Submitted to the Annals of Applied Statistics
arXiv: math.PR/0000000

MODELING TEMPORAL GRADIENTS IN REGIONALLY AGGREGATED CALIFORNIA ASTHMA HOSPITALIZATION DATA

By Harrison Quick, Sudipto Banerjee and Bradley P. Carlin

Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota 55455

APPENDIX A: DATA PREPROCESSING ALGORITHM

Recall that in the asthma hospitalization data, days with between 1 and 4 asthma hospitalizations are censored for privacy issues. In order to obtain reasonable estimates of the missing, we use an algorithm similar to Besag's Iterated Conditional Modes (1986), but instead calculate iterated conditional *means*. Assuming daily hospitalization counts for the i^{th} county follow a Poisson distribution with rate λ_i, the algorithm depends only on the county's observed counts and a prior distribution for λ_i. This will be done for each county, thus all censored values for a given county will receive the same estimate. We also assume $\lambda_i, i = 1, \ldots, M$ are independent and identically distributed.

A.1. Prior Specification. Let $Y_i(t_j) \overset{ind}{\sim} Pois(\lambda_i)$ be the number of hospitalizations on day j in county i and $\mathbf{Y}_i = (Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_D))$ denote the vector of values from the i^{th} county, where D is the number of days in the study, and let $\lambda_i \sim Gam(\alpha, \beta)$. Simple algebra shows

$$\lambda_i \mid \mathbf{Y}_i \sim Gam\left(\sum_j Y_i(t_j) + \alpha, \left(n + \frac{1}{\beta}\right)^{-1}\right), \tag{1}$$

which has $E[\lambda_i \mid \mathbf{Y}_i] = \left(\sum_j Y_i(t_j) + \alpha\right)\left(n + \frac{1}{\beta}\right)^{-1}$.

Due to confidentiality, daily counts between 1 and 4 are censored, so in order to continue our analysis, we require an accurate estimated count for these days. Conditioning on this information, it is easily shown that for all censored observations in the i^{th} county,

$$\begin{aligned} E\left[Y_i(t_j) \mid \lambda_i, Y_i(t_j) \in [1,4]\right] &= \frac{\sum_{y=1}^{4} \frac{y e^{-\lambda_i}\lambda_i^y}{y!}}{\sum_{y=1}^{4} \frac{e^{-\lambda_i}\lambda_i^y}{y!}} \\ &= \frac{1 + \lambda_i + \frac{1}{2}\lambda_i^2 + \frac{1}{6}\lambda^3}{1 + \frac{1}{2}\lambda_i + \frac{1}{6}\lambda_i^2 + \frac{1}{24}\lambda_i^3} \end{aligned} \tag{2}$$

Naively, one would estimate $E\left[Y_i(t_j) \mid \lambda_i, Y_i(t_j) \in [1,4]\right] = \frac{4+1}{2} = 2.5$, that is, by putting a flat prior over the range. Thus, we wish to choose hyperparameters α and β such that $\lambda^* = E[\lambda] = \alpha\beta$ satisfies $E\left[Y_i(t_j) \mid \lambda^*, Y_i(t_j) \in [1,4]\right] = 2.5$ in (2). While this is difficult to do in closed form, with numerical methods one can easily find that $\lambda^* \approx 2.9$. Choosing a prior variance of 100 gives $\beta = \left(\frac{100}{2.9}\right)$ and $\alpha = \frac{2.9^2}{100}$.

