

Bayesian Inference for Directional Conditionally Autoregressive Models

Minjung Kyung* and Sujit K. Ghosh†

Abstract. Counts or averages over arbitrary regions are often analyzed using conditionally autoregressive (CAR) models. The neighborhoods within CAR models are generally determined using only the inter-distances or boundaries between the sub-regions. To accommodate spatial variations that may depend on directions, a new class of models is developed using different weights given to neighbors in different directions. By accounting for such spatial anisotropy, the proposed model generalizes the usual CAR model that assigns equal weight to all directions. Within a fully hierarchical Bayesian framework, the posterior distributions of the parameters are derived using conjugate and non-informative priors. Efficient Markov chain Monte Carlo (MCMC) sampling algorithms are provided to generate samples from the marginal posterior distribution of the parameters. Simulation studies are presented to evaluate the performance of the estimators and are used to compare results with traditional CAR models. Finally the method is illustrated using data sets on local crime frequencies in Columbus, OH and on the elevated blood lead levels of children under the age of 72 months observed in Virginia counties for the year of 2000.

Keywords: Anisotropy; Bayesian estimation; Conditionally autoregressive models; Lattice data; Spatial analysis.

1 Introduction

In many studies, counts or averages over arbitrary regions, known as lattice or area data (Cressie 1993), are observed and spatial analysis is performed. Given a set of geographical regions, observations collected over regions nearer to each other tend to have similar characteristics, as compared to distant regions. In geography, this feature is known as *Tobler's First Law* (Miller 2004). From a statistical perspective, this feature is attributed to the fact that the autocorrelation between pairs of regions tends to be higher for regions near one another than for those farther apart. Thus, this spatial process observed over a lattice or a set of irregular regions is usually modeled using autoregressive models.

In general, given a set of sub-regions S_1, \dots, S_n , we consider a generalized linear model for the aggregated responses, $Y_i = Y(S_i)$, as

$$\begin{aligned} E[\mathbf{Y}|\mathbf{Z}] &= g(\mathbf{Z}) \\ \text{and } \mathbf{Z} &= \boldsymbol{\mu} + \boldsymbol{\eta}, \end{aligned} \tag{1}$$

*Department of Statistics, University of Florida, Gainesville, FL, <mailto:kyung@stat.ufl.edu>

†Department of Statistics, North Carolina State University, Raleigh, NC, <mailto:ghosh@stat.ncsu.edu>

where $\mathbf{Y} = (Y_1, \dots, Y_n) = (Y(S_1), \dots, Y(S_n))$, $\mathbf{Z} = (Z_1, \dots, Z_n) = (Z(S_1), \dots, Z(S_n))$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n) = (\mu(S_1), \dots, \mu(S_n))$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n) = (\eta(S_1), \dots, \eta(S_n))$. Here, $g(\cdot)$ is a suitable link function, $\boldsymbol{\mu}$ represents a vector of large-scale variations (or trends over geographical regions) and $\boldsymbol{\eta}$ denotes a vector of small-scale variations (or spatial random effects) with mean 0 and the variance-covariance matrix $\boldsymbol{\Sigma}$.

Usually, the large scale variations, μ_i 's, are modeled as a deterministic function of some explanatory variables (e.g., latitudes, longitudes and other area level covariates) using a parametric or semiparametric regression model (see [van der Linde *et al.* 1995](#)) involving a finite dimensional parameter $\boldsymbol{\beta}$. However, a more difficult issue is to develop suitable models for the spatial random effects η_i 's, as they are spatially correlated and model specifications are required to satisfy the positive definiteness condition of the induced covariance structure. Popular approaches to estimate such spatial covariances are based on choosing suitable parametric forms so that the $n \times n$ covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\omega})$ is a deterministic function of a finite dimensional parameter $\boldsymbol{\omega}$ and then $\boldsymbol{\omega}$ is estimated from data. It is essential that any such deterministic function should lead to a positive definite matrix for any sample size n and for all allowable parameter values $\boldsymbol{\omega}$. For example, several geostatistical models are available for point-reference observations assuming that the spatial process is weakly stationary and isotropic (see [Cressie 1993](#)). Several extensions to model nonstationary and anisotropic processes have also been recently developed (see [Higdon 1998](#); [Higdon *et al.* 1999](#); [Fuentes and Smith 2001](#); [Fuentes 2002, 2005](#); [Paciorek and Schervish 2006](#); [Hughes-Oliver *et al.* 2009](#)). Once a valid model for $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ is specified, parameter estimates can be obtained using maximum likelihood methods, weighted least squares methods or the posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{\omega})$ (see [Schabenberger and Gotway 2005](#)). Once the point-referenced data are aggregated to the sub-regions (S_i 's), the process representing the aggregated data is modeled using integrals of a spatial continuous process ([Journel and Huijbregts 1978](#)). In this paper, the focus is on the estimation of $\boldsymbol{\omega}$ with the model chosen for $\boldsymbol{\Sigma}$.

In practice there are two distinct approaches to develop models for spatial covariance based on areal data. A suitably aggregated geostatistical model directly specifies a deterministic function of the elements of the $\boldsymbol{\Sigma}$ matrix. On the contrary, the conditional autoregressive models involve specifying a deterministic function of elements of the inverse of the covariance, $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\omega})$ (e.g., see [Besag 1974](#); [Besag and Kooperberg 1995](#)). There have been several attempts to explore the possible connections between these approaches of spatial modeling (e.g., see [Griffith and Csillag 1993](#); [Rue and Tjelmeland 2002](#); [Hrafinkelsson and Cressie 2003](#)). Recently, [Song *et al.* \(2008\)](#) proposed that these Gaussian geostatistical models can be approximately represented by Gaussian Markov Random fields (GMRFs) and vice versa by using spectral densities. However so far most of the GMRFs that are available in literature do not specifically take into account the anisotropic nature of areal data.

In practice, statistical practitioners are accustomed to the exploration of relationships among variables, modeling these relationships with regression and classification models, testing hypothesis about regression and treatment effects, developing meaningful contrasts, and so forth ([Schabenberger and Gotway 2005](#)). For these spatial linear models, we usually assume a correlated relationship among sub-regions and study how a

particular region is influenced by its “neighboring regions” (Cliff and Ord 1981). Therefore, we consider generalized linear mixed models for the area aggregate data. In these models, the latent spatial process Z_i ’s can be treated as a random effect and to model it, conditionally autoregressive (CAR) models (Besag 1974, 1975; Cressie and Chan 1989) and simultaneously autoregressive (SAR) models (Ord 1975) have been used widely.

Gaussian CAR models have been used as random effects within generalized mixed effects models (Breslow and Clayton 1993; Clayton and Kaldor 1987). Because the Gaussian CAR process has the merit that under fairly general regularity conditions (e.g., positivity conditions etc.) lower dimensional conditional Gaussian distributions uniquely determine joint Gaussianity of the spatial CAR processes. Thus, the maximum likelihood (ML) and the Bayesian estimates can be easily obtained. However, one of the major limitations of the CAR model is that the neighbors are formed using some form of a distance metric and the effect of direction is completely ignored. In recent years, there have been some attempts to use different CAR models for different parts of the region. For instance, Reich *et al.* (2007) presented a novel model for periodontal disease and use separate CAR models for separate jaws. White and Ghosh (2008) used a stochastic parameter within the CAR framework to determine effects of the neighbors. Nevertheless, if the underlying spatial process is anisotropic, the magnitude of autocorrelation between the neighbors might be different in different directions. This limitation serves as our main motivation and an extension of the regular CAR process is proposed that can capture such inherent anisotropy. In this article, we focus on developing and exploring more flexible models for the spatial random effects η_i ’s and the newly proposed spatial process will be termed the directional CAR (DCAR) model.

In Section 2, we define the new spatial process and present statistical inferences for the parameters based on samples obtained from the posterior distribution of the parameters using suitable Markov chain Monte Carlo (MCMC) methods. In Section 3, the finite sample performance of the Bayesian estimators are explored using simulated data and the newly proposed DCAR models are compared to the regular CAR models in terms of popular information theoretic criteria and various tests. In Section 4, the proposed method is demonstrated and compared with regular CAR using data sets of the crime frequencies in Columbus, OH and of the elevated blood lead levels of children under the age of 72 months observed in Virginia in the year 2000. Finally, in Section 5, some possible extensions of the DCAR model are discussed.

2 Directional CAR models

In this section, we develop a new model for the latent spatial process, Z_i ’s, described in (1). For simpler illustration and notational simplicity, we assume that S_i are sub-regions in a two-dimensional space, i.e., $S_i \subseteq \mathbb{R}^2$, $\forall i$. However, the proposed model and associated statistical inference presented in this article can easily be extended to higher dimensional data. First, we consider how to define a neighbor structure that depends on the directions between centroids for any pair of sub-regions. Let $s_i = (s_{1i}, s_{2i})$ be a centroid of the sub-region S_i , where s_{1i} corresponds to the horizontal coordinate (x-axis)

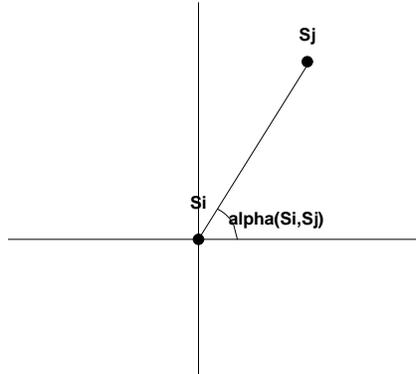


Figure 1: The angle (in radian) α_{ij}

and s_{2i} corresponds to the vertical coordinate (y-axis). The angle (in radians) between S_i and S_j is defined as

$$\alpha_{ij} = \alpha(S_i, S_j) = \begin{cases} \left| \tan^{-1} \left(\frac{s_{2j} - s_{2i}}{s_{1j} - s_{1i}} \right) \right| & \text{if } s_{2j} - s_{2i} \geq 0 \\ -(\pi - \left| \tan^{-1} \left(\frac{s_{2j} - s_{2i}}{s_{1j} - s_{1i}} \right) \right|) & \text{if } s_{2j} - s_{2i} < 0 \end{cases}$$

for all $j \neq i$. We consider directions of neighbors from the centroid of sub-region S_i 's. For example, in Figure 1, S_j is in the north-east (NE) region of S_i and hence $\alpha(S_i, S_j)$ is in $[0, \frac{\pi}{2})$. Let \mathcal{N}_i represent a set of indices (j 's) of neighborhoods for the i th region S_i that are based on some form of distance metric (say as in a regular CAR model). We can now create new sub-regions, for each i , as follows:

$$\begin{aligned} \mathcal{N}_{i1} &= \{j : j \in \mathcal{N}_i, 0 \leq \alpha_{ij} < \frac{\pi}{2}\}, \\ \mathcal{N}_{i2} &= \{j : j \in \mathcal{N}_i, \frac{\pi}{2} \leq \alpha_{ij} < \pi\}, \\ \mathcal{N}_{i3} &= \{j : j \in \mathcal{N}_i, \pi \leq \alpha_{ij} < \frac{3}{2}\pi\}, \\ \mathcal{N}_{i4} &= \{j : j \in \mathcal{N}_i, \frac{3}{2}\pi \leq \alpha_{ij} < 2\pi\}. \end{aligned}$$

