Harvard University
Harvard University Biostatistics Working Paper Series

Year 2006 *Paper* 61

Spatio-Temporal Analysis of Areal Data and Discovery of Neighborhood Relationships in Conditionally Autoregressive Models

Subharup Guha* Louise Ryan†

*Harvard University, sguha@hsph.harvard.edu
†Harvard School of Public Health and Dana-Farber Cancer Institute, lryan@hsph.harvard.edu

SPATIO-TEMPORAL ANALYSIS OF AREAL DATA AND DISCOVERY OF NEIGHBORHOOD RELATIONSHIPS IN CONDITIONALLY AUTOREGRESSIVE MODELS

Subharup Guha[1] and Louise Ryan[2]

ABSTRACT. Outcome data in disease mapping problems consist of counts or averages observed within areal units belonging to a geographical region. Conditionally autoregressive (CAR) models are often used to analyze these data. Although they are computationally convenient, most CAR models implementations rely on an assumed neighborhood structure, and account for spatio-temporal association through area-specific random effects that do not interact with temporal trends.

The separability assumption and the ad-hoc definition of the neighborhood structure may not be valid in practice. We propose a computationally feasible Bayesian approach that simultaneously investigates the spatio-temporal variability in the area-specific random effects and the latent neighborhood structure. Based on the concept of generalized distances, the neighborhood structure is modeled by a continuous-time Bernoulli process having Markovian dependence. Although areas separated by large generalized distances have small prior probabilities of being neighbors, no assumption is made about the particular form of the relationship, e.g. K nearest neighbors, sharing a common boundary. We describe an MCMC procedure for simulation-based inference. The proposed model is applied to analyze heart disease incidence rates in the Sydney Metropolitan Area of Australia, where we discover complex temporal trends in the random effects associated with the postcodes. The estimated neighborhood structure also differs considerably from that assumed

[1]Postdoctoral Research Fellow, Department of Biostatistics, Harvard School of Public Health.
[2]Chair and Professor of Biostatistics, Harvard School of Public Health.

in typical CAR model applications.

KEYWORDS. Adjacency matrix, disease mapping, epidemiology, Markov processes.

SHORT TITLE. CAR model extensions.

AMS CLASSIFICATION: Primary: 62M40; Secondary: 62M05.

1. Introduction. Outcome data in disease mapping problems consist of counts or averages observed in a geographical area having n units (e.g. postcodes, counties) with well-defined boundaries. Imagine that data on these areal units are observed repeatedly over a number of time periods $t_1, \ldots, t_T$. Denote the outcomes by Y_{it}, where area $i \in \mathcal{D} = \{1, \ldots, n\}$ and time $t \in \mathcal{T} = \{t_1, \ldots, t_T\}$. Additional information is available as covariates measured on each areal units which may be time-dependent. Often the data, possibly after some transformation, can be modeled as normal random variables. We would then assume $Y_{it} \stackrel{ind}{\sim} N(\eta_{it}, \sigma^2)$, with the *linear predictor* η_{it} defined as

$$\eta_{it} = o_{it} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_f + \boldsymbol{z}'_{it}\boldsymbol{\beta}_r + \theta_i. \tag{1}$$

In the above expression, o_{it} is a known (possibly zero) offset, $\boldsymbol{\beta}_f$ is the vector of fixed effects and $\boldsymbol{\beta}_r$ are random effects. Column vectors $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{it}$ denote the observed covariates and the area-specific random effects $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ capture the spatial variability.

When the data consist of disease or mortality counts for a rare disease, as in disease mapping problems, we might assume the Poisson regression model: $Y_{it} \sim Po(e^{\eta_{it}})$, with the offset o_{it} in (1) set equal to the log population at risk in area i at time t. Alternatively, the data may represent binary outcomes (presence or absence of a particular condition) or continuous measurements for which the normal assumption does not hold even after transformation.

Generalized linear mixed models (GLMMs) provide an elegant framework for modeling these data (Zeger and Karim, 1991). The likelihood function of a GLMM belongs to the exponential family: $Y_{it} \mid \omega_{it}, \varsigma \overset{ind}{\sim} h(Y_{it}, \varsigma) \cdot \exp\left\{(Y_{it}\,\omega_{it} - b(\omega_{it}))\,/a(\varsigma)\right\}$, where ς is a dispersion parameter and the conditional expectation $E[Y_{it} \mid \omega_{it}, \varsigma]$ equals $\mu_{it} = b'(\omega_{it})$ (McCullagh and Nelder, 1999, pp. 28). The conditional variance $Var[Y_{it} \mid \omega_{it}, \varsigma] = \Upsilon(\omega_{it})\,a(\varsigma)$, with the *variance function* $\Upsilon(\omega_{it})$ defined as $b''(\omega_{it})$. For an appropriate link function $g(\cdot)$, the linear predictor η_{it} is related to the mean μ_{it} as $\eta_{it} = g(\mu_{it})$. The likelihood function may be expressed as a function of η_{it} and the dispersion parameter ς:

$$Y_{it} \mid \eta_{it}, \varsigma \overset{ind}{\sim} h(Y_{it}, \varsigma) \cdot \exp\left\{(Y_{it}\,\omega(\eta_{it}) - b(\eta_{it})\,/a(\varsigma)\right\} \qquad (2)$$

The normal model is a special case with identity link and $\varsigma = \sigma^2$. Poisson regression corresponds to the log link and dispersion parameter $\varsigma = 1$. Logistic regression corresponds to a Bernoulli likelihood, logit link and dispersion parameter $\varsigma = 1$ (McCullagh and Nelder, 1999, p. 30). In Bayesian analyses, a normal prior is typically assumed for the GLMM fixed effects: $\boldsymbol{\beta}_r \sim N_p(\mu_\beta, \Sigma_\beta)$. A prior for the precision matrix of the random effects is $D^{-1} \sim Wishart\,(d_0, R_0)$, where the positive definite matrix R_0 is of order q and $d_0 \geq q$.

To achieve spatial smoothing, the region-specific random effects can be modeled using a conditionally autoregressive (CAR) structure (Besag, 1974; Clayton and Kaldor, 1987; Cressie and Chan, 1989; Besag et al., 1991; Bernardinelli et al., 1995; Waller et al., 1997). Alterative approaches include those of Wakefield and Morris (1999) and Banerjee et al. (2003), who propose a multivariate normal distribution for the random effects with the correlations determined by the intercentroidal distances. Böhning et al. (1999) propose using a discrete Poisson mixture. Knorr-Held and Rasser (2000) and Giudici et al. (2000) partition the geographical domain into a random number of clusters with constant relative risk.

CAR models are computationally convenient for analyzing areal data. The commonly used CAR formulation of Besag et al. (1991) relies on the assumption of a *neighborhood structure* among the areal units (refer to Banerjee, Carlin and Gelfand, 2004, pp. 79). An n by n matrix $\boldsymbol{W} = (w_{ij})$ called the *adjacency matrix* summarizes the neighborhood relationships between the areas. For example,

$$w_{ij} = \begin{cases} 1 & \text{if the areas } i \text{ and } j \text{ share a common boundary, } i \neq j \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Alternatively, we might set w_{ij} equal to a decreasing function (e.g. inverse) of the intercentroidal distance between the areas, or let $w_{ij} = 1$ if the intercentroidal distances fall within a certain threshold. Another possibility is to set $w_{ij} = 1$ for the K nearest neighbors of each area.

Given the adjacency matrix $\boldsymbol{W}$, the CAR model assumes that the distribution of θ_i conditional on the remaining set of random effects $\boldsymbol{\theta}_{-i}$ depends only on the neighbors of the area i. In particular, $\theta_i \mid \boldsymbol{\theta}_{-i} \sim N(\sum_j w_{ij}\, \theta_j / w_{i+}, \sigma^2 / w_{i+})$, where $w_{i+} = \sum_j w_{ij}$ is the total number of neighbors of area i. We denote the model as $CAR(\boldsymbol{W}, \sigma^{-2})$. On applying Brook's Lemma (Brook, 1964), it can be shown that the full conditionals induce the joint distribution

$$[\boldsymbol{\theta} \mid \boldsymbol{W}] \propto \exp\left\{ -\frac{1}{2\sigma^2} \boldsymbol{\theta}' \left(\boldsymbol{D} - \boldsymbol{W}\right) \boldsymbol{\theta} \right\} = \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i<j} w_{ij} (\theta_i - \theta_j)^2 \right\}$$

where $\boldsymbol{D}$ is the diagonal matrix with elements w_{i+}. The precision matrix $\boldsymbol{\Sigma}_\theta^{-1} = \boldsymbol{D} - \boldsymbol{W}$ is singular and so the above joint distribution is improper. Since it is a prior on the unobserved random effects and not a stochastic mechanism for the data, there are no conceptual problems for a Bayesian analysis if the posterior is proper. However, it is often of interest to use a proper CAR model. Several modifications in the literature remedy the prior impropriety, including Besag et al. (1991), Carlin and Banerjee (2003), Cressie (1991) and Leroux, Lei and Breslow (1998). The proper CAR model of Leroux et al. (1998) relies on the variance components $\boldsymbol{\nu} = (\sigma^{-2}, \lambda)$ and assumes that

$$\sigma^2 \boldsymbol{\Sigma}_\theta^{-1} = (1-\lambda)\boldsymbol{I} + \lambda \boldsymbol{W}^*, \tag{4}$$

where $0 \leq \lambda \leq 1$. When $\lambda = 1$, we obtain the improper CAR model in (4). The precision matrix $\boldsymbol{\Sigma}_\theta^{-1}$ is non-singular when $\lambda \neq 1$. When $\lambda = 0$, we obtain the independence model with no spatial structure.