A.2. The Algorithm. Now that we've found our prior distribution, we can outline how our estimation algorithm proceeds:

Step 1. Set $Y_i(t_j) = 2.5 \; \forall \; Y_i(t_j) \in [1,4]$

Step 2. Let $\hat{\lambda} = E[\lambda_i \mid \mathbf{Y}_i] = \left(\sum_j Y_i(t_j) + \alpha\right)\left(n + \frac{1}{\beta}\right)^{-1}$

Step 3. Let

$$Y_i^*(t_j) = \begin{cases} Y_i(t_j), & \text{if } Y_i(t_j) \notin [1,4] \\ E\left[Y_i(t_j) \mid \hat{\lambda}, Y_i(t_j) \in [1,4]\right] = \frac{1+\hat{\lambda}+\frac{1}{2}\hat{\lambda}^2+\frac{1}{6}\hat{\lambda}^3}{1+\frac{1}{2}\hat{\lambda}+\frac{1}{6}\hat{\lambda}^2+\frac{1}{24}\hat{\lambda}^3}, & \text{if } Y_i(t_j) \in [1,4] \end{cases}$$

Step 4. Calculate $r = \frac{\sum_j Y_i^*(t_j)}{\sum_j Y_i(t_j)}$ and let $Y_i(t_j) = Y_i^*(t_j), j = 1, 2, \ldots, D$. If $r \in [1-\delta, 1+\delta]$ for some prespecified error threshold δ (say, $\delta = 0.01$), then the algorithm is complete. Otherwise, return to Step 2 and repeat.

Convergence follows from Besag (1986) and can be demonstrated via simulation.

APPENDIX B: ALTERNATIVE MODELS

We would be remiss not to recognize the limitations of our separable model ("Model 1") and how, at least in theory, our constructive approach can obviate such limitations. The separable model imposes the same temporal correlation structure for every areal unit because each of the $v_i(t)$'s are assumed to follow independent Gaussian processes with the same correlation function for every region i. Letting each $v_i(t)$ have its *own* set of process parameters $\boldsymbol{\phi}_i$ yields

$$\textbf{Model 2}: \; K_Z(t,u) = A\tilde{R}(t,u;\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{N_s})A^T \text{ and } \Sigma_Z = (I_{N_t} \otimes A)\Sigma_v(I \otimes A^T)\,,$$

where $\tilde{R}(t,u;\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{N_s})$ is an $N_s \times N_s$ diagonal matrix with $\rho(t,u;\boldsymbol{\phi}_i)$ as the i-th diagonal element, A is a square-root matrix such that $AA^T = \sigma^2(D - \alpha W)^{-1}$, and Σ_v is the variance-covariance matrix for the $v_i(t_j)$'s, that is an $N_sN_t \times N_sN_t$ block matrix with $\tilde{R}(t_i,t_j;\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_{N_s})$ as the (i,j)-th block. Thus, we still assume that the elements of $\mathbf{v}(t)$ are independent Gaussian processes but are no longer identical. Now each region will be able to account for its own temporal correlation characteristics, including the smoothness of its temporal curve and the rate of decay of temporal correlation.

Model 2, while temporally non-stationary, still assumes the same spatial association structure across time. This, perhaps, is a less stringent limitation. It is not clear when one would envision a temporally evolving spatial association between the areal units, particularly for data with temporally short duration. However, if this were the case, then temporally evolving spatial associations can be modeled using $A(t)$ as a function of t. For instance, we could extend Model 2 to

$$\textbf{Model 3}: \; K_Z(t,u) = A(t)\tilde{R}(t,u;\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\ldots,\boldsymbol{\phi}_{N_s})A(u)^T \text{ and } \Sigma_Z = \tilde{A}\Sigma_v\tilde{A}^T \;,$$

where $A(t)$ is a square root of $\sigma^2(t)(D-\alpha(t)W)^{-1}$, $\sigma^2(t)$ and $\alpha(t)$ are continuous functions of time, and $\tilde{A}$ is an $N_sN_t \times N_sN_t$ block diagonal matrix with $A(t_i)$'s as its diagonal blocks. This raises the question of what type of function $\alpha(t)$ should be. Recall that $\alpha(t) \in (0,1)$ for all t. One could perhaps think of a polynomial-spline, say $\psi(t)$, on the logit-link of the smoothness parameter: $\log\left(\alpha(t)/(1-\alpha(t))\right) = \psi(t)$. Indeed $\psi(t)$ will likely have parameters so one will need to ascertain identifiability issues in such general settings. Even more generally, if one does not want to assume a CAR structure, we can keep the $A(t)$ as unknown functions and estimate them. For example, we could treat its diagonal elements as a log-Gaussian, and the lower-triangular elements are independent Gaussian processes.