These directional neighborhoods should be chosen carefully so that, for each i , they form a *clique*. Recall that a *clique* is any set of sites which either consists of a single site or else in which every site is a neighbor of every other site in the set (Besag 1974). This would allow us to show the existence of the spatial process by using the Hammersley-Clifford Theorem (Besag 1974, p.197-198) and to derive the finite dimensional joint

distribution of the process using only a set of (lower dimensional) full conditional distributions. For instance, if $j \in \mathcal{N}_{i1}$, then it should be ensured that $i \in \mathcal{N}_{j3}$. For the above four neighbor sets, we can combine each pair of the diagonally opposite neighbor sets to form a new neighborhood. It means that we can create $\mathcal{N}_{i1}^* = \mathcal{N}_{i1} \cup \mathcal{N}_{i3}$, and $\mathcal{N}_{i2}^* = \mathcal{N}_{i2} \cup \mathcal{N}_{i4}$ for $i = 1, \dots, n$. Now it is easy to check that if $j \in \mathcal{N}_{i1}^*$, then $i \in \mathcal{N}_{j1}^*$. Thus, we redefine two subsets of \mathcal{N}_i 's as follows:

$$\begin{aligned} \mathcal{N}_{i1}^* &= \{j : j \in \mathcal{N}_i \text{ and } (0 \leq \alpha_{ij} < \frac{\pi}{2} \text{ or } \pi \leq \alpha_{ij} < \frac{3}{2}\pi)\} \\ \mathcal{N}_{i2}^* &= \{j : j \in \mathcal{N}_i \text{ and } (\frac{\pi}{2} \leq \alpha_{ij} < \pi \text{ or } \frac{3}{2}\pi \leq \alpha_{ij} < 2\pi)\}. \end{aligned} \tag{2}$$

Then, each of \mathcal{N}_{i1}^* and \mathcal{N}_{i2}^* forms a clique and it can be shown that $\mathcal{N}_i = \mathcal{N}_{i1}^* \cup \mathcal{N}_{i2}^*$.

A centroid of the sub-region S_i might not be given or available in some situations, for example, neighbor relationships are defined via adjacencies instead of distances between centroids. In this case, the directions of neighbors for each sub-region S_i are not clear. One suggestion in this situation is that we might define the directions of neighbors intuitively based on the direction of adjacencies. For this topic, we need further study. However, throughout this paper, we assume that we can define the directions of neighbors for each sub-regions.

Before fitting a DCAR model, we would need to define these directional neighborhood just as we need to define the CAR weights before fitting a CAR model. Note that with defined directional adjacency, we can easily rotate the distance category boundaries while maintaining the clique. For example, we can easily define the different weights to the neighbors in the north-south region compared to those in the east-west.

The above scheme of creating new neighborhoods based on the inter-angles, α_{ij} 's can be extended beyond just two sub-neighborhoods so that each of the new sub-neighborhood forms a clique. For example, we can extend the directional cliques with 4 sub-sets of neighborhoods as

$$\begin{aligned} \mathcal{N}_{i1}^* &= \{j : j \in \mathcal{N}_i \text{ and } (0 \leq \alpha_{ij} < \frac{\pi}{4} \text{ or } \pi \leq \alpha_{ij} < \frac{5}{4}\pi)\} \\ \mathcal{N}_{i2}^* &= \{j : j \in \mathcal{N}_i \text{ and } (\frac{\pi}{4} \leq \alpha_{ij} < \frac{\pi}{2} \text{ or } \frac{5}{4}\pi \leq \alpha_{ij} < \frac{3}{2}\pi)\} \\ \mathcal{N}_{i3}^* &= \{j : j \in \mathcal{N}_i \text{ and } (\frac{\pi}{2} \leq \alpha_{ij} < \frac{3}{4}\pi \text{ or } \frac{3}{2}\pi \leq \alpha_{ij} < \frac{7}{4}\pi)\} \\ \mathcal{N}_{i4}^* &= \{j : j \in \mathcal{N}_i \text{ and } (\frac{3}{4}\pi \leq \alpha_{ij} < \pi \text{ or } \frac{7}{4}\pi \leq \alpha_{ij} < 2\pi)\}. \end{aligned}$$

However, it should be noted that anisotropic specifications for the geostatistical covariance functions are quite different from the directional specification of neighborhood cliques used to define the inverse of the covariance. In this regard, the directional adjustments within the CAR framework allow the anisotropy parameters to capture the local (neighboring) directional effects whereas the anisotropy parameters of a geostatistical model generally capture the overall global directional effects. Finally, it is possible to

increase the number of sub-neighborhoods to more than 2 or 4 sub-neighborhoods. However, we cautiously note that if we keep increasing the number of sub-neighborhoods, the number of parameter increases whereas the amount of observations available within a sub-neighborhood decreases. Thus, we need to restrict the number of sub-neighborhoods by introducing some form of a penalty term (e.g., via the prior distributions of anisotropy parameters) and use some form of information criterion to choose the number of sub-neighborhoods. This is an important but open issue within our DCAR framework. Hence for the rest of the article, for simplicity, we restrict our attention to the case with only two sub-neighborhoods as described in (2).

Based on subsets of the associated neighborhoods, \mathcal{N}_{i1}^* and \mathcal{N}_{i2}^* , we can construct directional weight matrices $\mathbf{W}^{(1)} = ((w_{ij}^{(1)}))$ and $\mathbf{W}^{(2)} = ((w_{ij}^{(2)}))$, respectively. For instance, we define the directional proximity matrices as $w_{ij}^{(1)} = 1$ if $j \in \mathcal{N}_{i1}^*$ and $w_{ij}^{(2)} = 1$ if $j \in \mathcal{N}_{i2}^*$. Notice that $\mathbf{W} = \mathbf{W}^{(1)} + \mathbf{W}^{(2)}$ reproduces the commonly used proximity matrix as in a regular CAR model.

In order to model the large-scale variations, we assume a canonical generalized linear model, $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, where \mathbf{x}_i 's are vectors of predictor variables specific to the sub-region S_i and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)^T$ is a vector of regression coefficients. Notice that nonlinear regression functions, including smoothing splines and polynomials, can be re-written in the above canonical form (e.g., see Wahba 1977; van der Linde *et al.* 1995). From model (1) it follows that

$$\mathbb{E}[\mathbf{Z}] = \mathbf{X}\boldsymbol{\beta} \quad \text{and} \quad \text{Var}[\mathbf{Z}] = \boldsymbol{\Sigma}(\boldsymbol{\omega}), \quad (3)$$

where $\boldsymbol{\omega}$ denotes the vector of spatial autocorrelation parameters and other variance components. Notice that along with (3), the model (1) can be used for discrete responses using a generalized linear model framework (Schabenberger and Gotway 2005, p.353). Now, we develop a model for $\boldsymbol{\Sigma}(\boldsymbol{\omega})$ that accounts for anisotropy.

Let δ_1 and δ_2 denote the directional spatial effects corresponding to \mathcal{N}_{i1} 's and \mathcal{N}_{i2} 's, respectively. We define the distribution of Z_i conditional on the rest of Z_j 's for $j \neq i$ using only the first two moments:

$$\begin{aligned} \mathbb{E}[Z_i | Z_j = z_j, j \neq i, \mathbf{x}_i] &= \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{k=1}^2 \delta_k \sum_{j=1}^n w_{ij}^{(k)} (z_j - \mathbf{x}_j^T \boldsymbol{\beta}) \\ \text{Var}[Z_i | Z_j = z_j, j \neq i, \mathbf{x}_i] &= \frac{\sigma^2}{m_i}, \end{aligned} \quad (4)$$

where $w_{ij}^{(k)} \geq 0$ and $w_{ii}^{(k)} = 0$ for $k = 1, 2$ and $m_i = \sum_{j=1}^n w_{ij}$.

The joint distribution based on a given set of full conditional distributions can be derived using *Brook's Lemma* (Brook 1964) provided the positivity condition is satisfied (e.g., see Besag 1974; Besag and Kooperberg 1995). For the DCAR model, by construction, it follows that each of \mathcal{N}_{i1}^* and \mathcal{N}_{i2}^* defined in (2) forms a clique for $i = 1, \dots, n$. Thus, it follows from the Hammersley-Clifford Theorem that the latent spatial process Z_i of a DCAR model exists and is a Markov Random Field (MRF). Therefore, we can

derive the exact joint distribution of the DCAR process, Z_i 's, by assuming that each of the full conditional distribution is a Gaussian distribution.

2.1 Gaussian DCAR models

The Gaussian CAR model has been used widely as a suitable model for the latent spatial process Z_i . In this section, to derive the joint distribution of the Z_i 's from a set of given full conditional distributions, we use Brook's Lemma.

Assume that the full conditional distributions of Z_i 's are given as

$$Z_i | Z_j = z_j, j \neq i, \mathbf{x}_i \sim N \left(\mathbf{x}_i^T \boldsymbol{\beta} + \sum_{k=1}^2 \delta_k \sum_{j=1}^n w_{ij}^{(k)} (z_j - \mathbf{x}_j^T \boldsymbol{\beta}), \frac{\sigma^2}{m_i} \right), \tag{5}$$

where $w_{ij}^{(k)}$ for $k = 1, 2$ are the directional weights. It can be shown that this latent spatial DCAR process Z_i 's is a MRF. Thus, by Brook's Lemma and the Hammersley-Clifford Theorem, it follows that the finite dimensional joint distribution is a multivariate Gaussian distribution given by

$$\mathbf{Z} \sim N_n \left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \left(\mathbf{I} - \delta_1 \mathbf{W}^{(1)} - \delta_2 \mathbf{W}^{(2)} \right)^{-1} \mathbf{D} \right),$$

where $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ and $\mathbf{D} = \text{diag}(\frac{1}{m_1}, \dots, \frac{1}{m_n})$. For simplicity, we denote the variance-covariance matrix of the DCAR process by $\boldsymbol{\Sigma}_{\mathbf{Z}} \equiv \sigma^2 (\mathbf{I} - \delta_1 \mathbf{W}^{(1)} - \delta_2 \mathbf{W}^{(2)})^{-1} \mathbf{D}$.