We generically denote improper and proper CAR models by $\boldsymbol{\theta} \sim CAR(\boldsymbol{W}, \boldsymbol{\nu})$. For a proper CAR model, the vector $\boldsymbol{\nu}$ may include additional variance components besides σ^2. For example, for the CAR model of Leroux et al. (1998), $\boldsymbol{\nu} = (\sigma^2, \lambda)$.

Spatio-temporal modeling. In many applications, spatio-temporal analyses of areal data include independent temporal terms (day-of-week effects or more sophisticated time trends) as the fixed/random effects in the linear predictor (1). This approach assumes separability of the spatial and temporal components, because the residual spatial variability represented by $\boldsymbol{\theta}$ is assumed to be unchanged over time. The assumption may not be reasonable in practice, and we may wish to introduce time dependent spatial random effects θ_{it_m}, where $i = 1, \ldots, n$ and $m = 1, \ldots, T$. Health surveillance studies, where the random effects are important for detecting and monitoring infectious disease outbreaks, are some important real life examples.

Kottas, Duan and Gelfand (2006) model the region-specific random effects using a dynamic formulation of the spatial Dirichlet process (Gelfand, Kottas and MacEachern, 2005). Within the CAR framework, Banerjee et al. (2004, pp. 284) assume $(\theta_{1t_m}, \ldots, \theta_{nt_m}) \stackrel{ind}{\sim} CAR(\boldsymbol{W}, \sigma^2_{t_m})$, where $\sigma^{-2}_{t_m} \stackrel{ind}{\sim}$ gamma(c, d). In the aforementioned CAR model-based approaches, the adjacency matrix $\boldsymbol{W}$ is assumed to be time-invariant and is defined in an ad-hoc manner. There has been relatively little work on discovering the latent neighborhood structure, even though inferences on the remaining model parameters rely on the neighborhood assumption. Moreover, since they represent similar areas that spatially borrow strength, study of the neighborhood relationships and their temporal variation can be useful for effective detection and control of disease. In investigations where it is not directly relevant, the underlying structure can be integrated out to estimate the parameters of interest.

Recently, Ma, Carlin and Banerjee (2006) have applied ideas related to statistical social analysis (Wang and Wong, 1987; Hoff et al., 2002) to model the neighborhood structure as $w_{ij}|p_{ij} \sim$ Bernoulli(p_{ij}), where $p_{ij} = \text{logit}(\boldsymbol{r}'_{ij}\boldsymbol{\phi})$ and $\boldsymbol{r}_{ij}$ is a vector of covariates believed to influence the neighborhood relationship.

Section 2 of this paper proposes a computationally feasible approach for the simultaneously modeling of the spatio-temporal variability in the region-specific random effects and the latent neighborhood structure. The technique is motivated by the concept of generalized distances introduced in Section 2.1. Section 2.2 mentions desirable properties of a reasonable spatio-temporal model for the adjacency indicators. Section 2.3 develops the basic theoretical framework. Section 2.4 applies the theoretical results to the disease mapping problem, and shows that the proposed model satisfies the theoretical requirements listed in Section 2.2. Section 2.5 specifies the Bayesian hierarchical model. An MCMC procedure is described in Section 3. In Section 4, the inference procedure is applied to analyze data on emergency room visits due to ischemic heart disease.

2. A model for time dependent area-specific random effects.

Let $\boldsymbol{\theta} = (\theta_{1t_1}, \ldots, \theta_{1t_T}, \ldots, \theta_{nt_1}, \ldots, \theta_{nt_T})$ be the vector of time dependent region-specific random effects. Let $\boldsymbol{W}^*$ be the random adjacency matrix of dimension nT corresponding to $\boldsymbol{\theta}$, so that

$$\boldsymbol{\theta} | \boldsymbol{W}^*, \boldsymbol{\nu} \sim CAR(\boldsymbol{W}^*, \boldsymbol{\nu}), \tag{5}$$

where, as before, vector $\boldsymbol{\nu}$ denotes the variance components associated with a proper CAR model. The adjacency matrix is symmetric because of the symmetry in the neighborhood relationships.

Given an area $i \in \mathcal{D}$ and time $t \in \{t_1, \ldots, t_m\}$, the *temporal neighborhood* $\mathbb{T}(i,t)$ is assumed to be a subset of $\mathcal{D} \times \mathcal{T}$, and the *spatial neighborhood* $\mathbb{S}(i,t)$ is taken to be a *random* subset of $\mathcal{D}$. The temporal and spatial neighborhoods jointly determine how the region-specific random effects mutually borrow strength. Specifically, the conditional distribution of θ_{it} depends on $\boldsymbol{\theta} - \{\theta_{it}\}$ only through the subset $\big\{\theta_{jt'} \mid (j,t') \in \mathbb{T}(i,t)\big\} \cup \big\{\theta_{jt'} \mid j \in \mathbb{S}(i,t) \text{ and } t' = t\big\}$.

Because of the one-dimensional nature of time, the temporal neighborhoods can be defined in a straightforward manner and may be assumed to be known. In order to achieve sparsity in the adjacency matrix $\boldsymbol{W}^*$, we assume that at time t_1, $\mathbb{T}(i,t_1) = \{(i,t_2)\}$ for $i \in \mathcal{D}$. This implies that the conditional distribution of the random effect θ_{it_1} depends only on θ_{it_2} and the spatial neighbors $\{\theta_{jt_1} \mid j \in \mathbb{S}(i,t_1)\}$ at time t_1. Similarly, when $m > 1$, we assume that $\mathbb{T}(i,t_m) = \{(i,t_{m-1}), (i,t_{m+1})\}$, implying that the conditional distribution of θ_{it_m} depends on $\theta_{it_{m-1}}$, $\theta_{it_{m+1}}$ and the spatial neighbors $\{\theta_{jt_m} \mid j \in \mathbb{S}(i,t_m)\}$ at time t_m.

There is considerable amount of flexibility in the definition of the spatial neighborhoods, and ad-hoc neighborhood structures like (3) are often used in practice. It is this component of the neighborhood structure that we regard as unknown and therefore random. It is convenient to represent the spatial neighborhoods at time t as the random adjacency matrix $\boldsymbol{W}_t = (W_{ijt})$, where

$$W_{ijt} = \begin{cases} 1 & \text{if } (j,t) \in \mathbb{S}(i,t) \\ \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The proposed model for the adjacency indicators $\{W_{ijt} : t \geq 0, \text{ fixed } i, j \in \mathcal{D}\}$ involves the concept of "generalized distance", which is defined below.

2.1. Generalized distance. Two areal units are considered to be "neighbors" if they are similar in some manner. Physical proximity is an important factor, although other factors like similarity in socioeconomic status, population densities, etc., may also play an important role. In the latter case, the neighborhood status may change as the demographic characteristics of the areas change over time. For example, two areal units that share a common boundary may initially have very similar patterns of demographic covariates and hence be regarded as neighbors. Subsequently, if the living conditions in one area changes drastically relative to the other (for example, if there is a flu outbreak, or a rapid decline in the standard of living due to the construction of a new highway), the areas may no longer be neighbors in the general sense because they have ceased to resemble each other with respect to the demographic characteristics.

For $i = 1, \ldots, n$ and time $t \in \mathcal{T}$, let $\boldsymbol{r}_{it} = (r_{it1}, \ldots, r_{itp})$ denote the set of p covariates that influence the neighborhood relationship. The (time-invariant) coordinates of the centroids of the areas, or some other measure of the position of the areal units, are typically included in the vector $\boldsymbol{r}_{it}$. For continuous covariates, the *generalized distance* Δ_{ijt} between the areas i and j at time t may be defined as

$$\Delta_{ijt} = \sum_{m=1}^{p} \phi_m |r_{imt} - r_{jmt}| \quad (7)$$

for some positive constants $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$. Another option in the case of continuous covariates is the Mahalanobis distance, $\left(\sum_{m=1}^{p} \phi_m \left(r_{imt} - r_{jmt}\right)^2\right)^{1/2}$. If some of the covariates are discrete, we may use a different distance measure (e.g. Manhattan distance) and combine the distances corresponding to the different types of covariates to obtained the overall generalized distances. For times $t \geq 0$, the generalized distances are positive and uniformly bounded away from zero if the intercentroidal distances are included in (7).

