCMC code for SANOVA models was written in R; MCAR models were analyzed in WinBUGS. For each model we ran three parallel chains for 10,000 iterations and discarded the first 2,000 iterations as “burn-in”. The combined draws of the three chains are used for posterior summaries. Trace plots showed good convergence for all the models.

C Mean-structure transformation for the Poisson case

The Poisson model in the Θ parameterization has high posterior correlations among the θ s, which leads to slow MCMC mixing. To see this, let $\tilde{y}_{ij} = \log y_{ij} - \log E_{ij}$; then from the normal approximation to the Poisson,

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{y}}_{Nn} \\ \mathbf{0}_{Nn} \end{bmatrix} \approx X\Theta + \mathbf{e} = \begin{bmatrix} X_D \\ I_{Nn} \end{bmatrix} \Theta + \mathbf{e} \quad (2)$$

where $\mathbf{e} \sim N(\mathbf{0}, \Gamma)$ with

$$\Gamma = \begin{bmatrix} \text{diag}(1/y_i) & 0 & 0 \\ 0 & 10^6 I_n & 0 \\ 0 & 0 & \text{diag} \frac{1}{\tau_{D^{(-)}}} \end{bmatrix} \quad (\text{see section ??}).$$

As a result,

$$\begin{aligned} \text{cov}(\Theta | \tilde{\mathbf{Y}}, \Gamma) &\approx (X'(\text{cov}(\tilde{\mathbf{Y}}))^{-1}X)^{-1} \\ &= \left[(X'_D | I_{Nn}) \Gamma^{-1} \begin{pmatrix} X_D \\ I_{Nn} \end{pmatrix} \right]^{-1} \\ &= [X'_D \text{diag}(y_i) X_D + A^{-1}]^{-1} \\ &\equiv \Upsilon, \end{aligned} \quad (3)$$

where

$$A = \begin{bmatrix} 10^6 I_n & 0 \\ 0 & \text{diag} \frac{1}{\tau_{D^{(-)}}} \end{bmatrix}.$$

Although A is diagonal, $X'_D \text{diag}(y_i) X_D$ is not. The approximate data-case precision matrix $\text{diag}(y_i)$ amplifies the off-diagonal entries of Υ substantially for the Minnesota cancer data because E_i and y_i range over a few orders of magnitude, so the posterior correlations among the θ s are large. To minimize this effect, we re-parameterized as follows. Let $\Delta = (\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_p)'$, where \mathbf{e}_i is the i^{th} eigenvector of Υ for a particular choice of τ ; we discuss this below. Therefore, Δ is a $p \times p$ matrix, where $p = \text{dim}(\Theta) = Nn$ in the Poisson simulations. By the spectral decomposition, if $\lambda_1, \dots, \lambda_p$ are the eigenvalues corresponding to $\mathbf{e}_1, \dots, \mathbf{e}_p$,

$$\begin{aligned} \text{Cov}(\Theta | \tilde{\mathbf{Y}}, \Gamma) &\approx \Upsilon \\ &= \Delta' \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \Delta. \end{aligned}$$

Let $\Xi = \Delta \Theta$; then

$$\tilde{\mathbf{Y}} \approx X \Theta + \mathbf{e} = X \Delta' \Delta \Theta + \mathbf{e} = X^* \Xi + \mathbf{e}. \quad (4)$$

From (3),

$$\begin{aligned}
Cov(\Xi|Y, \Theta, \Gamma) &= \Delta cov(\Theta|Y, \Gamma) \Delta' \\
&\approx \Delta \Upsilon \Delta' \\
&= \Delta \Delta' \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \Delta \Delta' \\
&= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix}.
\end{aligned}$$

We can then sample Ξ instead of Θ . Ξ and Θ are one-to-one, so a draw of Ξ gives a draw of Θ . Since A is a function of τ , Δ is also a function of τ . Any MCMC routine will update τ frequently, so it is not obvious how to specify a general Δ to use in this transformation. Fortunately, τ had only a small effect on off-diagonal entries of Υ in our datasets, so we used $\tau = (100, 100, 100)$ to calculate a value of Δ which was then fixed and used in the MCMC. With this τ , the conditional posterior correlations among the Ξ s were mostly below 0.1, which is good enough to solve the mixing problem.

D Estimating $H_A^{(+)}$ from MCAR1

We used the posterior median of Ω , $\hat{\Omega}$, which had spectral decomposition

$$\begin{aligned}
\hat{\Omega} &= \begin{bmatrix} 11.85 & -2.21 & -5.05 \\ -2.21 & 5.24 & -2.09 \\ -5.05 & -2.09 & 6.88 \end{bmatrix} \\
&= \begin{pmatrix} 0.86 & 0.26 & -0.45 \\ -0.08 & -0.78 & -0.62 \\ -0.51 & 0.57 & -0.65 \end{pmatrix} \begin{bmatrix} 15.06 & 0 & 0 \\ 0 & 7.47 & 0 \\ 0 & 0 & 1.43 \end{bmatrix} \begin{pmatrix} 0.86 & -0.08 & -0.51 \\ 0.26 & -0.78 & 0.57 \\ -0.45 & -0.62 & -0.65 \end{pmatrix}. \tag{5}
\end{aligned}$$

Let Λ be the eigenvector matrix of $\hat{\Omega}$, the left-most matrix in the second row of (5). Then in the design matrix in (10), for the data cases (X_D) replace $H_A^{(+)} \otimes V^{(-)}$ with $\Lambda \otimes V^{(-)}$. Unlike HA_1 and HA_2 , this specification of $H_A^{(+)}$ has no county main effect.