While further modeling of these functions using stochastic processes is conceivable in Bayesian contexts, such models will be excessively complex, both for estimation and interpretation, and without much foreseeable inferential gains in most practical settings. This is especially true in our current context, where we are more interested in estimating gradients from models. Gradient analysis can be performed seamlessly once the posterior distribution for any of the above models has been estimated.

APPENDIX C: MCMC IMPLEMENTATION DETAILS

Let $D_\epsilon = diag\{\tau_i^2\}_{i=1}^M$ and $X = (X(t_1)^T, X(t_2)^T, \ldots, X(t_{N_t})^T)^T$ be our $(MN_t) \times p$ design matrix, where $X(t_j) = (x_1(t_j)^T, x_2(t_j)^T, \ldots, x_M(t_j)^T)^T$ denotes the $M \times p$ design matrix for the j^{th} timepoint. Recall $x_i(t_j)$ is the i^{th} region's $p \times 1$ column vector of covariates for the j^{th} timepoint. For notational convenience, we let $[\theta \mid \cdot]$ denote the distribution of any parameter θ given all other parameters and the data.

C.1. Gibbs update for $\boldsymbol{\beta}$.

The full conditional for $\boldsymbol{\beta}$ is

$$\begin{aligned} [\boldsymbol{\beta} \mid \cdot] &\propto N(\boldsymbol{\beta} \mid \boldsymbol{\mu}_\beta, \Sigma_\beta) \times N(\mathbf{Y} \mid X\boldsymbol{\beta} + \mathbf{Z}, I \otimes D_\epsilon) \\ &= N\left(\boldsymbol{\beta} \mid \boldsymbol{\mu}_{\beta\mid\cdot}, \Sigma_{\beta\mid\cdot}\right) \end{aligned} \tag{3}$$

where

$$\begin{aligned}\Sigma_{\beta\,|\,\cdot} &= [\Sigma_\beta^{-1} + X^T(I\otimes D_\epsilon)^{-1}X]^{-1} \\ &= [\Sigma_\beta^{-1} + X^T(I\otimes D_\epsilon^{-1})X]^{-1} \qquad (4)\\ \boldsymbol{\mu}_{\beta\,|\,\cdot} &= \Sigma_{\beta\,|\,\cdot}[\Sigma_\beta^{-1}\boldsymbol{\mu}_\beta + X^T(I\otimes D_\epsilon^{-1})(\mathbf{Y}-\mathbf{Z})] \qquad (5)\end{aligned}$$

If we take $\Sigma_\beta^{-1} = \mathbf{0}$ (i.e., a flat prior on $\boldsymbol{\beta}$), then (4) and (5) become

$$\begin{aligned}\Sigma_{\beta\,|\,\cdot} &= [X^T(I\otimes D_\epsilon^{-1})X]^{-1} \qquad (6)\\ \boldsymbol{\mu}_{\beta\,|\,\cdot} &= [X^T(I\otimes D_\epsilon^{-1})X]^{-1}X^T(I\otimes D_\epsilon^{-1})(\mathbf{Y}-\mathbf{Z}) \qquad (7)\end{aligned}$$