2.2. Properties of a reasonable model. For any given pair of areas i and j, we model the adjacency indicators $\{W_{ijt} : t \geq 0\}$ as time-dependent Bernoulli random variables. Due to the symmetry of the neighborhood relationships, we have $w_{ijt} = w_{jit}$. The adjacency matrix $\boldsymbol{W}^*$ corresponding to $\boldsymbol{\theta}$ therefore consists of $n(n-1)/2 \cdot T$ free parameters. We expect the adjacency indicators to possess the following distributional properties:

(i) The neighborhood relationships should evolve "smoothly" over time, in the sense that areas that are (not) neighbors at time t tend (not) to be neighbors at time $t + \triangle t$ when the difference $\triangle t$ is small. This is formally expressed by stating that for any given distinct times $0 \leq t < s$ and pairs of areas i and j, we expect that $\lim_{s \downarrow t} \Pr\left(W_{ijs} = w_2 \mid W_{ijt} = w_1\right)$ equals 1 when $w_1 = w_2$, and equals 0 if $w_1 \neq w_2$, for all $w_1, w_2 \in \{0, 1\}$.

(ii) Information about the value of W_{ijt} should convey progressively little information about W_{ijs} as the difference $s - t$ increases.

(iii) For a given pair of areas i and j, suppose that that the adjacency variable W_{ijt} is marginally distributed as $\boldsymbol{\rho}^{(ij)}(t)$ so that $W_{ijt} = w$ with probability $\rho_w^{(ij)}(t)$, where $w = 0, 1$. The probability $\rho_1^{(ij)}(t)$ should be a decreasing, continuous function of the generalized distance Δ_{ijt}.

Section 2.3 introduces a general class of binary valued non-homogenous Markov processes that satisfy Property *(i)–(ii)*. Section 2.4 applies the processes to disease mapping problems and demonstrates that Property *(iii)* is satisfied.

2.3. A non-homogeneous Markov process for adjacency indicators. Let $\{U_t : t \geq 0\}$ be a family of random variables taking values in the finite state space $\mathcal{S} = \{0, 1\}$. The process $\{U_t : t \geq 0\}$ is a Markov process if

$$\Pr\left(U_{t_n} = w \mid U_{t_1} = w_1, \ldots, U_{t_{n-1}} = w_{n-1}\right) = \Pr\left(U_{t_n} = w \mid U_{t_{n-1}} = w_{n-1}\right)$$

for all $w, w_1, \ldots, w_{n-1} \in \mathcal{S}$ and any sequence of times $t_1 < \ldots < t_n$. For any given pair of times $0 \leq t \leq s$, the transition probability matrix $\boldsymbol{P}(t, s)$ has the typical element

$$\boldsymbol{P}_{w_1, w_2}(t, s) = \Pr\left(U_s = w_2 \mid U_t = w_1\right), \quad w_1, w_2 \in \mathcal{S}.$$

A set of transition matrices $\{\boldsymbol{P}(t, s) : 0 \leq t \leq s\}$ satisfies the following properties: *(i)* $\boldsymbol{P}(t, s)$ should be row-stochastic, *(ii)* $\boldsymbol{P}(t, t) = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix of order K, and *(iii)* the matrices should satisfy the Chapman-Kolmogorov equation, $\boldsymbol{P}(t, s) = \boldsymbol{P}(t, u) \cdot \boldsymbol{P}(u, s)$, for all $0 \leq t \leq u \leq s$. A homogenous Markov process is one for which $\boldsymbol{P}(t, s) = \boldsymbol{P}(0, s - t)$ for all $0 \leq t \leq s$. Grimmett and Stirzaker (2001, pp. 256) discusses homogenous Markov processes in detail. Relaxing the homogeneity assumption can lead to some very difficult problems (e.g. refer to Grimmett and Stirzaker, 2001, pp. 274–281).

Let $\lambda_0(t)$ and $\lambda_1(t)$ be two positive, continuous functions such that $\lim_{t\to 0} \lambda_w(t) < \infty$ for $w \in \mathcal{S}$. For $w_1, w_2 \in \mathcal{S}$, we assume that the transition probabilities of the Markov process $\{U_t : t \geq 0\}$ satisfy

$$\boldsymbol{P}_{w_1,w_2}(t, t+h) = \begin{cases} \lambda_{w_1}(t) \cdot h + o(h) & \text{if } w_1 \neq w_2 \\ 1 - \lambda_{w_1}(t) \cdot h + o(h) & \text{if } w_1 = w_2 \end{cases} \tag{8}$$

The main result relies on the following indefinite integrals:

$$\begin{aligned} \Lambda_w(t) &= \int \lambda_w(s)\, ds, \quad \text{and} \\ \Pi_w(t) &= \int e^{\Lambda_0(s)+\Lambda_1(s)}\, \lambda_w(s)\, ds, \quad w \in \mathcal{S}. \end{aligned}$$

Proposition. Let $\lambda_w(t)$ be a positive, continuous function such that $\lim_{t\downarrow 0} \lambda_w(t)$ is finite for $w \in \mathcal{S}$. For the Markov process (8), the diagonal elements of the transition matrix $\boldsymbol{P}(0, t)$ are given by

$$P_{w,w}(0,t) = e^{-\Lambda_0(t)-\Lambda_1(t)} \left(\Pi_{1-w}(t) + c_w\right), \quad w \in \mathcal{S},$$

where the constant c_w is determined by the boundary condition: $P_{w,w}(0, 0) = 1$.

Proof. Let $\boldsymbol{G}(t)$ be a 2×2 matrix with element $G_{w_1,w_2}(t) = \lambda_{w_1}(t)$ if $w_1 \neq w_2$ and $G_{w_1,w_2}(t) = -\lambda_{w_1}(t)$ if $w_1 = w_2$, where $w_1, w_2 \in \mathcal{S}$. Clearly,

$$\boldsymbol{G}(t) = \lim_{h\downarrow 0} \frac{1}{h} \left(\boldsymbol{P}(t, t+h) - \boldsymbol{I}\right).$$

The matrices $\{\boldsymbol{P}(0, t) : t \geq 0\}$ satisfy Kolmogorov's forward equation, $\frac{d}{dt}\boldsymbol{P}(0, t) = \boldsymbol{P}(0, t) \cdot \boldsymbol{G}(t)$. Therefore, for $w \in \mathcal{S}$, we obtain the ordinary differential equation

$$\frac{d}{dt} P_{w,w}(0, t) = -\left(\lambda_0(t) + \lambda_1(t)\right) \cdot P_{ww}(0, t) + \lambda_{1-w}(t)$$

whose solution has the afore-mentioned form.

Corollary 1. When $\lambda_w(t) \equiv \lambda_w$ is a fixed, positive number for $w \in \mathcal{S}$, we obtain the *two-state alternating Markov process.* The diagonal elements of the transition matrix $\boldsymbol{P}(0,t)$ are then

$$P_{w,w}(0,t) = \frac{\lambda_{1-w}}{\lambda_0+\lambda_1} + \frac{\lambda_w}{\lambda_0+\lambda_1} e^{-(\lambda_0+\lambda_1)t} \qquad w \in \mathcal{S},$$

Given the vector $\boldsymbol{\rho}_0$ of marginal probabilities of U_0, we can compute the vector of marginal probabilities of U_t as $\boldsymbol{\rho}_t = \boldsymbol{P}^T(0,t) \cdot \boldsymbol{\rho}_0$. For any $0 \le t < s$, the transition matrix $\boldsymbol{P}(t,s) = \boldsymbol{P}^{-1}(0,t) \cdot \boldsymbol{P}(0,s)$ by the Chapman-Kolmogorov equation.

Since $\boldsymbol{P}(t,s)$ varies continuously as a function of t and s, $\lim_{s \downarrow t} \boldsymbol{P}_{w_1,w_2}(t,s)$ equals 1 if $w_1 = w_2$, and equals 0 if $w_1 \neq w_2$ (*Property (i) of Section 2.2*). Property *(ii)* of Section 2.2 is guaranteed under the further assumption that $\int_0^\infty \lambda_w(s)\, ds = \infty$ for $w \in \mathcal{S}$, as stated in the following result.

Corollary 2. Suppose that $\int_0^\infty \lambda_w(s)\, ds = \infty$ for $w \in \mathcal{S}$. Then

$$\lim_{t \to \infty} \boldsymbol{P}_{0,w}(0,t)/\boldsymbol{P}_{1,w}(0,t) = 1, \qquad w \in \mathcal{S}.$$

Proof. We have $\lim_{t\to\infty} \Lambda_w(t) = \infty$. Therefore, when $w = 0$,

$$\begin{aligned}
\lim_{t \to \infty} \boldsymbol{P}_{0,0}(0,t)/\boldsymbol{P}_{1,0}(0,t) &= \lim_{t \to \infty} \frac{e^{-\Lambda_0(t)-\Lambda_1(t)}\,(\Pi_1(t)+c_0)}{1-e^{-\Lambda_0(t)-\Lambda_1(t)}\,(\Pi_0(t)+c_1)} \\
&= \lim_{t \to \infty} \frac{e^{-\Lambda_0(t)-\Lambda_1(t)} \cdot \Pi_1(t)}{1-e^{-\Lambda_0(t)-\Lambda_1(t)} \cdot \Pi_0(t)} \\
&= 1,
\end{aligned}$$

since $\Pi_0(t) + \Pi_1(t) = e^{\Lambda_0(t)+\Lambda_1(t)} + c$, where c does not depend on t. The case is similarly proved when $w = 1$.