For a proper Gaussian model, the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ needs to be positive definite. First, notice that if we use the standardized directional proximity matrices $\tilde{\mathbf{W}}^{(k)} = ((\tilde{w}_{ij}^{(k)} = \frac{w_{ij}^{(k)}}{m_i}))$, $k = 1, 2$, it can be easily shown that $\boldsymbol{\Sigma}_{\mathbf{Z}}$ is symmetric. Thus, the finite dimensional joint distribution is given by

$$\mathbf{Z} \sim N_n \left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \left(\mathbf{I} - \delta_1 \tilde{\mathbf{W}}^{(1)} - \delta_2 \tilde{\mathbf{W}}^{(2)} \right)^{-1} \mathbf{D} \right), \tag{6}$$

Next, we derive a sufficient condition that ensures that the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$ is non-singular and hence making it a positive definite matrix. As \mathbf{D} is a diagonal matrix, we only require suitable conditions on $\tilde{\mathbf{W}}^{(k)}$ and on δ_k for $k = 1, 2$. The following results provides a sufficient condition:

Lemma 1. Let $\mathbf{A} = (a_{ij})$ be a $n \times n$ symmetric matrix. If $a_{ii} > \sum_{j \neq i} |a_{ij}|$ for all i , then \mathbf{A} is positive definite.

Proof: See Ortega 1987, P.226. \square

Lemma 2. Let $\mathbf{A} = \mathbf{I} - \sum_{k=1}^K \delta_k \tilde{\mathbf{W}}^{(k)}$ be an $n \times n$ matrix where $\sum_{k=1}^K \tilde{\mathbf{W}}^{(k)}$ is a symmetric matrix with non-negative entries, diagonal $\mathbf{0}$ and each row sum equal to unity. If $\max_{1 \leq k \leq K} |\delta_k| < 1$, then the matrix \mathbf{A} is positive definite.

Proof: Let a_{ij} denote the (i, j) th element of \mathbf{A} . Notice that for each $i = 1, 2, \dots, n$, we have

$$\sum_{j \neq i} |a_{ij}| = \sum_{j \neq i} \left| \sum_{k=1}^K \delta_k \tilde{w}_{ij}^{(k)} \right| \leq \sum_{k=1}^K |\delta_k| \sum_{j \neq i} \tilde{w}_{ij}^{(k)} < \sum_{k=1}^K \sum_{j \neq i} \tilde{w}_{ij}^{(k)} = 1 = a_{ii}$$

Hence it follows from Lemma 1 that \mathbf{A} is positive definite. \square

Notice that when $\delta_1 = \delta_2 = \rho$, $\text{DCAR}(\delta_1, \delta_2, \sigma^2)$ reduces to $\text{CAR}(\rho, \sigma^2)$ and hence the regular CAR model is nested within the DCAR model provided we use a prior that puts positive mass on the line $\delta_1 = \delta_2$. The next step of our statistical analysis is to estimate the unknown parameters of the DCAR model based on the observed responses and the explanatory variables, so that it enables us to stabilize estimates within the regions using the estimated spatial correlation. In the next section, we discuss Bayesian methods for the spatial autoregressive models.

2.2 Parameter estimation using Bayesian methods

With the Gaussian DCAR model of the latent spatial process Z_i 's, we describe how to estimate parameters and associated measures of uncertainties based on Bayesian methods.

Bayesian inference about the unknown parameters has been considered for statistical models for which the likelihood functions are analytically intractable, because of possibly high-dimensional parameters or due to the fact that the likelihood function involves high-dimensional integrations (e.g., when Y_i 's are discrete valued). In the Gaussian DCAR model, because the likelihood function may involve high-dimensional integration, posterior estimation is not easy to achieve analytically. In particular, the joint posterior density of δ_1 and δ_2 does not have a closed form. Also, when a generalized mixed model is used with the random spatial effects having a DCAR model, analytical exploration of the posterior distribution becomes almost prohibitive. Thus, the Gaussian DCAR model leads to an intractable posterior density and numerical methods are needed for inference about unknown parameters.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\delta}^T)^T$, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2)$. The posterior density $\pi(\boldsymbol{\theta}|\mathbf{z})$ is proportional to the product of the prior distribution $\pi(\boldsymbol{\theta})$ of unknown parameters and the sampling density of data \mathbf{Z} given $\boldsymbol{\theta}$. Therefore, by using Markov chain Monte Carlo (MCMC) methods, we can obtain samples from the path of Markov chains whose stationary density is the posterior density.

For the DCAR process Z_i 's, under the joint multivariate Gaussian distribution, the likelihood function is given by

$$L(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}) \propto |\sigma^2 \mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\}, \quad (7)$$

where $\mathbf{A}^*(\boldsymbol{\delta}) = \mathbf{I} - \delta_1 \tilde{\mathbf{W}}^{(1)} - \delta_2 \tilde{\mathbf{W}}^{(2)}$ and $\mathbf{D} = \text{diag}(\frac{1}{m_1}, \dots, \frac{1}{m_n})$. We consider a class of

prior distributions that ensure that the posterior distribution is proper even when the priors are improper. A class of such prior distribution is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta}) &\equiv 1 \\ \pi(\sigma^2|\boldsymbol{\delta}) &\propto \left(\frac{1}{\sigma^2}\right)^{a+1} e^{-\frac{b}{\sigma^2}} \quad a, b > 0 \text{ and} \\ \pi(\boldsymbol{\delta}) &= \frac{1}{4} \mathbf{I}[\max(|\delta_1|, |\delta_2|) < 1]. \end{aligned}$$

As the prior distribution of $\boldsymbol{\beta}$ is not proper, we need to ensure that the posterior is proper. Given prior distributions above, the joint posterior distribution of $\boldsymbol{\theta}$ can be shown to have the following form:

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}) &\propto L(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z})\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\sigma^2|\boldsymbol{\delta})\pi(\boldsymbol{\delta}) \\ &\propto (\sigma^2)^{-n/2-a-1} |\mathbf{A}^*(\boldsymbol{\delta})^{-1}\mathbf{D}|^{-1/2} \\ &\quad \exp\left\{-\frac{1}{2\sigma^2} [(z - \mathbf{X}\boldsymbol{\beta})^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta})(z - \mathbf{X}\boldsymbol{\beta}) + 2b]\right\} \\ &\quad \times \mathbf{I}[\max(|\delta_1|, |\delta_2|) < 1]. \end{aligned} \tag{8}$$

Here, if $\boldsymbol{\delta}$ is known, like the regular posterior distribution in a regression model, the posterior distribution of $\boldsymbol{\beta}$ given σ^2 and $\boldsymbol{\delta}$ is a Gaussian distribution with complicated form of the mean and variance, and the posterior of σ^2 given $\boldsymbol{\delta}$ is an inverse gamma distribution. For the Gaussian DCAR model, assume that the design matrix \mathbf{X} is full rank and the variance-covariance matrix $\boldsymbol{\Sigma}_Z$ is positive definite. Thus, based on the sufficient conditions given by Sun *et al.* (1999), one can easily deduce that the posterior distribution of $\boldsymbol{\theta}|\mathbf{z}$ is proper.

We can also consider the conditionally conjugate priors for $\boldsymbol{\beta}$ and σ^2 . Given the values of $\boldsymbol{\delta}$, the likelihood function of the DCAR model (7) is like a regression model with mean $\mathbf{X}\boldsymbol{\beta}$ and variance-covariance $\sigma^2\mathbf{A}^*(\boldsymbol{\delta})^{-1}\mathbf{D}$. Thus, a conditional conjugate prior given $\boldsymbol{\delta}$ can be considered in two stages according to

$$\begin{aligned} \boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta} &\sim N(\boldsymbol{\beta}_0, \sigma^2\mathbf{A}^*(\boldsymbol{\delta})^{-1}\mathbf{D}) \\ \sigma^2|\boldsymbol{\delta} &\sim IG(a_0, b_0). \end{aligned}$$

Given the conditionally conjugate prior distributions and a marginal prior for $\boldsymbol{\delta}$, the joint posterior distribution of $\boldsymbol{\theta}$ is given by

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}) &\propto L(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z})\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\delta})\pi(\sigma^2|\boldsymbol{\delta})\pi(\boldsymbol{\delta}) \\ &\propto (\sigma^2)^{-(n/2+p/2+a_0)-1} \times \mathbf{I}[\max(|\delta_1|, |\delta_2|) < 1] \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2} \left[\tilde{s} \right. \right. \\ &\quad \left. \left. + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \left(\mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X} + (\mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D})^{-1} \right) (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) \right] \right\} \end{aligned} \tag{9}$$

where

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X} + (\mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D})^{-1})^{-1} \left((\mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D})^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X} \hat{\boldsymbol{\beta}} \right), \\ \tilde{s} &= \widehat{\sigma}^2 (n - p) + 2b_0 + \left(\boldsymbol{\beta}_0 - \tilde{\boldsymbol{\beta}} \right)^T (\mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D})^{-1} \boldsymbol{\beta}_0 + \left(\hat{\mathbf{b}} - \tilde{\boldsymbol{\beta}} \right) \mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X} \hat{\boldsymbol{\beta}}, \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{z} \\ \widehat{\sigma}^2 &= \frac{(\mathbf{z} - \mathbf{X} \hat{\mathbf{b}})^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) (\mathbf{z} - \mathbf{X} \hat{\mathbf{b}})}{n - p}\end{aligned}$$

and p is the dimension of $\boldsymbol{\beta}$. Then the conditional posterior distributions of the parameters are given by

$$\begin{aligned}\boldsymbol{\beta} | \sigma^2, \boldsymbol{\delta}, \mathbf{Z} &\sim N(\tilde{\boldsymbol{\beta}}, \mathbf{X}^T \mathbf{D}^{-1} \mathbf{A}^*(\boldsymbol{\delta}) \mathbf{X} + (\mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D})^{-1}) \text{ and} \\ \sigma^2 | \boldsymbol{\delta}, \mathbf{Z} &\sim IG((n/2 + p/2 + a_{10}), \tilde{s}).\end{aligned}$$

However, as we discussed above, there is no closed form for the posterior distribution of $\boldsymbol{\delta}$. Therefore, we need numerical methods to obtain the posterior summaries (e.g., suitable moments) of $\boldsymbol{\theta}$.

In order to obtain samples from the path of the Markov chains, we need to consider the starting values of each parameters. In the DCAR model of the latent spatial process Z_i 's, the parameter space Θ of $\boldsymbol{\theta}$ can be defined as $\Theta = \{\boldsymbol{\theta} | \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 \in (0, \infty), \mathbf{I}[\max(|\delta_1|, |\delta_2|) < 1]\}$. Within Θ , we can choose several starting points for chains and run parallel chains. After burn-in, we obtain approximate samples from the posterior density $\pi(\boldsymbol{\theta} | \mathbf{z})$.