Due to computational complexities, the following algebraic simplifications are helpful

$$\begin{aligned}X^T(I\otimes D_\epsilon^{-1})X &= \sum_{j=1}^{N_t} X(t_j)^T D_\epsilon^{-1} X(t_j)\\ &= \sum_{j=1}^{N_t}\sum_{i=1}^{M} \frac{1}{\tau_i^2}\mathbf{x}_i(t_j)\mathbf{x}_i(t_j)^T\\ &= \sum_{i=1}^{M}\frac{1}{\tau_i^2}\sum_{j=1}^{N_t}\mathbf{x}_i(t_j)\mathbf{x}_i(t_j)^T\end{aligned}$$

and

$$\begin{aligned}X^T(I\otimes D_\epsilon^{-1})(\mathbf{Y}-\mathbf{Z}) &= \sum_{j=1}^{N_t} X(t_j)^T D_\epsilon^{-1}(\mathbf{Y}(t_j)-\mathbf{Z}(t_j))\\ &= \sum_{j=1}^{N_t}\sum_{i=1}^{M}\frac{1}{\tau_i^2}\mathbf{x}_i(t_j)(\mathbf{y}_i(t_j)-\mathbf{z}_i(t_j))^T\\ &= \sum_{i=1}^{M}\frac{1}{\tau_i^2}\sum_{j=1}^{N_t}\mathbf{x}_i(t_j)(\mathbf{y}_i(t_j)-\mathbf{z}_i(t_j))^T.\end{aligned}$$

C.2. Gibbs update for Z. Due to the potential for computational problems associated with manipulating an $(MN_t)\times(MN_t)$ matrix for large values of M or N_t, updates for $\mathbf{Z}$ are performed on $\widetilde{\mathbf{Z}} = (\mathbf{z}_1^T, \mathbf{z}_2^T, \ldots, \mathbf{z}_M^T)^T$. The full conditional for $\widetilde{\mathbf{Z}}$ is

$$[\widetilde{\mathbf{Z}}\,|\,\cdot] \propto N(\widetilde{\mathbf{Z}}\,|\,\mathbf{0}, (D-\alpha W)^{-1}\otimes\sigma^2 R(\phi)) \times \prod_{i=1}^{M} N(\mathbf{y}_i\,|\,X_i\boldsymbol{\beta}+\mathbf{z}_i, \tau_i^2 I) \qquad (8)$$

thus, the full conditionals for each $\mathbf{z}_i, i = 1,2,\ldots,M$ are

$$[\mathbf{z}_i\,|\,\cdot] \propto N\left(\mathbf{z}_i\,|\,\boldsymbol{\mu}_{\mathbf{z}_i\,|}, \Sigma_{\mathbf{z}_i\,|}\right) \qquad (9)$$

where

$$\Sigma_{\mathbf{z}_i\,|\,\cdot} = \left[\frac{R(\phi)^{-1}}{\frac{\sigma^2}{m_i}} + \frac{I}{\tau_i^2}\right]^{-1}$$

and

$$\boldsymbol{\mu}_{\mathbf{z}_i\,|\,\cdot} = \Sigma_{\mathbf{z}_i\,|\,\cdot}\left[\frac{\alpha}{m_i}\sum_{j\sim i} w_{ij}\left(\frac{R(\phi)^{-1}}{\frac{\sigma^2}{m_i}}\right)\mathbf{z}_j + \frac{(\widetilde{\mathbf{y}}_i - X_i\boldsymbol{\beta})}{\tau_i^2}\right]$$

C.3. Gibbs update for τ_i^2. The full conditional for τ_i^2 is

$$\begin{aligned}
[\tau_i^2\,|\,\cdot] &\propto IG(\tau_i^2\,|\,a_\tau, b_\tau) \times \prod_{j=1}^{N_t} N(\mathbf{Y}(t_j)\,|\,X(t_j)\boldsymbol{\beta} + \mathbf{Z}(t_j), D_\epsilon) \\
&= IG(\tau_i^2\,|\,a_{\tau\,|\,\cdot}, b_{\tau\,|\,\cdot}) \qquad (10)
\end{aligned}$$

where $a_{\tau\,|\,\cdot} = a_\tau + \frac{N_t}{2}$ and $b_{\tau\,|\,\cdot} = b_\tau + \frac{1}{2}\sum_{j=1} N_t(y_i(t_j) - \mathbf{x}_i(t_j)'\boldsymbol{\beta} - z_i(t_j))^2$.