An alternative interpretation. The Markov process (8) is equivalent to the following stochastic mechanism. Suppose $U_t = w_{t_1}$ at time t_1. Due to the Markov assumption, the values of the adjacency variables previous to the time t_1 do not affect the future evolution of the process. Let $t_2 > t_1$ be the first instance (downstream of $t = t_1$) that the process switches to the state $1 - w_{t_1}$. Then the difference $t_2 - t_1$ is distributed as the first arrival time of a non-homogenous Poisson process with intensity function $\lambda_{w_1}(t)$. At time $t = t_2$, the system switches to the Poisson process with intensity function $\lambda_{1-w_1}(t)$.

2.4. Application to disease mapping problems. For every pair of areas i and j belonging to the region $\mathcal{D}$ and separated by a generalized distance of Δ_{ijt} at time t, we set

$$\lambda_0^{(ij)}(t) = \gamma e^{-\Delta_{ijt}} \text{ and } \lambda_1^{(ij)}(t) = \gamma \left(1 - e^{-\Delta_{ijt}}\right), \tag{9}$$

for some $\gamma > 0$. When the intercentroidal distance is included in (7), the intensity functions $\lambda_0^{(ij)}(t)$ and $\lambda_1^{(ij)}(t)$ satisfy the conditions stated in the Proposition and its corollaries because the generalized distance is uniformly bounded above zero.

We assume that the process $\{W_{ijt} : t \geq 0 \text{ and } i, j \text{ fixed}\}$ is distributed according to the Markov process (8) corresponding to functions (9). The prior specification is completed by assuming independent distributions for the initial variables $\{W_{ij0} : i, j \in \mathcal{D}\}$. The Markov process depends on the vector $\boldsymbol{\gamma}$ and vector $\boldsymbol{\phi}$ appearing in (7), and is denoted by $\mathcal{M}_{ij}(\boldsymbol{\gamma}, \boldsymbol{\phi})$. The adjacency indicators $\boldsymbol{w}_{ij} = (W_{ijt_1}, \ldots, W_{ijt_T})$ are regarded as realizations of $\mathcal{M}_{ij}(\boldsymbol{\gamma}, \boldsymbol{\phi})$ at the times $t_1, \ldots, t_T$.

Invoking the afore-mentioned interpretation in terms of arrival times of non-homogenous Poisson processes, we see why specification (9) is appropriate. Suppose the generalized distance between areas i and j increases gradually with time as the demographic conditions change, so that the areas grow increasingly "dissimilar". For large t, we have $\lambda_0^{(ij)}(t) < \lambda_1^{(ij)}(t)$, and so the process $\{W_{ijt} : \text{large } t \text{ and } i, j \text{ fixed}\}$

tends to stay for longer durations in the state 0 ("not neighbors") than in the state 1 ("neighbors"). The marginal probabilities $\rho_0^{(ij)}(t) > \rho_1^{(ij)}(t)$ for large t, and Property *(iii)* of Section is satisfied. The process similarly adapts to non-monotone changes in generalized distance.

As a second example, imagine that the generalized distances between the n areal units remain constant over time. Consider pairs of areas (i, j) and (i, j') such that $\Delta_{ijt} < \Delta_{ij't}$ for all t. (For instance, the intercentroidal distance is the only covariate involved in definition (7), and area j' is situated farther away from area i than area j.) Then the intensity functions in (9) satisfy $\lambda_0^{(ij)}(t) > \lambda_0^{(ij')}(t)$ and $\lambda_1^{(ij)}(t) < \lambda_1^{(ij')}(t)$ for all t. Therefore, the process $\{W_{ij't} : t \geq 0 \text{ and } i, j' \text{ fixed}\}$ has a greater tendency than the process $\{W_{ijt} : t \geq 0 \text{ and } i, j \text{ fixed}\}$ to stay in the state 0. The marginal probabilities satisfy $\rho_0^{(ij)}(t) < \rho_0^{(ij')}(t)$. This is precisely the requirement of Property *(iii)* of Section 2.2.

The intensity functions (9) satisfy $\lambda_0^{(ij)}(t) + \lambda_1^{(ij)}(t) = \gamma$, which results in the following simplifications:

$$P_{0,0}^{(ij)}(0,t) = e^{-\gamma t}\left(\gamma \int_0^t e^{\gamma u}\left(1 - e^{-\Delta_{iju}}\right) du + 1\right),$$
$$P_{1,1}^{(ij)}(0,t) = e^{-\gamma t}\left(\gamma \int_0^t e^{\gamma u} e^{-\Delta_{iju}} du + 1\right).$$

In real applications, the generalized distances and intensity functions are only available at the times $t_1, \ldots, t_T$. The intensity functions may be assumed to be piecewise constant or they may be linearly interpolated. More sophisticated interpolation techniques may be used if necessary.

Special cases. When the data are observed at uniformly spaced time intervals, it is convenient to rescale the time axis and parameters of the Markov process so that $\mathcal{T} = \{1, \ldots, T\}$. Additionally, if the generalized distance depends only on the intercentroidal distances d_{ij} between areas i and j, we have $\Delta_{ijt} = \phi\, d_{ij}$. Then the intensity functions $\lambda_0^{(ij)}(t)$ and $\lambda_1^{(ij)}(t)$ are time invariant and we obtain a homogeneous continuous-time Markov process in (8). For any $t \in \{1, \ldots, T-1\}$, the transition matrix $\boldsymbol{P}^{(ij)}(t, t+1)$ equals $\boldsymbol{P}^{(ij)}(1, 2)$. The off-diagonal elements of the matrix $\boldsymbol{P}^{(ij)}(1, 2)$ are given by

$$
\begin{aligned}
P_{0,1}^{(ij)}(1,2) &= e^{-\phi d_{ij}} \left(1 - e^{-\gamma}\right), \\
P_{1,0}^{(ij)}(1,2) &= \left(1 - e^{-\phi d_{ij}}\right) \cdot \left(1 - e^{-\gamma}\right).
\end{aligned} \tag{10}
$$

Further, if the initial marginal distribution $\boldsymbol{\rho}^{(ij)}(1)$ at time $t = 1$ is taken to be the stationary distribution of the Markov chain, then we obtain for any t:

$$
\rho_0^{(ij)}(t) = 1 - e^{-\phi d_{ij}}, \quad \rho_1^{(ij)}(t) = e^{-\phi d_{ij}}.
$$

The apriori probability of the areas i and j being neighbors does not depend on the time t or on γ, and it decreases as the distance d_{ij} increases (*Property (iii) of Section 2.2*).

2.5. The Bayesian hierarchical model. For areas $i, j \in \mathcal{D} = \{1, \ldots, n\}$ and times $t \in \mathcal{T} = \{t_1, \ldots, t_T\}$, we assume

$$
\begin{aligned}
Y_{it} \mid \eta_{it}, \varsigma &\overset{ind}{\sim} h(Y_{it}, \varsigma) \cdot \exp\left\{(Y_{it}\, \omega(\eta_{it}) - b(\eta_{it}) / a(\varsigma)\right\} \\
\eta_{it} &= o_{it} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_f + \boldsymbol{z}'_{it}\boldsymbol{\beta}_r + \theta_{it} \\
\boldsymbol{\beta}_r &\sim N_p(\mu_\beta, \Sigma_\beta) \\
\boldsymbol{\theta} \mid \boldsymbol{W}^*, \boldsymbol{\nu} &\sim CAR(\boldsymbol{W}^*, \boldsymbol{\nu}) \\
(W_{ijt_1}, \ldots, W_{ijt_T}) &\overset{ind}{\sim} \mathcal{M}_{ij}(\gamma, \boldsymbol{\phi})
\end{aligned} \tag{11}
$$

Flat priors are assumed for the fixed effects $\boldsymbol{\beta}_f$, $\log \gamma$, $\log \boldsymbol{\phi}$ and variance components $\boldsymbol{\nu}$.

3. Simulation-based inference. Conditional on the remaining model parameters, the fixed and random effects are updated using the Metropolis-Hastings procedure of Zeger and Karim (1991). Conditional on the adjacency matrix $\boldsymbol{W}^*$, the variance components $\boldsymbol{\nu}$ of the proper CAR model are similarly updated by making normal proposals within Metropolis-Hastings steps. Some components of $\boldsymbol{\nu}$ may be transformed so that the conditional posterior more closely resembles a normal distribution.