2.3 A simulation study

In order to study the finite sample performance of Bayesian estimators, we conduct a simulation study. In this simulation study, we focus on the behavior of Gaussian DCAR model of the latent spatial process $\mathbf{Z} = (Z_1, \dots, Z_n)$ in (6).

Mardia and Marshall (1984) conducted a simulation study with 10×10 unit spacing lattice, based on samples generated from a normal distribution with mean zero and a spherical covariance model. The sampling distribution of the MLE's of the parameters were studied based on 300 Monte Carlo samples. Following a similar setup, for our simulation study, we selected a 15×15 unit spacing lattice and generated $N = 100$ data sets each of size $n = 225$ from a multivariate normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and the variance-covariance matrix $\sigma^2 \mathbf{A}^*(\boldsymbol{\delta})^{-1} \mathbf{D}$, where $\mathbf{A}^*(\boldsymbol{\delta}) = \mathbf{I} - \delta_1 \tilde{\mathbf{W}}^{(1)} - \delta_2 \tilde{\mathbf{W}}^{(2)}$. The \mathbf{X} matrix was chosen to consist of the coordinates of latitude and longitude in addition to a column of ones to represent an intercept. The true value of the parameters were fixed at $\boldsymbol{\beta} = (1, -1, 2)^T$ and $\sigma^2 = 2$. For the above mentioned DCAR model, to study the behavior of posterior distributions for δ_1 and δ_2 , we consider four different test cases

of δ 's:

Case 1: $\delta_1 = -0.95$ & $\delta_2 = -0.97$

Case 2: $\delta_1 = -0.30$ & $\delta_2 = 0.95$

Case 3: $\delta_1 = -0.95$ & $\delta_2 = 0.97$

Case 4: $\delta_1 = 0.95$ & $\delta_2 = 0.93$.

Following Lemma 2, we restrict our choice of δ_1 and δ_2 to satisfy the sufficient condition, $\max_{1 \leq k \leq K} |\delta_k| < 1$. Thus, for the near boundary values of δ_1 and/or δ_2 , like -0.95 and 0.93 , there might be some unexpected behavior of the sampling distribution. Thus, we generate data sets with two negative near boundary weights of δ (Case 1) and two positive near boundary weights of δ (Case 4) in order to explore extreme cases, respectively. In our applications we have found that such extreme values are quite common within CAR or DCAR models (see Section 4). Besag and Kooperberg (1995) also discussed similar situations in their paper. We also consider settings with one positive near boundary weight assigned to one direction and one negative near boundary weight assigned to the other direction for extremely different weighted situations (Case 3). Moreover, we give somewhat mild weight to one direction and positive boundary weight to a different direction to study the behavior of a strong positive spatial effect in one direction only (Case 2). Thus, with extreme boundary values of δ , we study the sampling distributions of the directional spatial effect parameters.

As we discussed in Section 2.2, for the Bayesian estimates, we consider three sets of initial values and run three parallel chains. We use a burn-in of $B = 1000$ for each of the three chains followed by $M = 2000$ iterations. This scheme produces a sample of 6000 (correlated) values from the joint posterior distribution of the parameter vector.

As Bayesian estimation involves the use of computationally intensive MCMC methods, we studied the finite sample performance of Bayes estimates with only $N = 100$ repetitions. The (coordinatewise) posterior median of the parameter vector is used as a Bayes estimate because of its robustness as compared to the posterior mean, especially when the posterior distribution is skewed. Also, for each coordinate of the parameter vector, we computed a 95% equal tail *credible set* (CS) using 2.5 and 97.5 percentiles as an interval estimate. Then, we computed the 95% nominal coverage probability (CP) by using the following rule: $95\% \text{ CP} = \frac{1}{N} \sum_{i=1}^N I(\theta_0 \in 95\% \text{ CS})$.

We summarize the sampling distribution of parameters numerically in Table 1. The bias represents the empirical bias of posterior medians as compared to the true value, the Monte Carlo Standard Error (MCSE) is the standard error of the posterior medians, the p-value is based on testing the average of posterior medians against the true value and 95% CP represents the percentage of times the true value was included within the 95% CS. All of these summaries are based on 100 replications. We observe that for all the four choices of δ , there are no significant biases in these Bayesian estimates with MC repetitions of size 100. (e.g., all p-values are bigger than 0.18). When the true δ_1 or δ_2 is positive, the bias of the Bayesian estimate tends to be slightly negative, except for δ_2 in Case 4. For Case 3, the nominal 95% coverage probabilities (CP's) of δ_1 and δ_2 are away from 0.95 and the MCSE's are not small. Also, from Figure 2, we observe that

	δ_1	δ_2	σ^2	δ_1	δ_2	σ^2
True	-0.95	-0.97	2.00	-0.30	0.95	2.00
bias	0.23	0.27	0.17	0.15	-0.21	0.11
MCSE	0.22	0.20	0.20	0.32	0.18	0.21
P-value	0.29	0.19	0.40	0.64	0.24	0.60
95% CP	0.99	1.00	0.92	0.97	0.93	0.92
	δ_1	δ_2	σ^2	δ_1	δ_2	σ^2
True	-0.95	0.97	2.00	0.95	0.93	2.00
bias	0.20	-0.18	0.26	-0.01	0.03	-0.05
MCSE	0.15	0.17	0.21	0.03	0.03	0.37
P-value	0.18	0.28	0.21	0.74	0.28	0.89
95% CP	0.89	0.93	0.84	0.98	0.89	0.67

Table 1: Finite sample performance of posterior estimates of the parameters of DCAR models (based on 100 replications).

for the extremely differently positively (δ_1) and negatively (δ_2) weighted situations, the posterior estimates seem to estimate true values with somewhat less precision. However, the distribution of δ is skewed when the true value is in the boundary. It might be the reason why we get somewhat larger values of MCSEs. For Case 4, the nominal 95% CP's of δ_1 and δ_2 are 0.98 and 0.89, respectively, and biases and MCSE's are smaller than those for any other cases. Thus, Bayesian methods based on posterior medians tend to estimate the true value quite well even when the true values of δ_1 and δ_2 are near the positive boundary. The higher than nominal coverage probability of the Bayesian interval estimates based on equal tail CS may be due to the skewness of the sampling distribution that we have observed in our empirical studies. Alternatively, a 95% HPD interval can be obtained using the algorithm of [Hyndman \(1996\)](#). It was observed that the posterior distributions of δ_1 or δ_2 are skewed to the right and to the left for the negative extreme value and the positive extreme value, respectively.

The bias of the posterior median of σ^2 tends to be slightly negative for Case 4, but for other cases, the biases are slightly positive. Again note that these biases are not statistically significant (all four p-values are greater than 0.21). Thus, in these cases, Bayesian estimation tends to estimate the true value quite well. However, for Case 4, the MCSE of the Bayesian estimates of σ^2 is bigger than those for other cases and the nominal coverage is only 0.67 as compared to a targeted value of 0.95.

For the estimation of the β 's, the estimates had small MCSE, and did not have any significant bias, except for β_0 in Case 4. Also we observed that the posterior distributions of the β 's are fairly symmetric (results not reported due to lack of space but are available in [Kyung 2006](#)).

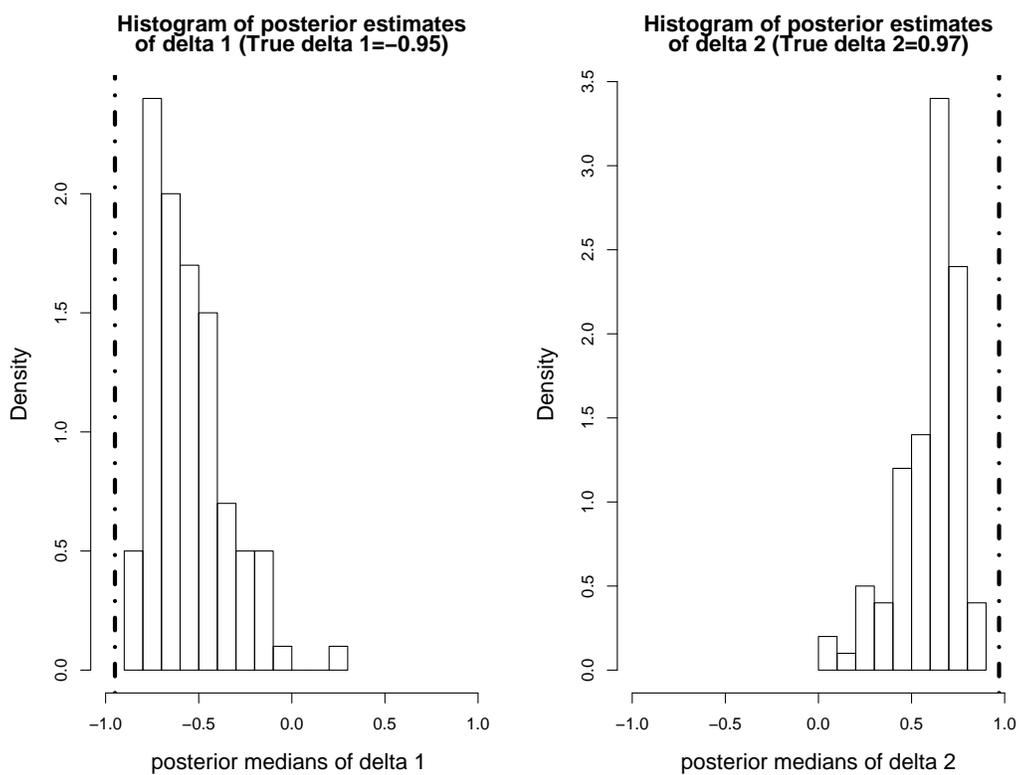


Figure 2: Histogram of 100 posterior estimate of δ_1 and δ_2 based on the DCAR process data with true $\delta_1 = -0.95$ and $\delta_2 = 0.97$ (Posterior median of $M = 6000$ Gibbs samples with 100 replications).