C.4. Gibbs update for σ^2. The full conditional for σ^2 is

$$\begin{aligned}
[\sigma^2\,|\,\cdot] &\propto IG(\sigma^2\,|\,a_\sigma, b_\sigma) \times N(\mathbf{Z}\,|\,\mathbf{0}, \sigma^2 R(\phi) \otimes (D - \alpha W)^{-1}) \\
&= IG(\sigma^2\,|\,a_{\sigma\,|\,\cdot}, b_{\sigma\,|\,\cdot}) \qquad (11)
\end{aligned}$$

where $a_{\sigma\,|\,\cdot} = a_\sigma + \frac{MN_t}{2}$,

$$\begin{aligned}
b_{\sigma\,|\,\cdot} &= b_\sigma + \frac{\mathbf{Z}^T[R(\phi)^{-1} \otimes (D - \alpha W)]\mathbf{Z}}{2} \\
&= b_\sigma + \frac{1}{2}\sum_{i=1}^{N_t}\sum_{j=1}^{N_t} r^{ij}(\phi)\mathbf{Z}(t_i)^T(D - \alpha W)\mathbf{Z}(t_j),
\end{aligned}$$

and $r^{ij}(\phi)$ is the $(ij)^{th}$ element in $R(\phi)^{-1}$.

C.5. Metropolis update for ϕ. The full conditional for ϕ is

$$[\phi\,|\,\cdot] \propto U(\phi\,|\,a_\phi, b_\phi) \times N(\mathbf{Z}\,|\,\mathbf{0}, \sigma^2 R(\phi) \otimes (D - \alpha W)^{-1}). \qquad (12)$$

To generate this using a Metropolis step, we transform to

$$\psi = \text{logit}\left[\frac{\phi - a_\phi}{b_\phi - \phi}\right]$$

and use a normal random walk proposal. Here, we let $a_\phi = 3/(N_t - 1)$ and $b_\phi = 10$. These were chosen such that $3/\phi \in (0, N_t - 1)$, based on the effective temporal range of an exponential covariance function.

C.6. Metropolis update for α. The full conditional for α is

$$[\alpha \,|\, \cdot] \propto Beta(\alpha \,|\, a_\alpha, b_\alpha) \times N(\mathbf{Z} \,|\, \mathbf{0}, \sigma^2 R(\phi) \otimes (D - \alpha W)^{-1}). \tag{13}$$

To generate this using a Metropolis step, we transform to

$$\gamma = \text{logit}\left[\frac{\alpha}{1-\alpha}\right]$$

and use a normal random walk proposal. Because we desired a moderately informative prior on α such that its expected value was 0.9, we chose to set $a_\alpha = 9$ and $b_\alpha = 1$.

APPENDIX D: COMPARISON TO DISCRETE-TIME MODEL

In settings where inferential interest is desired only at the resolution at which the data are collected, it is true that several of the discrete-time models in the literature will suffice in our setting. In order to compare such models to our continuous-time Gaussian process model, we must restrict our attention to a monthly level analysis and compute finite differences, defined as

$$\frac{Z_i(t_{j+1}) - Z_i(t_j)}{t_{j+1} - t_j}.$$

For this comparison, for simplicity we limit our focus to the first 7 years of data (i.e., $N_t = 84$) and chose to model the spatiotemporal random effects in our discrete-time model using a first-order autoregressive (AR(1)) process, such that

$$\mathbf{Z}(t_j) = \rho \mathbf{Z}(t_{j-1}) + \mathbf{w}(t_j), j = 1, \ldots, N_t, \tag{14}$$

where $\mathbf{Z}_i(0) = 0$ for all i and $\mathbf{w}(t_j) \sim CAR(\gamma_j)$.