Updating the adjacency matrix $\boldsymbol{W}^*$

The prior distribution of the adjacency indicator W_{ijt_m}, for $t_m \in \mathcal{T} - \{t_1, t_T\}$ and any pair of areas $i, j \in \mathcal{D}$, is

$$\begin{aligned}
&\Pr\left(W_{ijt_m} = w \mid \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right) \\
&\quad = \Pr\left(W_{ijt_m} = w \mid W_{ijt_{m-1}} = w_{m-1}, W_{ijt_{m+1}} = w_{m+1}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right) \\
&\quad \propto P^{(ij)}_{w_{m-1},w}(t_{m-1}, t_m) \cdot P^{(ij)}_{w,w_{m+1}}(t_m, t_{m+1}), \quad w \in \mathcal{S}. \qquad (12)
\end{aligned}$$

Similar expressions are obtained when $t = t_1$ or $t = t_T$. Let $\boldsymbol{F} = (1-\lambda)(\boldsymbol{D}^* - \boldsymbol{W}^*) + \lambda \boldsymbol{I}$, where $\boldsymbol{D}^*$ is the diagonal matrix of the row sums of $\boldsymbol{W}^*$. For the proper CAR model (4), we have

$$[\boldsymbol{\theta} \mid W_{ijt_m} = w, \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\nu}] \propto |\boldsymbol{F}| \cdot \exp\left\{-\frac{1}{2\sigma^2} w(\theta_i - \theta_j)^2\right\}, \quad w \in \mathcal{S}. \qquad (13)$$

The full conditional of W_{ijt_m} is therefore

$$\begin{aligned}
&\Pr\left(W_{ijt_m} = w \mid \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right) \\
&\propto [\boldsymbol{\theta} \mid W_{ijt_m} = w, \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\nu}] \cdot \Pr\left(W_{ijt_m} = w \mid \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right), \quad w \in \mathcal{S}. \qquad (14)
\end{aligned}$$

When n and T are small, we can evaluate the full conditional of W_{ijt_m} using the above expressions and update the adjacency indicators using Gibbs sampling. The calculation can be intensive when nT is large because of the determinant involved in (13). Therefore, for large nT, we assume

$$[\boldsymbol{\theta} \mid W_{ijt_m} = w, \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\theta}, \boldsymbol{\nu}] \approx \exp\left\{-\frac{1}{2\sigma^2} w(\theta_i - \theta_j)^2\right\},$$

which corresponds to an improper CAR model (Ma, Carlin and Banerjee, 2006). We then have the following approximation to the full conditional:

$$\begin{aligned} &\Pr\left(W_{ijt_m} = w \mid \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\theta}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right) \\ \approx\ &\exp\left\{-\frac{1}{2\sigma^2} w(\theta_i - \theta_j)^2\right\} \cdot \Pr\left(W_{ijt_m} = w \mid \boldsymbol{W}^* - \{W_{ijt_m}\}, \boldsymbol{\gamma}, \boldsymbol{\phi}\right), \quad w \in \mathcal{S}. \end{aligned} \tag{15}$$

Approximation (15) drastically reduces the computational cost of generating proposals because it avoids the computation of determinants. All n^2 adjacency indicators for a given value of $t \in \mathcal{T}$ are jointly proposed by this method, and accepted or rejected in a single Metropolis step. The acceptance rates exceed 70% in most applications.

For a more formal description of the Metropolis Hastings step, let $\boldsymbol{W}_1^*$ be the augmented adjacency matrix corresponding to the joint proposals and $\boldsymbol{W}_0^*$ be the current value of the adjacency matrix. Let $q(\boldsymbol{W}_1^* \mid \boldsymbol{W}_0^*)$ be the probability of proposing $\boldsymbol{W}_1^*$ from $\boldsymbol{W}_0^*$ based on approximation (15). Let $\boldsymbol{F}_1 = (1-\lambda)(\boldsymbol{D}_1^* - \boldsymbol{W}_1^*) + \lambda \boldsymbol{I}$, where $\boldsymbol{D}_1^*$ is the diagonal matrix of the row sums of $\boldsymbol{W}_1^*$. Similarly, define the matrices $\boldsymbol{D}_0^*$ and $\boldsymbol{F}_0$ corresponding to the matrix $\boldsymbol{W}_0^*$. Then the proposed move is accepted with probability

$$\varpi(\boldsymbol{W}_0^*, \boldsymbol{W}_1^*) = \min\left\{1, \frac{q(\boldsymbol{W}_0^* \mid \boldsymbol{W}_1^*) \cdot |\boldsymbol{F}_1|^{1/2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\theta}'\boldsymbol{F}_1\boldsymbol{\theta}\right\}}{q(\boldsymbol{W}_1^* \mid \boldsymbol{W}_0^*) \cdot |\boldsymbol{F}_0|^{1/2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\theta}'\boldsymbol{F}_0\boldsymbol{\theta}\right\}}\right\}$$

Computing the determinants of $\boldsymbol{F}_0$ and $\boldsymbol{F}_1$ can be intensive when dimension nT is large. Section A1 of the Appendix describes an efficient technique.

Updating the region-specific random effects

The following description assumes the proper CAR model of Leroux, Lei and Breslow (1998). The updates are similarly derived for other versions of the CAR model. Let the square matrix $\boldsymbol{W}_{t_m}$ of dimension n, consisting of only zeros and ones, summarize the purely spatial neighborhood relationships among the n areal units at time t_m. Assume that the i^{th} element of the diagonal matrix $\boldsymbol{D}_{t_m}$ equals the total number of neighbors (spatial as well as temporal) of the area i at time t_m. Define $\boldsymbol{\Sigma}_{t_m}^{-1} = \sigma^{-2}\left[(1-\lambda)\cdot(\boldsymbol{D}_{t_m} - \boldsymbol{W}_{t_m}) + \lambda\cdot \boldsymbol{I}\right]$. Define $\varepsilon_{t_m i} = \theta_{it_2}$ if $m = 1$ and $\varepsilon_{t_m i} = \theta_{it_{m-1}} + \theta_{it_{m+1}}$ if $m > 1$. Let the i^{th} element of the vector $\boldsymbol{\alpha}_{t_m}$ equal $\sigma^{-2}(1-\lambda)\cdot\varepsilon_{t_m i}$.

Applying the Laplace approximation (expression (18) of Section A2 in the Appendix), it can be shown that for $m = 1, \ldots, T$, the posterior distribution of the block $\boldsymbol{\theta}_{t_m} = (\theta_{1t_m}, \ldots, \theta_{nt_m})'$, conditional on the remaining random effects, is approximately multivariate normal with precision $\boldsymbol{U}_{t_m}^{-1} + \boldsymbol{\Sigma}_{t_m}^{-1}$, where $\boldsymbol{U}_{t_m}$ is the diagonal matrix of working weights corresponding to $\boldsymbol{\theta}_{t_m}$. Therefore, for $m = 1, \ldots, T$:

(i) Let $\boldsymbol{\theta}_{t_m}^0$ be the current value of $\boldsymbol{\theta}_{t_m}$. Compute the precision $(\boldsymbol{U}_{t_m}^0)^{-1} + \boldsymbol{\Sigma}_{t_m}^{-1}$ and generate the proposal $\boldsymbol{\theta}_{t_m}^* \sim N_n\left(\boldsymbol{\theta}_{t_m}^0, c_n^2\left[(\boldsymbol{U}_{t_m}^0)^{-1} + \boldsymbol{\Sigma}_{t_m}^{-1}\right]^{-1}\right)$. The constant c_n is chosen so that the acceptance rates in step *(ii)* are approximately 25% (Gelman, Roberts and Gilks, 1995).

(ii) Using the proposal $\boldsymbol{\theta}_{t_m}^*$, compute $(\boldsymbol{U}_{t_m}^*)^{-1} + \boldsymbol{\Sigma}_{t_m}^{-1}$ and vector $\boldsymbol{\alpha}_{t_m}^*$. Accept

the proposal with probability

$$\varpi(\boldsymbol{\theta}^0_{t_m}, \boldsymbol{\theta}^*_{t_m}) =$$

$$\min\left\{1, \frac{[Y_{it} \mid \eta^*_{it}, \varsigma] \cdot N_q\left(\boldsymbol{\theta}^*_{t_m} | \boldsymbol{\Sigma}^0_{t_m} \cdot \boldsymbol{\alpha}^0_{t_m}, \boldsymbol{\Sigma}^0_{t_m}\right) \cdot N_q\left(\boldsymbol{\theta}^0_{t_m} \mid \boldsymbol{\theta}^*_{t_m} \cdot, c_q^2 \left[(\boldsymbol{U}^*_{t_m})^{-1} + \boldsymbol{\Sigma}^{-1}_{t_m}\right]^{-1}\right)}{[Y_{it} \mid \eta^0_{it}, \varsigma] \cdot N_q\left(\boldsymbol{\theta}^0_{t_m} | \boldsymbol{\Sigma}^*_{t_m} \cdot \boldsymbol{\alpha}^*_{t_m}, \boldsymbol{\Sigma}^*_{t_m}\right) \cdot N_q\left(\boldsymbol{\theta}^*_{t_m} \mid \boldsymbol{\theta}^0_{t_m} \cdot, c_q^2 \left[(\boldsymbol{U}^0_{t_m})^{-1} + \boldsymbol{\Sigma}^{-1}_{t_m}\right]^{-1}\right)}\right\}$$

Updating the parameters of the Markov process

It is convenient to work with the transformation $\boldsymbol{\zeta} = (\log \gamma, \log \boldsymbol{\phi})$, since the posterior more closely resembles a $(p+1)$-variate normal distribution. Conditional on the adjacency indicators, the posterior density $[\boldsymbol{\zeta} \mid \boldsymbol{W}^*]$ equals $[\boldsymbol{W}^* \mid \boldsymbol{\zeta}]$ due to the flat priors for $\boldsymbol{\zeta}$. With L denoting the conditional posterior $[\boldsymbol{\zeta} \mid \boldsymbol{W}^*]$, the Metropolis-Hastings updates consist of the following steps:

(i) Let $\boldsymbol{\zeta}_0$ denote the current value of $\boldsymbol{\zeta}$. Generate the proposal $\boldsymbol{\zeta}_1 \sim N_{p+1}\left(\boldsymbol{\zeta}_0, c^2_{p+1}\boldsymbol{V}_0\right)$, where $\boldsymbol{V}_0^{-1} = -\partial^2 L/\partial\boldsymbol{\zeta}^2 \mid_{\boldsymbol{\zeta}_0}$ and constant c_{p+1} is chosen so that the acceptance rates in step *(ii)* are approximately 25%.