3 Comparing the performances of DCAR and CAR models using an information criterion

In Section 2.1 we have shown that the DCAR model is a generalization of the CAR model and hence the DCAR model is expected to provide a reasonable fit to a given data set possibly at the cost of loss of efficiency, especially when the data arise from a CAR model. So it is of interest to explore the loss (gain) in efficiency of a DCAR model over the regular CAR model when the data arises from a CAR (DCAR) model. There are several criteria (e.g., information criteria, cross-validation measures, hypotheses tests, etc.) to compare the performances across several competing models. Given the popularity of the Deviance Information Criterion (DIC) originally proposed by Spiegelhalter *et al.* (2002) we use DIC to compare the performance of fitting DCAR and CAR models to data generated from a CAR model and then also from a DCAR model. Another advantage of using DIC is that this criterion is already available within the WinBUGS software. To calculate DIC, first we define the deviance as

$$D(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}) + 2 \log h(\mathbf{z}),$$

where $h(\mathbf{z})$ is standardizing function of the data \mathbf{z} only and remains the same across all competing models. In general it is difficult to find the normalizing function $h(\mathbf{z})$ for models involving spatial random effects. However given that we are interested in the differences of DIC between the models, we may use the following definition of deviance by dropping the $h(\mathbf{z})$ term:

$$D(\boldsymbol{\theta}) = -2 \log L(\boldsymbol{\theta}|\mathbf{X}, \mathbf{z}),$$

as the normalizing term will cancel anyway when we take the difference between two DIC's with same h function. Based on the deviance, the definition of the effective number of parameters, denoted by p_D , is defined as:

$$p_D = E[D(\boldsymbol{\theta})|\mathbf{z}] - D(E[\boldsymbol{\theta}|\mathbf{z}]) = \bar{D} - D(\bar{\boldsymbol{\theta}}),$$

where $\bar{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\mathbf{y}]$ is the posterior mean of $\boldsymbol{\theta}$. The DIC is then defined as

$$DIC = D(\bar{\boldsymbol{\theta}}) + 2p_D.$$

In theory, we select the model with the smaller DIC values. DIC and p_D are easily computed by MCMC methods. We consider two cases based on data generated from a (i) CAR model and (ii) DCAR model.

3.1 Results based on data generated from CAR model

With samples from a Gaussian CAR process, we fit both CAR and DCAR models, respectively. Notice that if there is no directional difference in the observed spatial data, then the estimate of δ_1 should be very similar to the estimate of δ_2 . Thus we might expect very similar estimates for δ_1 and δ_2 based on a sample from a CAR process

	δ_1	δ_2	ρ	δ_1	δ_2	ρ
True	-0.95	-0.95	-0.95	-0.25	-0.25	-0.25
bias	0.01	0.01	0.02	0.14	0.15	0.08
MCSE	0.02	0.02	0.00	0.29	0.26	0.33
P-value	0.67	0.66	0.00	0.63	0.57	0.81
95% CP	1.00	1.00	1.00	1.00	1.00	0.99
	δ_1	δ_2	ρ			
True	0.00	0.00	0.00			
bias	-0.02	0.00	-0.01			
MCSE	0.30	0.26	0.30			
P-value	0.95	1.00	0.97			
95% CP	1.00	1.00	1.00			
	δ_1	δ_2	ρ	δ_1	δ_2	ρ
True	0.25	0.25	0.25	0.95	0.95	0.95
bias	-0.10	-0.09	0.01	-0.15	-0.06	-0.06
MCSE	0.31	0.26	0.32	0.11	0.08	0.04
P-value	0.85	0.72	0.98	0.17	0.45	0.14
95% CP	0.99	1.00	0.98	1.00	1.00	1.00

Table 2: Performance of Bayesian estimates of δ_1 's, δ_2 's (DCAR) and ρ 's (CAR) based on data generated from CAR model

because $CAR(\rho, \sigma^2) = DCAR(\rho, \rho, \sigma^2)$. In fact, it might be a good idea to use a prior on (δ_1, δ_2) which allows for a positive mass on the diagonal line $\delta_1 = \delta_2$ to capture a CAR model with positive probability. To study the performance of the model as function of the key parameter ρ of the CAR model, we consider five different values of ρ : Case 1: $\rho = -0.95$, Case 2: $\rho = -0.25$, Case 3: $\rho = 0$, Case 4: $\rho = 0.25$ and Case 5: $\rho = 0.95$. For each case, we generate 100 sets of data each of sample size $n = 225$ from a CAR model with ρ taking values as one of above five cases, while the other parameters (β and σ) are fixed at their true values (see Section 2.3). First we compare the posterior estimates of the ρ when we fitted a CAR model and that of δ_1 and δ_2 when we fitted a DCAR model to data generated from one of the five CAR models. In Table 2 we compare the bias (of the posterior median), Monte Carlo Standard Error (MCSE) of the posterior median, the p-value for testing the null hypothesis that the (MC) average of the 100 posterior medians is the same as the true value, and the 95% nominal coverage probability (CP) of the 95% posterior intervals constructed by computing 2.5% and 97.5% percentiles of the posterior distribution of the parameters.

From the results (presented under the columns ρ in Table 2) based on the posterior estimates (median and 95% equal-tail intervals) obtained by fitting a CAR model, we observe that for all cases, the biases of ρ are slightly positive except Case 5, but such empirical biases are not statistically significant (all p-values being bigger than 0.06). The nominal 95% CP's of ρ 's are higher than their targeted value of 0.95 for all cases. For all cases, the biases of posterior medians of σ^2 tend to be slightly positive. However,

DGP Fit	CAR($\rho = -0.95$)		CAR($\rho = -0.25$)		CAR($\rho = 0.00$)	
	CAR	DCAR	CAR	DCAR	CAR	DCAR
PCD(DIC)	100%	0%	51%	49%	34%	66%
P-value	0.64		0.50		0.47	
DGP Fit	CAR($\rho = 0.25$)		CAR($\rho = 0.95$)			
	CAR	DCAR	CAR	DCAR		
PCD(DIC)	55%	45%	100%	0%		
P-value	0.50		0.68			

Table 3: Comparison of DIC between CAR and DCAR models with data sets from CAR process (PCD = Percentage of Correct Decision)

again we found that these empirical biases in all cases are not significant because all calculated p-values are at least 0.5 (results not reported). Finally, with regard to the performance of the posterior medians of β 's, we did not find any significant biases (all p-values being bigger than 0.32). We have not presented the details of these results (for σ^2 and β 's) due to lack of space, but detailed results are available online in the doctoral thesis of the first author (Kyung 2006). Next we compare the results obtained by fitting a DCAR model to the same data sets generated from the five CAR models (as described above).

Performance of DCAR model under mis-specification

In Table 2 we also present the bias of the posterior medians of δ_1 and δ_2 , their MCSE's, p-values (for testing $\delta_1 = \delta_2 = \rho$), and the 95% CP's of the 95% equal-tailed posterior intervals, when a DCAR model is fitted to each of the same 100 data sets generated from each of five CAR models (as described in the previous section). Although DCAR is not the true model that generated the data in these cases, we observe that for all five cases, the biases of posterior medians of δ_1 and δ_2 are marginally positive for Cases 1, 2 and 3, whereas, the biases are slightly negative for Cases 4 and 5. However, these biases are not statistically significant (all p-values being greater than 0.09). This indicates that even when the data arise from a CAR model, the posterior medians of δ 's can well approximate the true ρ value of the CAR model. As expected, the MCSE's of the posterior medians of δ 's are slightly larger than that of the ρ 's, but such loss in efficiency for fitting an incorrect model is not prominent either. Finally, in terms of maintaining the nominal coverage of the posterior intervals, the results from both model fits are comparable. Thus, in summary, when we fit a DCAR model to data sets generated from a CAR model, the posterior estimates obtained from the DCAR model are approximately unbiased and there is no big loss in efficiency.

In addition to comparing the parameter estimates based on fitting both CAR and DCAR models to data generated from a CAR model we have also used DIC (as defined earlier in this section) to compare the overall performance of these models. By a data generating process (DGP) we mean the true model that generates data for our simulation

DGP	CAR($\rho = -0.95$)	CAR($\rho = -0.25$)	CAR($\rho = 0.00$)
$E\left(\frac{\text{Var}(\text{DCAR})}{\text{Var}(\text{CAR})}\right)$	1.009(0.000)	0.999(0.003)	0.998(0.003)
DGP	CAR($\rho = 0.25$)	CAR($\rho = 0.95$)	
$E\left(\frac{\text{Var}(\text{DCAR})}{\text{Var}(\text{CAR})}\right)$	0.999(0.004)	1.009(0.001)	

Table 4: The average ratio of posterior variances for DCAR and CAR models: Average($\text{Var}(\text{DCAR})/\text{Var}(\text{CAR})$) based on Gibbs sampler from data sets of CAR process

study and we use the notation FIT to denote the model that was fitted to a simulated data set. So in this case CAR is the DGP while FIT can be either a CAR or a DCAR model. We measure the performance of the FIT by computing the percentage of correct decisions (PCD) made by DIC in selecting one of the two models. In other words, PCD represents the percentage of the times the DIC value, based on fitting a CAR model, is lower than that of fitting a DCAR model to the same sets of data obtained from a CAR model. We also report the p-values based on performing a two sample test that compares the average values of the DICs (over 100 replications) between CAR and DCAR models when the true data is generated from a CAR model.

From Table 3, we observe that when the DGP is a CAR with $\rho = -0.95$ (negative boundary) and $\rho = 0.95$ (positive boundary), the PCD based on DIC is 100% which means that the DIC correctly identifies a CAR model all the times when data are generated from a CAR model with $\rho = \pm 0.95$. However for other cases, the PCD's are not that strongly in favor of a CAR model (when compared against a DCAR model) even when the data arise from a CAR model. Thus, when the spatial dependence in a CAR model is weak, DIC will not be able to distinguish between the CAR and DCAR models. Again such a phenomenon is expected as DCAR nests CAR when $\delta_1 = \delta_2 = \rho$ and this is further evidenced by looking the p-values, which suggest that we can not reject the null hypotheses that the DIC values are the same for both models.

For the measure of relative efficiency, the average ratio of posterior variances for DCAR and CAR models based on data sets of the CAR processes are reported in Table 4. From Table 4, we observe that there are no differences between the posterior variance for DCAR and for the CAR models based on the Gibbs sampler from data sets of each CAR process. Again such a phenomenon is expected, as DCAR nests CAR when $\delta_1 = \delta_2 = \rho$.

3.2 Results based on data generated from DCAR model

In this section, our DGP (data generating process, as defined in earlier sections) is a DCAR model while the FIT is again either a CAR or a DCAR model. Here again we use the data sets generated from four DCAR models (as defined in Section 2.3) but fit a CAR model in addition to the DCAR models that we fitted earlier (see Section 2.3 for details). In this case it is of interest to find out how the posterior estimates of ρ of a CAR model behave, especially when the data arise from a DCAR model with δ values

DGP: DCAR	$\rho_0 = \frac{\delta_1 + \delta_2}{2}$			
	$\delta_1 = -0.95,$ $\delta_2 = -0.97$	$\delta_1 = -0.30,$ $\delta_2 = 0.95$	$\delta_1 = -0.95,$ $\delta_2 = 0.97$	$\delta_1 = 0.95,$ $\delta_2 = 0.93$
True	-0.96	0.33	0.01	0.94
bias	0.25	0.07	0.03	0.03
MCSE	0.13	0.31	0.38	0.02
P-value	0.09	0.83	0.94	0.13
95% CS	1.00	0.99	0.98	0.85

Table 5: Fitting a CAR model to data generated from the DCAR process

well separated (e.g., for the cases 2 and 3 of Section 2.3).