Based on both model fit diagnostics, our Gaussian process model outperforms the AR(1) model, with a difference in DIC of 597 and a difference in D-S scoring rule of 2358. Motivated by the investigation of Los Angeles County in our paper, we focus our analysis here as well. While both models fit the data (open circles) well, Figure 1 reveals the AR(1) model fails to capture the large peaks in December of three of the years (1991, 1993, and 1995). This appears to be due to the model limiting the impact of the random effects; while the curve in the right-middle panel of Figure 1 has the same peaks and valleys as its Gaussian process counterpart, its variability is substantially diminished. This lack of variability is further illustrated in the temporal changes (gradients or finite differences) in the bottom panel of the figure.

While the results shown here may be due to our choice of discrete-time model, we believe that, in general, results from our model will be consistent with those from discrete-time models when

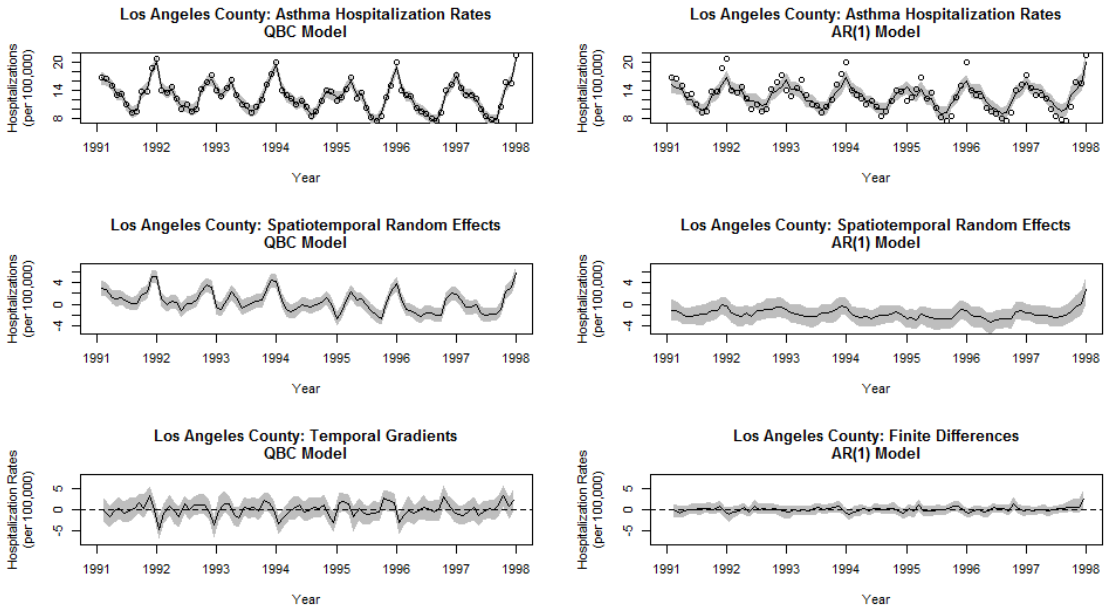

Fig 1. *Comparison of our Gaussian process model and an AR(1) model. The plots in the top panel show the fitted asthma hospitalization rates (with 95% CI bounds) plotted against the observed rates. The middle panel displays the temporal profiles for the spatiotemporal random effects from each model. In the bottom panel, we compare the predicted temporal gradients computed by our model to the finite differences between adjacent timepoints for the spatiotemporal random effects.*

inferential interest is limited to the resolution of the data. In addition, the flexibility of our model also permits inference at a finer resolution directly from the posterior predictive distribution, which discrete-time models will be unable to do without prespecifying missing values at unobserved time points prior to model fitting.

Harrison Quick
Sudipto Banerjee
Bradley P. Carlin
Division of Biostatistics
University of Minnesota
Minneapolis, MN 55425
E-mail: quic0038@umn.com