(ii) Accept the proposal with probability

$$\varpi(\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1) = \min\left\{1, \frac{[\boldsymbol{W}^* \mid \boldsymbol{\zeta}_1] \cdot N_{p+1}\left(\boldsymbol{\zeta}_0 \mid \boldsymbol{\zeta}_1, c^2_{p+1}\boldsymbol{V}_1\right)}{[\boldsymbol{W}^* \mid \boldsymbol{\zeta}_0] \cdot N_{p+1}\left(\boldsymbol{\zeta}_1 \mid \boldsymbol{\zeta}_0, c^2_{p+1}\boldsymbol{V}_0\right)}\right\}$$

where $\boldsymbol{V}_1^{-1} = \partial^2 L/\partial\boldsymbol{\zeta}^2 \mid_{\boldsymbol{\zeta}_1}$.

Posterior inference. The post–burn-in MCMC sample can be used to estimate various features of the posterior distribution. For example, by focussing only on the simulated values of the fixed effects, we compute the Monte Carlo average, $\hat{E}[\boldsymbol{\beta}_f|\boldsymbol{Y}]$. This estimate of the fixed effects marginalizes over the remaining model parameters, including the neighborhood structure. The adjacency indicators can be estimated as follows. For the areas i, $j \in \mathcal{D}$ and time $t \in T$, let the k^{th} MCMC iterate of

W_{ijt} be denoted by $w_{ijt}^{(k)}$. A simulation-based estimate of the posterior probability $P\left(W_{ijt}=1 \mid \boldsymbol{Y}\right)$ is then

$$\hat{P}\left(W_{ijt}=1 \mid \boldsymbol{Y}\right)=\frac{1}{N} \sum_{k=1}^{N} w_{ijt}^{(k)}$$

We declare $\hat{W}_{ijt}=1$ if $\hat{P}\left(W_{ijt}=1 \mid \boldsymbol{Y}\right)>0.5$, and declare $\hat{W}_{ijt}=0$ otherwise. The procedure is applied to estimate all random elements of the matrix $\boldsymbol{W}^{*}$.

4. Illustration. The Spatial Environmental Epidemiology in New South Wales project yielded outcome data on ischemic heart disease (IHD) abstracted from daily separation records from all public and private hospitals in New South Wales, Australia, during the period July 1, 1996 to June 30, 2001. Population data were obtained from census information collected by the Australian Bureau of Statistics (ABS) and inter-censal estimates, called Estimated Residential Populations (ERPs), provided for July 1st of each non-census year. Patient reported residential postcode was used to assign the geographical location of hospitalization for IHD. Available data included the postcode of residence, date of hospitalization, patient age and patient gender. Patient age was grouped into one of the following categories: younger than 20 years, 13 different 5-year intervals (20-24 years, 24-29 years etc. up to 80-84 years), plus a 15^{th} category of 85 years or older.

In this investigation, we focus on the Sydney Metropolitan Area consisting of 249 postcodes. In addition to estimating the neighborhood structure and the random effects associated with the postcodes, we wish to explore the association of IHD with an index of socioeconomic disadvantage called the SEIFA (Socio-Economic Indexes for Areas), provided for each postcode by ABS. A low SEIFA score reflects relatively low educational attainment and income, high unemployment, and jobs in relatively unskilled occupations. The SEIFA scores are re-centered around zero to justify the specification of independent priors on the fixed effects. Available information includes the location of the postcode centroids. Refer to Burden et al. (2005), Guha and Ryan (2006) and Guha et al. (2006) for analyses relying on adjacency matrix (3). We partition the 1820-day interval into 20 "quarters" consisting of 91 days each. The random effect associated with postcode i ($i = 1, \ldots, 249$) is assumed to remain constant during each quarter. We denote by θ_{it} the random effect associated with postcode i during quarter t ($t = 1, \ldots, 20$). Let j ($j = 1, \ldots, 1820$) index the day of the five-year period and k ($k = 1, \ldots, 30$) index the social categories consisting of unique combinations of age category and gender.

Denote by Y_{ijk} the number of IHD hospitalizations in the i^{th} postcode ($i = 1, \ldots, 249$), among the N_{ijk} subjects at risk on day j and belonging to the k^{th} social category. Let seifa_{ij} and dens_{ij} respectively denote the SEIFA index and population density of postcode i on day j. We assume a Poisson GLMM with log link:

$$Y_{ijk} \sim Po(\mu_{ijk}),$$

$$\log(\mu_{ijk}) = \log(N_{ijk}) + \delta_k + \alpha_w + \beta_0 + \beta_{1k} \cdot \text{seifa}_{ij} + \beta_2 \cdot \text{dens}_{ij} + \theta_{it},$$

where δ_k is the fixed effect associated with the k^{th} social category (with δ_1 being the reference group), α_w is the fixed effect of the w^{th} day of the week (with $w = 1$ representing Mondays being the reference group), β_0 is the model intercept, β_{1k} is the interaction term between socioeconomic status and the k^{th} social category, and β_2 is the linear predictor associated with the population density.

Assuming the proper CAR model of Leroux, Lei and Breslow (1998) with precision matrix (4), we define the generalized distance between areas i and i' as $\Delta_{ii't} = \phi \, d_{ii'}$ where $d_{ii'}$ is the intercentroidal distance. We also assume that the

Parameter	Mean	SD
intercept	-17.98	0.005
density/10^5	-1.210	0.1146
Variance components		
σ^{-2}	13.20	1.072
λ	0.13	0.023
Markov process parameters		
γ	0.03	0.009
ϕ	21.88	5.881

Table 1: Estimated posterior means and standard deviations of selected model (16) parameters.

marginal distribution $\boldsymbol{\rho}^{(ij)}(1)$ at quarter 1 equals the stationary distribution of the Markov process $\mathcal{M}_{ij}(\gamma, \phi)$. The hierarchical model is then

$$\begin{aligned}
Y_{it} \mid \eta_{it} &\stackrel{ind}{\sim} Po(e^{\eta_{ijk}}) \\
\eta_{it} &= \log(N_{ijk}) + \delta_k + \alpha_w + \beta_0 + \beta_{1k} \cdot \text{seifa}_{ij} + \beta_2 \cdot \text{dens}_{ij} + \theta_{it} \\
\boldsymbol{\theta} \mid \boldsymbol{W}^*, \nu &\sim CAR(\boldsymbol{W}^*, \nu) \qquad (16) \\
(W_{ij,1}, \ldots, W_{ij,20}) &\stackrel{ind}{\sim} \mathcal{M}_{ij}(\gamma, \phi) \\
\boldsymbol{\nu} &= (\sigma^{-2}, \lambda) \\
\sigma^{-2} &\sim \text{gamma}(1, 1)
\end{aligned}$$

Flat priors are assumed for the fixed effects, $\tau = \log(\lambda/(1-\lambda))$, $\log\gamma$ and $\log\phi$. The MCMC algorithm described in Section 3 is applied to simulate posterior samples. An initial set of 10,000 draws was discarded as burn-in and the remaining 100,000 samples were used for inference.

Table 1 displays the estimates and posterior standard deviations for the model intercept β_0, coefficient β_2 associated with the population density, variance components $\boldsymbol{\nu} = (\sigma^{-2}, \lambda)$ of the CAR model of Leroux et al. (1998), and parameters ϕ and γ related to the latent Markov process.

For any given quarter, the posterior probabilities that two postcodes are neighbors decreases rapidly with their intercentroidal distance. This is seen in the upper panel of Figure 1, where we observe that most of the postcodes are separated by distances greater than 20 kilometers (km), but the posterior probabilities of being neighbors tend to be very small at these distances. During quarter 1, the barplot on the bottom left panel of Figure 1 shows that there are 536 postcode pairs estimated as neighbors ($\hat{W}_{ij1} = 1$) at intercentroidal distances less than 8 km, but only 15 pairs are estimated as neighbors at distances exceeding 22 km. For comparison, the 249 postcodes in the Sydney Metropolitan Area correspond to a total of 30,876 distinct postcode pairs. The results are similar for the remaining quarters 2 through 20.

The solid line in the bottom right panel of Figure 1 represents the average number of neighbors per postcode for quarters 1 through 20. In contrast, the dashed line represents the time-independent average number of neighbors per postcode corresponding to neighborhood structure (3), suggesting that the estimated neighborhood structure considerably differs from standard neighborhood assumptions.

Figure 2 plots the estimated random effects for four randomly selected postcodes. The dashed lines represent margins of two posterior standard deviations. The graph indicates that arbitrary temporal trends in the region-specific random effects are captured by the proposed model, unlike models that account for temporal variation by including a linear or quadratic term in the mean structure.

The left panel of Figure 3 plots the estimated main effects of the age/gender categories with the youngest age group of the male subpopulation as the reference group. The IHD rates increase sharply with age. Although the estimates follow similar trends in both genders, males typically have higher IHD rates than females of the same age. The day-of-week effects are plotted in the right panel of Figure

3. The effect for Monday, the reference group, is found to be significantly higher than the remaining days of the week. Possible explanations for the sharp decrease over the weekend include job-related stress on weekdays, and the fact that people tend to ignore warning health signs during weekends, resulting in a high number of emergency room visits on Mondays.