Performance of CAR model under mis-specification

Based on generating 100 data sets from different DCAR models, we observed that the posterior median of ρ seems to estimate the average of the true values of δ_1 and δ_2 of the DCAR models. Therefore, we define a *pseudo-true* value of ρ as $\rho_0 = \frac{\delta_1 + \delta_2}{2}$ and compare the performance of the posterior median of ρ to this so-called “pseudo-true” value of ρ_0 . In Table 5 we list the empirical bias of posterior median of ρ , the MCSE of these posterior medians, the p-value for testing the null hypothesis $\rho = \rho_0$ and the 95% nominal CP based on the 95% equal-tail posterior intervals of ρ when a CAR model is fitted to four DCAR models (as described in Section 2.3). It is clear from the results reported in this table that the ρ parameter of the CAR model attempts to estimate $(\delta_1 + \delta_2)/2$ of the DCAR model and thus will lead to a misleading conclusion, especially when δ 's are of opposite signs but with large absolute values (e.g., cases 2 and 3 of Section 2.3). In other words, when there are strong spatial dependencies possibly in orthogonal directions, the CAR model would fail to capture such dependencies as opposed to a DCAR model. On the other hand, when the DGP is a CAR model, the DCAR model still provides a very reasonable approximation to that DGP (see the results on Section 3.1). This is one of the main advantages of fitting a DCAR model over a regular CAR model.

In Table 6, we compare the performance of DIC in choosing the correct model (which is a DCAR model in this case) when we fitted both CAR and DCAR models. The numbers reported in this table have similar interpretations as in Table 3. As expected, for the cases 1 and 4 (where $\delta_1 \approx \delta_2$), the DIC more often chooses the CAR model as the best parsimonious model even when the data arise from a DCAR model. However, the p-values (for testing the null hypothesis of no difference in average DIC values indicate that such DIC values are not statistically significantly different. On the other hand when the DCAR model is sharply different from a CAR model (e.g., in cases 2 and 3, where $\delta_1 \delta_2 < 0$), the DIC correctly picks DCAR as the better model more frequently (e.g. 99% of the times in case 3) as compared to a CAR model. Moreover, the p-values suggest that in these two cases the DIC values obtained by fitting CAR and DCAR

DGP	DCAR($\delta_1 = -0.95, \delta_2 = -0.97$)		DCAR($\delta_1 = -0.30, \delta_2 = 0.95$)	
Fit	CAR	DCAR	CAR	DCAR
PCD(DIC)	91%	9%	30%	70%
P-value	0.59		0.03	
DGP	DCAR($\delta_1 = -0.95, \delta_2 = 0.97$)		DCAR($\delta_1 = 0.95, \delta_2 = 0.93$)	
Fit	CAR	DCAR	CAR	DCAR
PCD(DIC)	1%	99%	56%	44%
P-value	0.01		0.73	

Table 6: Comparison of DIC between CAR and DCAR models with data sets from DCAR process (PCD = percentage of correct decisions)

DGP	DCAR($\delta_1 = -0.95, \delta_2 = -0.97$)	DCAR($\delta_1 = -0.30, \delta_2 = 0.95$)
$E\left(\frac{\text{Var}(\text{DCAR})}{\text{Var}(\text{CAR})}\right)$	1.002(0.003)	0.994(0.007)
DGP	DCAR($\delta_1 = -0.95, \delta_2 = 0.97$)	DCAR($\delta_1 = 0.95, \delta_2 = 0.93$)
$E\left(\frac{\text{Var}(\text{DCAR})}{\text{Var}(\text{CAR})}\right)$	0.983(0.010)	1.041(0.247)

Table 7: The average ratio of posterior variances for DCAR and CAR models: $\text{Average}(\text{Var}(\text{DCAR})/\text{Var}(\text{CAR}))$ based on Gibbs sampler from data sets of DCAR process

models are significantly different in favor of the DCAR model when the DGP is indeed a DCAR model.

The average ratio of posterior variances for DCAR and CAR models based on data sets of the DCAR process are reported in Table 7 for the measure of relative efficiency. From Table 7, we observe that there are no differences in the posterior variances for DCAR and for CAR models based on the Gibbs sampler For Cases 1 and 4 (where $\delta_1 \approx \delta_2$). Also, for Case 2, the posterior variances are not different for DCAR and CAR models. However, for the extreme case (Case 3), the posterior variances of the DCAR model are smaller than that of the CAR model. Thus, when there are strong spatial dependencies, possibly in orthogonal directions, the DCAR model would capture such dependencies more precisely than a CAR model.

From our extensive simulation studies we can make the following fairly general conclusions: (i) DCAR models provide a reasonably good fit and approximately unbiased parameter estimates even when the data arise from a CAR model, (ii) CAR models cannot provide an adequate fit for data sets arising from a DCAR model, especially when there are strong spatial dependencies in opposite directions, (iii) DIC performs reasonably well in choosing a parsimonious model when CAR and DCAR models are compared.

4 Data analysis

We illustrate the fitting of DCAR and CAR models using real data sets. For each data set, we consider a linear regression model with iid errors and correlated errors (modeled by CAR and DCAR processes). We obtain the Gibbs sampler of ρ , σ^2 , $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ and $\boldsymbol{\delta} = (\delta_1, \delta_2)$ under different modeling assumptions. We consider the following models:

$$Z_i = \mathbf{X}_i \boldsymbol{\beta} + \epsilon_i \quad i = 1, \dots, n$$

Model 1. $\epsilon_i \sim N(0, \sigma^2)$: iid errors

Model 2. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{I} - \rho \tilde{\mathbf{W}})^{-1} \mathbf{D})$: CAR errors

Model 3. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{I} - \delta_1 \mathbf{W}^{(1)} - \delta_2 \mathbf{W}^{(2)})^{-1} \mathbf{D})$: DCAR errors ,

where $Z_i = f(Y_i)$ and $f(\cdot)$ is a transformation function of the response Y_i .

In addition to using DIC to compare the models with CAR and DCAR error structures, we also computed a cross-validation measure (leave-one-out mean square predictive error (MSPE)). This is defined as follows:

$$\text{MSPE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{-i})^2,$$

where $\hat{y}_{-i} = E(Y_i | y_{-i})$ is the posterior predictive mean of Y_i obtained by fitting a model based on a reduced data set consisting of all (n-1) observations leaving out the i th observation y_i .

4.1 Crime distribution in Columbus, Ohio

We illustrate the performance of fitting CAR and DCAR models to a real data set for estimating the crime distribution in Columbus, Ohio collected during the year of 1980. The original data set can be found in Table 12.1 of [Anselin \(1988, p.189\)](#). Using this interesting data set, [Anselin \(1988\)](#) illustrated the presence of separate levels of spatial dependencies by fitting two separate regression curves with simultaneous autoregressive (SAR) error models for the east and west sides of Columbus city. As a result, the author concluded that when a SAR error model is used, there exists structural instability in terms of the regression models. In this paper, we fit proposed models to this data set. Each of the models is a single regression curve but allow spatial anisotropy in the errors by modeling the errors as a CAR or DCAR model.

The data set consists of the observations collected in 49 contiguous Planning Neighborhoods of Columbus, Ohio. Neighborhoods correspond to census tracts, or aggregates of a small number of census tracts. In this data set, the crime variable represents the total number of residential burglaries and vehicle thefts per thousand households (henceforth denoted by Y_i for the i th neighborhood). As possible predictors for crime variable, we use the income level and housing values for each one of these 49 neighborhoods. The income and housing values are measured in thousands of dollars.

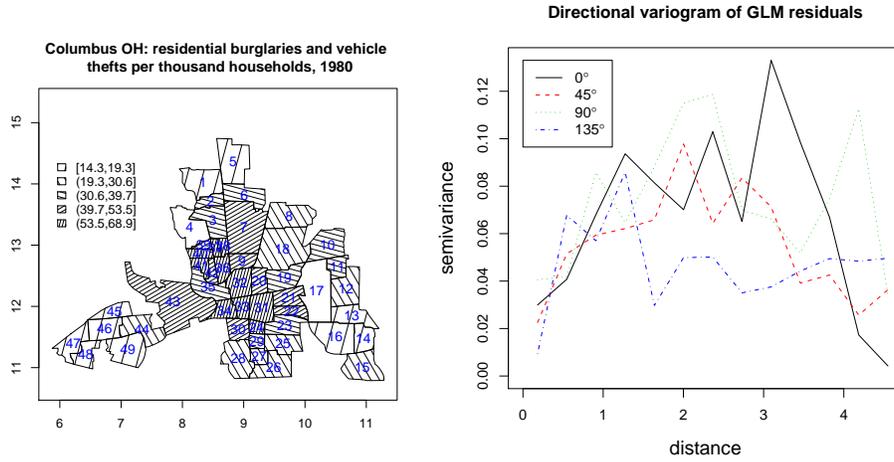


Figure 3: The crime distribution of 49 neighborhoods in Columbus, OH, and the correlogram of the deviance residuals after fitting a Poisson regression model.

As a part of our preliminary exploratory data analysis, in Figure 3, we plot the the crime counts divided into 5 intervals, based on 20% quantiles. During our initial analysis we observed that Y_4 and Y_{17} have extremely small values and hence could possibly be eliminated as outliers or incorrectly recorded values (as these two values were less than 2.5% percentile of the Y_i 's). For rest of the analysis, we use the remaining $n = 47$ neighborhoods for our analysis. From the map in Figure 3 we observe that there seems to be a relatively higher crime frequencies in NW/SE direction than those frequencies in its orthogonal direction, though such differences in crime distribution are not strikingly evident from this plot.

From the estimated directional spatial correlogram in Figure 3, it appears that spatial correlations are not as strong. However, as the distance between neighbors increases, the estimated directional spatial correlation is different from those in different directions. There might be hidden effects of the different directional spatial correlation, thus, we assume a Gaussian DCAR spatial structure.

As our response variable (the crime variable) is a count variable, we assume that $Y_i \sim Poisson(\lambda_i)$ for $i = 1, \dots, n$. Also, let x_{1i} and x_{2i} represent the housing value and the income both in thousand dollars, respectively. Thus $\mathbf{x}_i = (1, x_{1i}, x_{2i})^T$ represents the intercept and predictors for neighborhood i . We consider three over-dispersed Poisson regression models using the latent variables Z_i 's as follows:

$$Y_i \sim Poisson(\lambda_i)$$

$$\log(\lambda_i) = Z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \quad \boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T, \quad i = 1, \dots, n$$

Posterior estimates consisting of the posterior median (denoted by Est. in the table)

Parameter	iid		CAR		DCAR	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
ρ	-	-	0.974	0.021	-	-
δ_1	-	-	-	-	0.962	0.031
δ_2	-	-	-	-	0.960	0.032
σ^2	2.576	-	0.358	0.250	0.230	0.183
β_0	4.568	0.074	4.197	0.264	4.147	0.243
β_1	-0.003	0.001	-0.004	0.003	-0.004	0.003
β_2	-0.064	0.006	-0.056	0.012	-0.055	0.011
DIC	-		335.76		336.79	

Table 8: Posterior estimates based on fitting different models to the crime frequency data.