Figure 4 displays the estimated interactions of age/gender category and SEIFA for males. The estimates for females had a very similar trend. Most of the interaction terms are negative indicating that people with high socioeconomic disadvantage (i.e. smaller SEIFA values) are at higher risk for IHD. The interactions increase with age until they are positive for the older age groups. The increase reflects the greater susceptibility to IHD for older people, for whom socioeconomic status appears to play a lesser role in determining the risk of heart disease.

5. Conclusion. The paper proposes an extension to the CAR model that simultaneously allows the spatio-temporal analysis of areal data and the discovery of neighborhood relationships among areal units in a domain. The neighborhood structure is modeled as a continuous-time Bernoulli process with Markovian dependence. The hierarchical model has several desirable theoretical properties, flexibly captures time trends in spatially correlated data, and is especially useful in applications like health surveillance where the study of spatial relationships and area-specific random effects influences the effective detection and control of disease. In investigations where the latent neighborhood structure is not directly relevant, it is marginalized to estimate the parameters of interest. The methodology is applied to analyze outcome data on ischemic heart disease in the Sydney Metropolitan Area, where the temporal trends in the spatial random effects and the estimated neighborhood structure differ considerably from the assumptions made in typical CAR analyses.

REFERENCES

Banerjee, S., Wall, M. M. and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4, 123-142.

Banerjee, S., Carlin, B. and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Florida.

Bernardinelli, L., Clayton, D. G. and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14, 2411–2432.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc., Ser. B*, 36, 192–236.

Besag, J., Mollié, A. and York, J. (1991), "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, 43, 1–20.

Bohning, D., Dietz, E. and Schlattmann, P. (2000). Space-time mixture modelling of public health data. *Statistics in Medicine*, 19, 2333–2344.

Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51, 481–483.

Burden S., Guha, S. Morgan, G., Ryan, L. and Young L. (2005). Spatio-temporal Analysis of Ischemic Heart Disease in NSW, Australia. *Environmental and Ecological Statistics*, 12, 427–448.

Carlin, B.P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatially correlated survival data. In *Bayesian Statistics 7*. Oxford: Oxford

University Press, 45–64.

Clayton, D. G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43, 671-681.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, Revised Edition, Wiley, New York.

Cressie, N. A. C. and Chan N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, 84, 393-401.

Gelfand, A. E., Kottas, A. and MacEachern, S. N. (2005). Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing. *Journal of the American Statistical Association*, 100, 1021-1035.

Grimmett and Stirzaker (2001). *Probability and Random Processes*. Third Edition, Oxford University Press.

Giudici, P., Knorr-Held, L. And Rasser, G. (2000). Modelling categorical covariates in Bayesian disease mapping by partition structures. *Statistics in Medicine*, 19, 2579–2593.

Guha, S. and Ryan, L. M. (2006). Gauss-Seidel Estimation of Generalized Linear Mixed Models with Application to Poisson Modeling of Spatially Varying Disease Rates. Submitted.

Guha, S., Ryan, L. M., Morgan, G., Beard, J. (2006). Age and gender modify the association between socio-economic factors and heart disease in New South Wales, Australia. Submitted.

Harville, D. A. (1997). *Matrix algebra from a statistician's perspective*. Springer-Verlag Inc., New York.

Hoff, P. D., Raftery, A. E. and Handcock, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.*, 97, 1090–1098.

Knorr-Held, L. and Rasser, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56, 13–21.

Kottas A., Duan J. and Gelfand A. E. (2006). Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures. Technical Report, Department of Applied Math and Statistics, University of California, Santa Cruz.

Leroux, B. G., Lei, X. and Breslow, N. (2000), Estimation of spatial disease rates in small areas: a new mixed model for spatial dependence, *The IMA Volumes in Mathematics and its Applications, Statistical Models in Epidemiology, the Environment, and Clinical Trials* (Halloran, M. E. and Berry, D. eds.), 116, 179–191, Springer-Verlag, New York.

Ma, H., Carlin, B. P. and Banerjee, S. (2006). Hierarchical and joint site-edge methods for Medicare hospice service region boundary analysis, Research Report 2006-010, Division of Biostatistics, University of Minnesota.

McCullagh, P. and Nelder, J. A. (1999), *Generalized Linear Models* (2nd ed.), CRC Press LLC, Boca Raton, Florida.

Wakefield, J. C. and Morris, S. E. (1999). An application of spatial errors-in-variables modelling: investigating the relationship between ischaemic heart disease and water constituents. *Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting* (Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. eds.), p. 657–684, Oxford University Press.

Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92, 607617.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs.

J. Amer. Statist. Assoc., 82, 8–19.

Zeger, S. L. and Karim, M. R. (1991), "Generalized linear models with random effects: A Gibbs sampling approach," *Journal of the American Statistical Association*, 86, 79–86.

APPENDIX

A1. Computing the determinant of matrix $\boldsymbol{F}$. Let $\boldsymbol{F} = (1-\lambda)(\boldsymbol{D}^* - \boldsymbol{W}^*) + \lambda \boldsymbol{I}$, where $\boldsymbol{W}^*$ is the augmented adjacency matrix of dimension nT and $\boldsymbol{D}^*$ is the diagonal matrix of row sums of $\boldsymbol{W}^*$. For $m = 1, \ldots, T$, let $\boldsymbol{W}_m$ be the *spatial* adjacency matrix of the n areas at time t_m and $\boldsymbol{D}_m$ be the diagonal matrix consisting of the row sums of $\boldsymbol{W}_m$. Define the matrices $\boldsymbol{R}_1 = (1-\lambda)(\boldsymbol{D}_1 - \boldsymbol{W}_1) + \boldsymbol{I}$, $\boldsymbol{R}_m = (1-\lambda)(\boldsymbol{D}_m - \boldsymbol{W}_m) + (2-\lambda)\boldsymbol{I} - (1-\lambda)^2 \boldsymbol{R}_{m-1}^{-1}$ for $m = 2, \ldots, T$, and $\boldsymbol{R}_T = (1-\lambda)(\boldsymbol{D}_T - \boldsymbol{W}_T) + \boldsymbol{I} - (1-\lambda)^2 \boldsymbol{R}_{T-1}^{-1}$. Then

$$|\boldsymbol{F}| = \prod_{m=1}^{T} |\boldsymbol{R}_m|.$$

A2. Laplace approximation. For models belonging to the exponential family, the Laplace approximation applies a linearized version of the link function $g(\cdot)$ to the data $Y_{1t_1}, \ldots, Y_{nt_T}$. The *working value* for each case is defined as

$$y_{it_m} = \eta_{it_m} + \frac{\partial \eta_{it_m}}{\partial \mu_{it_m}} \cdot (Y_{it_m} - \mu_{it_m}), \quad i = 1, \ldots, n \text{ and } m = 1, \ldots, T. \tag{17}$$

The *working weight* is defined as $u_{it_m} = \{\Upsilon(\mu_{it_m})\}^{-1} (\partial \mu_{it_m} / \partial \eta_{it_m})^2$ where the variance function $\Upsilon(\cdot)$ is defined in Section 1. With these definitions, we obtain (Harville, 1997) the following approximation:

$$y_{it_m} \overset{ind}{\sim} N\left(\eta_{it_m}, u_{it_m}^{-1}\right) \tag{18}$$

where $(y_{1t_1}, \ldots, y_{nt_m})$ represents the vector of working values, and not the data. For the special case of normal likelihoods, the approximation is exact with $Y_{it_m} = y_{it_m}$.