		MSPE
Model 1	(iid error)	0.084
Model 2	(CAR error)	0.053
Model 3	(DCAR error)	0.050

Table 9: Mean Squared Predicted Error of Leave-one-out method (MSPE)

and the posterior standard deviation (denoted by Std.Err in the table) of the parameters under these models are displayed in Table 8. In this table, we observe that for all models, the posterior estimates of the regression coefficients (β 's) are very similar across all three models. As expected, the negative posterior medians of the β_1 and β_2 indicate that crime frequencies are expected to be lower in neighborhoods with higher income level and housing values. Next we turn our attention to the error part of the three models. First, significantly lower values of the posterior medians of σ^2 under both the CAR and DCAR models indicate that greater variability is explained by the models with spatially correlated errors (i.e., by the CAR and DCAR models in this case) than the corresponding model with independent errors. This is further evidenced from the “deviance residual” (as defined by McCullagh and Nelder 1989) plot in Figure 4 which also suggests that these residuals are not randomly scattered around the horizontal line at the origin. Among the spatially correlated error models, the difference between the DCAR model and the CAR model is negligible. Also, from the scatterplot of the predicted values from DCAR spatial structure versus those from CAR in Figure 4, we observe that there are many points which are far from the straight line. The straight line has slope 1, so if the predicted values are similar to the original data, the points are close to the straight line. However, the predicted values from DCAR and CAR are not different from each other, which is also evident by comparing their corresponding DIC values.

In addition to using DIC to compare the models with CAR and DCAR error structures, we also computed cross-validation measures like leave-one-out mean square predictive error (MSPE). Here $\hat{z}_{-i} = E(Z_i|z_{-i})$ is the ML predictive mean of Z_i obtained

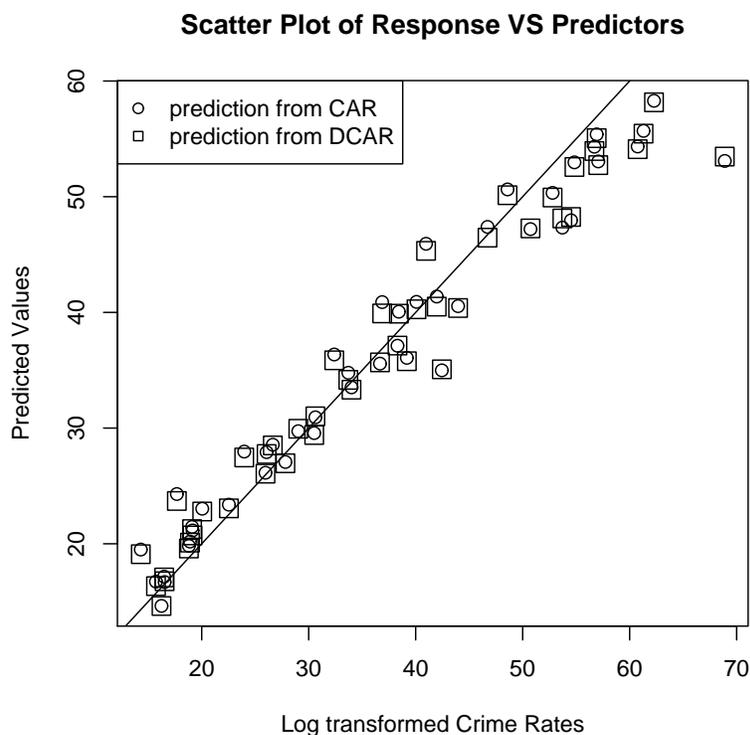


Figure 4: Scatterplot of regional estimated frequencies from DCAR versus those from CAR for the log-transformed crime frequencies. The straight line has slope 1. Thus, if the predicted values are similar to the original data, points are close to the straight line.

by fitting a model based on a reduced data set consisting of all $(n-1)$ observations leaving out the i th observation y_i . In Table 9 we present the MSPEs for three models. Again it is evident that the spatially correlated error models perform much better than the independent error model. Among the two spatial models, DCAR performs slightly better than the CAR model in terms of having lower MSPE. Thus, we conclude that although there is possibly no separate directional spatial correlations, there is a strong spatial correlation on either side of the neighborhoods.

4.2 Elevated Blood Lead Levels in Virginia

We also illustrate the fitting of the CAR and DCAR models using a second data set, estimating the rate, per thousand, of children under the age of 72 months with elevated blood lead levels observed in Virginia in the year 2000. As predictors for the rate of children with elevated blood lead levels, we consider the median housing value in

\$100,000 and the number of children under 17 years of age living in poverty in 2000, per 100,000 children at risk. These observations were collected in 133 counties in Virginia in the year 2000, with coordinates being the centroids of each county. The aggregated data for each county are counts: The number of children under 6 years of age with elevated blood levels in county i and the number of children under 6 years of age tested. In [Schabenberger and Gotway \(2005\)](#), the original data set was used to illustrate the percentage of children under the age of 6 years with elevated blood lead levels by using a Poisson-based generalized linear model (GLM) and a Poisson-based generalized linear mixed model (GLMM) in the analysis of spatial data. [Schabenberger and Gotway \(2005\)](#) illustrated spatial dependence by comparing predictions from a marginal spatial GLM, a conditional spatial GLMM, a marginal spatial GLM using geostatistical variance structure, and a marginal GLM using a CAR variance structure. For the CAR variance structure, they used binary sets of neighbors which share a common border. They mentioned that because of this choice of adjacency weights, the model with the CAR variance smoothes the data much more than the model with the geostatistical variance.

Instead of using a generalized linear model for count data, we consider the Freeman-Tukey (FT) square-root transformation for the Y_i 's. There are zero values in some counties, and the FT square-root transformation shows more stability than the usual square-root transformation ([Freeman and Tukey 1950](#); [Cressie and Chan 1989](#)). With the FT square-root transformed elevated blood lead level rate, we assume a Gaussian distribution with CAR and DCAR spatial structure. For the neighbor structure, we compute distances among the centroids of each geographical group as measured in latitude and longitude. So as not to have any counties reporting zero neighbors, we include counties whose distance is within a 54.69 radius of another county.

For this data set, we denote $Z_i = \sqrt{1000 * Y_i / T_i} + \sqrt{1000 * (Y_i + 1) / T_i}$ for $i = 1, 2, \dots, n$, where Y_i is the number of children under the age of 72 months with elevated blood lead levels observed and T_i is the number of children under the age of 72 months who have been tested in Virginia in the year 2000. Thus, Z_i is a FT square-root transformed elevated blood lead level rate of sub-area S_i . There exists significant correlation between the median housing value in \$100,000 and the number of children under 17 years of age living in poverty in 2000, per 100,000 children at risk. Thus, we only include the centered housing value in \$100,000 (X).

We plot the the FT square-root transformed elevated blood lead level rate that are divided into 5 intervals of the 20% quantiles in [Figure 5](#). In [Figure 5](#), it appears that spatial correlations in the northeast (NE) direction seems strong. However, from the estimated correlogram in [Figure 5](#), we observe that the spatial correlations in four different directions do not seem to be very different from each other. But, there seems to be different amounts of correlation for the 45^0 and 135^0 compared to no directional correlation. Thus, we assume a DCAR process as a hidden spatial structure.

The posterior estimates, with standard deviations under iid error, CAR error, and DCAR error models are displayed in [Table 10](#). In this table, we observe that for all models, the posterior mode of the intercept (β_0) are very similar across all three models. However, the estimate of the regression coefficient of median housing value under iid

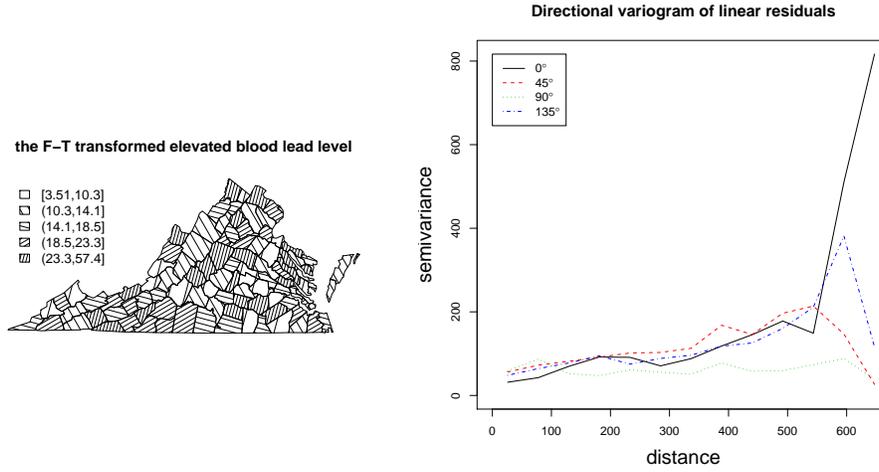


Figure 5: The elevated blood lead levels rate per thousand of children under the age of 72 months observed in Virginia in the year 2000 the correlogram of the deviance residuals after fitting linear model.

Parameter	iid		CAR		DCAR	
	Est.	Std.Err.	Est.	Std.Err.	Est.	Std.Err.
ρ	-	-	0.792	0.120	-	-
δ_1	-	-	-	-	0.450	0.236
δ_2	-	-	-	-	0.896	0.105
σ^2	88.52	11.15	574.1	72.170	564.0	74.2
β_0	17.46	0.822	18.78	2.822	17.42	2.315
β_1	-0.624	2.103	-3.295	3.017	-3.072	2.756
DIC	-		940.532		938.854	

Table 10: Bayesian estimates based on fitting different models to the elevated blood lead level data.