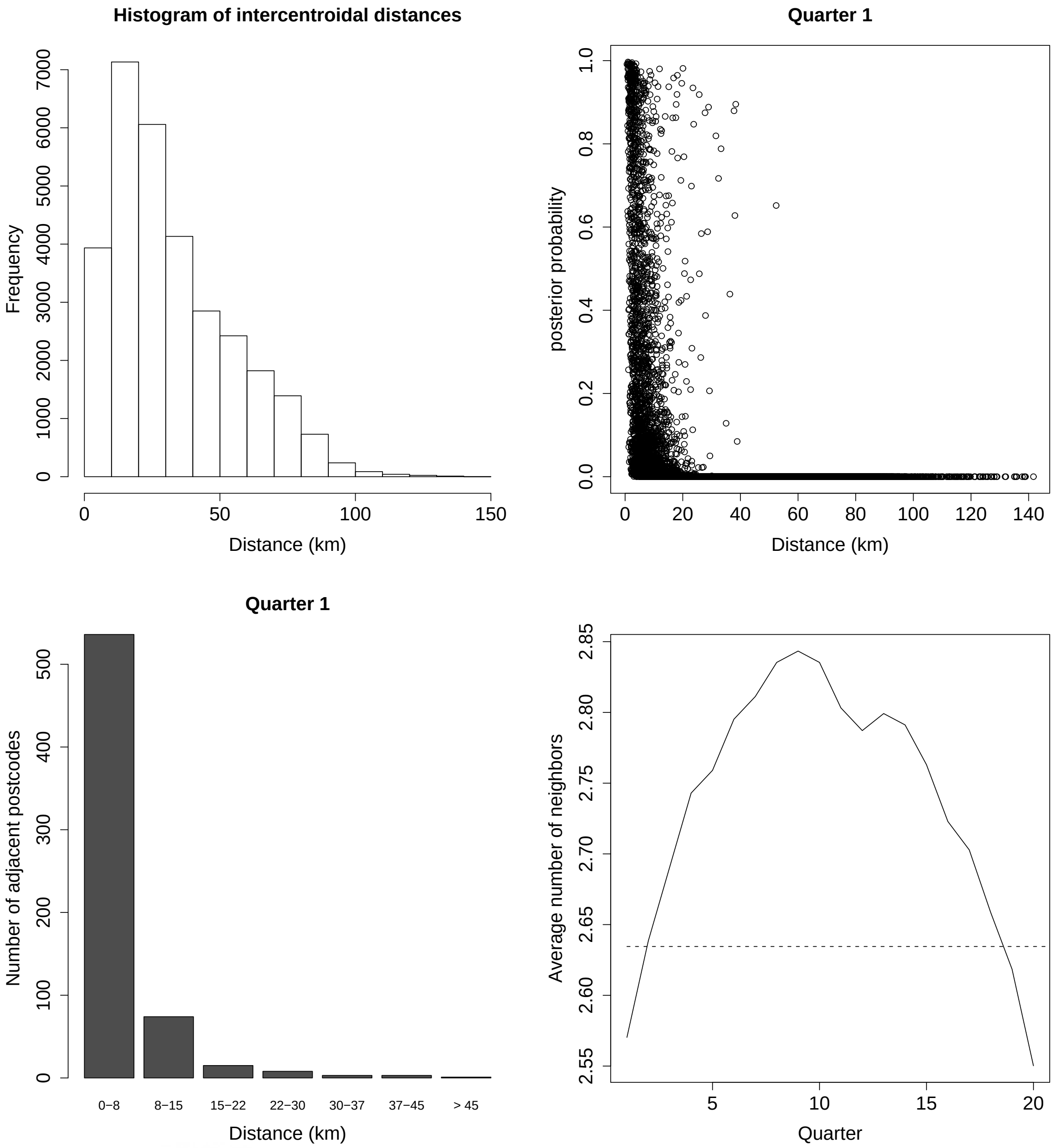

Figure 1: The upper left panel displays the histogram of intercentroidal distances for postcodes belonging to the Sydney Metropolitan Area. For every pair of postcodes belonging to the Sydney Metropolitan Area, the upper right panel plots the estimated posterior probability of adjacency during quarter 1 versus the intercentroidal distance. The bottom left panel displays a barplot of the estimated number of adjacent postcodes for different categories of intercentroidal distances. The solid line in the bottom right panel represents the number of neighbors per postcode for quarters 1 through 20, while the dashed line represents the corresponding value for neighborhood structure (3).

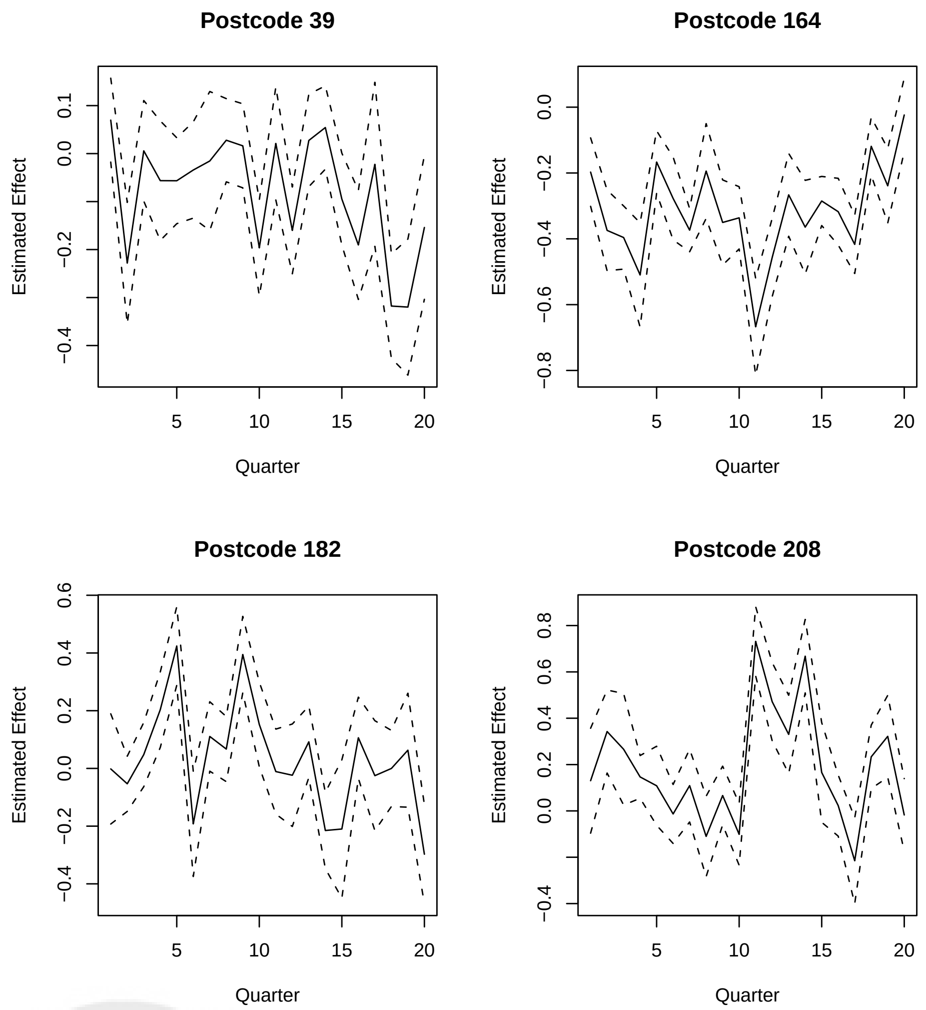

Figure 2: Estimated random effects for quarters 1 through 20 for four randomly chosen postcodes belonging to the Sydney metropolitan area. The dashed lines indicate margins of two posterior standard deviations.

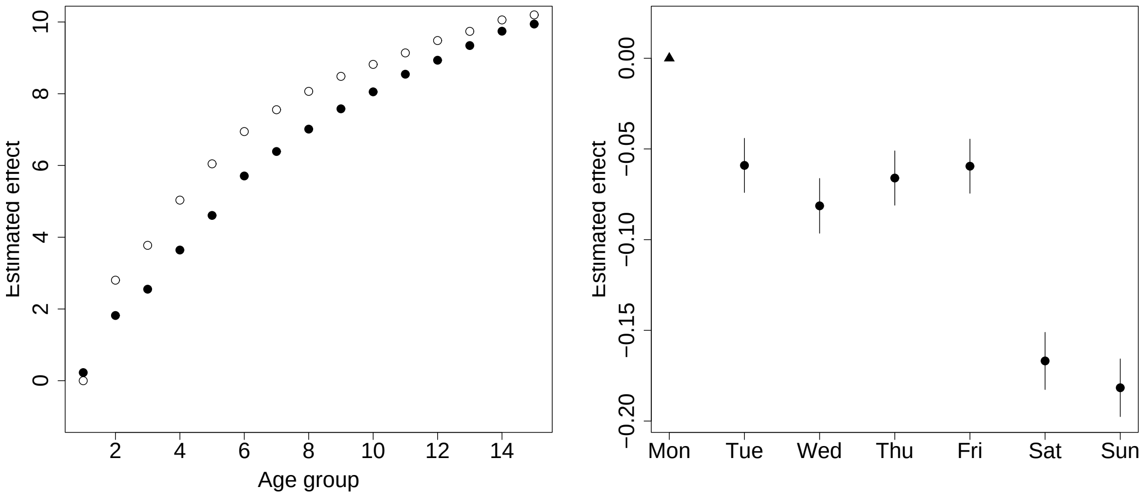

Figure 3: The left panel plots the estimated effects of social category for the 15 age groups of both genders. The open circles represent males and the solid circles represent females. The right panel displays the estimated day-of-week effects. Monday is the reference group. The lines represent intervals of two posterior standard deviations.

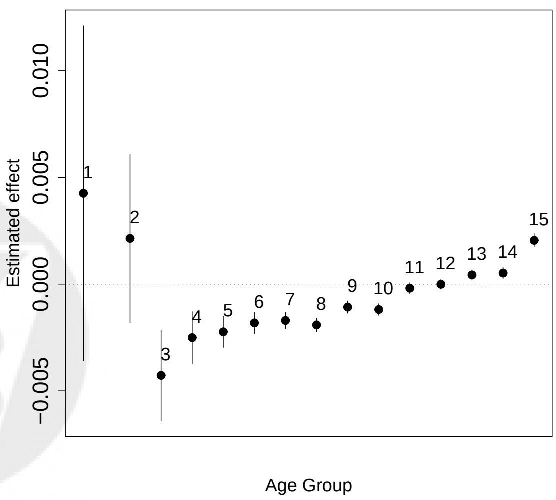

Figure 4: Estimated SEIFA interactions for the 15 age groups of the male subpopulation. The lines represent margins of two posterior standard deviations.