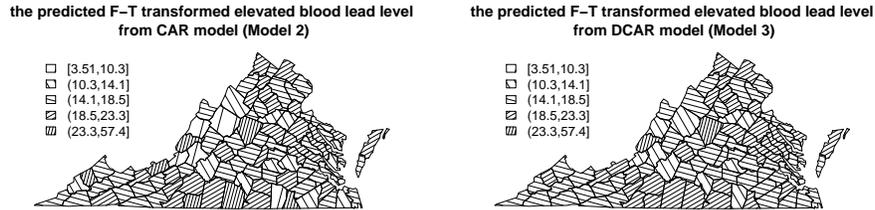


Figure 6: Predicted elevated blood lead levels rate of children in Virginia (Model 2 and Model 3)

errors are different from the posterior estimates of CAR errors and DCAR errors. As expected, the negative posterior medians of the β_1 indicate that the rates per thousand of children under the age of 72 months with elevated blood lead levels are expected to be lower at neighborhoods with higher housing values. The estimate of the error term (σ^2) with independent errors is significantly lower than the corresponding estimates under spatially correlated errors. However, the posterior mode of β_1 (-0.624) is not significant under iid errors, having a large standard error (2.103). As we discussed in Section 3.2, the posterior mode of ρ (0.792) seems to estimate the average of the true values of δ_1 and δ_2 (0.450 and 0.896) of the DCAR models. There exists a positive spatial relationship for elevated blood lead levels among counties in Virginia. However, there exist different amounts of positive spatial correlation among neighbors in the northeast-southwest and the northwest-southeast directions. The spatial correlation among neighbors in northeast-southwest direction ($\hat{\delta}_2 = 0.896$) is stronger than in northwest-southeast direction ($\hat{\delta}_1 = 0.450$). Among the spatially correlated error models, DCAR explains slightly more variability than the CAR, though the difference between these models is negligible, which is also evident by comparing their corresponding DIC values. This is further evidenced from the residual plots in Figure 6 which also suggest that the residuals based on the DCAR error model appears not to have a trend over the study region. Also in Figure 7, we observe that most of predicted values from the DCAR spatial structure are bigger than those from CAR. This means that for the FT-transformed elevated blood lead levels, the DCAR model captures more variability than the CAR model in stabilizing estimates within the regions using the estimated spatial correlation.

To compare the models with CAR and DCAR error structures, we also computed leave-one-out mean square predictive error (MSPE). In Table 11 we present the MSPEs for three models. Again it is evident that the spatially correlated error models perform

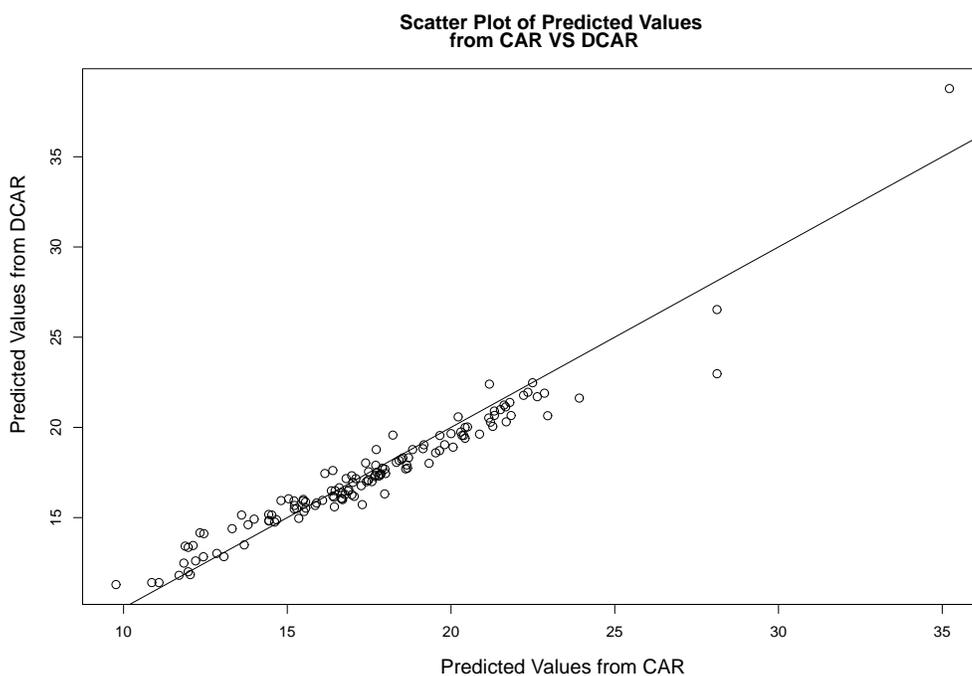


Figure 7: Scatterplot of regional estimated rates from DCAR versus those from CAR for the FT-transformed original elevated blood lead level rates. The straight has slope 1. Thus, if the predicted values from DCAR are similar to the predicted values of CAR, points are close to the straight line.

	MSPE
Model 1	88.155
Model 2	66.822
Model 3	64.269

Table 11: Mean Squared Predicted Error (Elevated blood lead level data)

much better than the independent error model. Among the two spatial models, DCAR (64.269) performs slightly better than the CAR model (66.822) in terms of having lower MSPE, but the difference is negligible. Thus, we conclude that there are strong spatial correlations with some evidence of differing strengths of correlation in different directions.

5 Extensions and future work

DCAR models capture the directional spatial dependence in addition to distance specific correlation, thus they are an extension of regular CAR models, which can often fail to capture strong but directionally orthogonal spatial correlations. The DCAR model is also found to be nearly as efficient as the CAR model even when data are generated from the CAR model. However, CAR models usually fail to capture the directional effects when data are generated from DCAR or other anisotropic models, particularly when the anisotropy is pronounced.

Our model proposed in (6) can be extended to M ($M \geq 2$) directions, and can be expressed as

$$\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I} - \sum_{k=1}^M \delta_k \tilde{\mathbf{W}}^{(k)})^{-1}\mathbf{D}),$$

where $\tilde{\mathbf{W}}^{(k)}$ denotes the matrices of weights specific to k th directional effect. In this paper we used only $M = 2$ sub-neighborhoods for a simpler illustration. However, we note that if we keep increasing the number of sub-neighborhoods, the number of parameters increases, and the amount of observations available within a sub-neighborhood decreases. Thus, we need to restrict the number of sub-neighborhoods by introducing a penalty term (or prior) and use some form of information criterion to choose the number of sub-neighborhoods. This is an important but open issue within our DCAR framework and we leave its further exploration as a part of our future research.

References

- Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers. 694
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal

- of the Royal Statistical Society, Series B. 36, 192-236 (with discussion). 676, 677, 678, 680
- Besag, J., 1975. Spatial analysis of non-lattice data. *The Statistician* 24, 179-195. 677
- Besag, J and Kooperberg, C., 1995. On Conditional and Intrinsic Autoregression. *Biometrika* 82, 733-746. 676, 680, 685
- Breslow, N. E. and Clayton, D. G., 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 88, 9-25. 677
- 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. 680
- Clayton, D. and Kaldor, J., 1987, Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. *Biometrics*. 43, 671-681. 677
- Cliff, A. D. and Ord, J. K., 1981. *Spatial Processes: Models & Applications*. Pion Limited. 677
- Cressie, N., 1993. *Statistics for Spatial Data*. John Wiley & Sons, Inc. 675, 676
- Cressie, N. and Chan, N. H., 1989. Spatial modeling of regional variables. *Journal of the American Statistical Association*. 84, 393-401. 677, 698
- Freeman, M. F. and Tukey, J. W., 1950. Transformations related to the angular and the square root. *Annals of Mathematical Statistics*. 21, 607-611. 698
- Fuentes, M., 2002. Spectral methods for nonstationary spatial processes. *Biometrika*. 89, 197-210. 676
- Fuentes, M., 2005. A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*. 96, 30-54. 676
- Fuentes, M and Smith, R., 2001. A new class of nonstationary spatial models. Technical report, North Carolina State University, Department of Statistics. 676
- Griffith, D. A. and Csillag, F., 1993. Exploring Relationships Between Semi-Variogram and Spatial Autoregressive Models. *Papers in Regional Science*. 72, 283-295. 676
- Higdon, D., 1998. A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Journal of Environmental and Ecological Statistics*. 5, 173-190. 676
- Higdon, D., Swall, J. and Kern, J., 1999. Non-stationary spatial modeling. In *Bayesian Statistics* 6, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford: Oxford University Press, 761-768. 676
- Hrafinkelsson, B. and Cressie, N., 2003. Hierarchical modeling of count data with application to nuclear fall-out. *Journal of Environmental and Ecological Statistics*. 10, 179-200. 676

- Hughes-Oliver, J. M., Heo, T. Y. and Ghosh, S. K., 2009. An Autoregressive Point Source Model for Spatial Processes. *Environmetrics*. 20, 575-594. 676
- Hyndman, R. J., 1996. Computing and Graphing Highest Density Regions. *The American Statistician*. 50, 120-126. 686
- Journel, A. G. and Huijbregts, C. J., 1978. *Mining geostatistics*. London:Academic. 676
- Kyung, M., 2006. *Generalized Conditionally Autoregressive Models*. Ph. D. Thesis, North Carolina State University, Department of Statistics. 686, 690
- Mardia, K. V. and Marshall, R. J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*. 71, 135-146. 684
- McCullagh, P. and Nelder, J. A., 1989. *Generalized Linear Models*. Chapman and Hall, London. 696
- Miller, H. J., 2004. Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*. 94, 284-295. 675
- Ord, K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*. 70, 120-126. 677
- Ortega, J. M., 1987. *Matrix Theory*. New York:Plenum Press. 681
- Paciorek, C. J. and Schervish, M. J., 2006. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*. 17, 483-506. 676
- Reich, B. J., Hodges, J. S. and Carlin, B. P., 2007. Spatial analyses of periodontal data using conditionally autoregressive priors having two classes of neighbor relations. *Journal of the American Statistical Association*. 102, 44-55. 677
- Rue, H. and Tjelmeland, H., 2002. Fitting Gaussian Markov Random Fields to Gaussian Fields. *Scandinavian Journal of Statistics*. 29, 31-49. 676
- Schabenberger, O. and Gotway, C. A., 2005. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC. 676, 680, 698
- Song, H. R., Fuentes, M. and Ghosh, S., 2008. A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate Analysis*. 99, 1681-1697. 676
- Spiegelhalter, D. J., Best, N. J., Carlin, B. P. and van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*. 64, 583-639 (with discussion). 688
- Sun, D., Tsutakawa, R. K. and Speckman, P. L., 1999. Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika*. 86, 341-350. 683
- van der Linde, A., Witzko, K.-H. and Jöckel, K.-H., 1995. Spatio-temporal analysis of mortality using splines. *Biometrics*. 4, 1352-1360. 676, 680

Wahba, G., 1977. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM Journal on Numerical Analysis*. 14, 651-667. [680](#)

White, G. and Ghosh, S. K., 2008. A Stochastic neighborhood Conditional Autoregressive Model for Spatial Data. *Computational Statistics and Data Analysis*. 53, 3033-3046. [677](#)

Acknowledgments

The authors are grateful to a referee, the associate editor and the editor for their careful reading of the paper and their constructive comments. We also thank Professor George Casella, Department of Statistics, University of Florida for comments and suggestions that led to a much improved version of the paper.

