University of Wollongong

Research Online

Faculty of Informatics - Papers (Archive)

Faculty of Engineering and Information Sciences

2008

Some diagnostics for Markov random fields

Noel A. Cressie
University of Wollongong, ncressie@uow.edu.au

Prasenjit Kapat
Ohio State University

Follow this and additional works at: https://ro.uow.edu.au/infopapers

Part of the Physical Sciences and Mathematics Commons

Recommended Citation

Some diagnostics for Markov random fields

Abstract

The development of diagnostics to check the fit of a proposed Markov random field (MRP) to data is a very important problem in spatial statistics. In this article, the consequences of fitting a given MRF to spatial data are visualized using diagnostic plots. The Gaussian MRF known as the conditional autoregressive model is featured. Various types of departures of the data from the fitted MRF model are calculated, allowing locally influential observations to be highlighted using the MRF-Neighborhoods plot. Through a global summary statistic and the Model-Comparison plot, we compare MRF models that differ both in terms of the neighborhood structure and the parameterization of spatial dependence. © 2008 American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America.

Keywords

diagnostics, markov, random, fields, ERA2015

Disciplines

Physical Sciences and Mathematics

Publication Details

Journal of Computational and Graphical Statistics

Some Diagnostics for Markov Random Fields

Noel Cressie and Prasenjit Kapat

Noel Cressie is Professor of Statistics and Distinguished Professor
of Mathematical and Physical Sciences, The Ohio State University,
Columbus, OH 43210-1247 . Prasenjit Kapat is a Ph.D. student,
Department of Statistics, The Ohio State University, Columbus, OH
43210-1247 .
Version of record first published: 01 Jan 2012.

PLEASE SCROLL DOWN FOR ARTICLE

Some Diagnostics for Markov Random Fields

Noel CRESSIE and Prasenjit KAPAT

The development of diagnostics to check the fit of a proposed Markov random field (MRF) to data is a very important problem in spatial statistics. In this article, the consequences of fitting a given MRF to spatial data are visualized using diagnostic plots. The Gaussian MRF known as the conditional autoregressive model is featured. Various types of departures of the data from the fitted MRF model are calculated, allowing locally influential observations to be highlighted using the MRF-Neighborhoods plot. Through a global summary statistic and the Model-Comparison plot, we compare MRF models that differ both in terms of the neighborhood structure and the parameterization of spatial dependence.

Key Words: CAR model; Irregular lattice; Maximum likelihood estimation; Model-Comparison plot; MRF-Neighborhoods plot; Neighborhood structure; Regular lattice.

1. INTRODUCTION

Diagnostics are helpful in choosing and comparing models. They are also very useful in detecting outliers. One open challenge in the field of spatial statistics is to be able to evaluate a proposed Markov random field (MRF) with informative diagnostics.

When analyzing spatial lattice data, one often starts with the assumption that the data come from a given class of MRFs, defined according to a particular neighborhood structure and a finite number of parameters (e.g., Besag 1974; Cressie 1993, chap. 6; Banerjee, Carlin, and Gelfand 2004, sec. 3.3; Waller and Gotway 2004, sec. 9.5.3). Most MRF models are based on nearest neighbors (i.e., regions sharing a common boundary) or second nearest neighbors (i.e., nearest neighbors and *their* nearest neighbors); see Section 3.4 and Cressie (1993, p. 285). There have also been attempts to model the spatial dependencies using distance-based neighborhood models where parametric functions of distances between regions are used. Cressie and Chan (1989), Conlon and Waller (2000), and Hrafnkelsson and Cressie (2003) modeled regression coefficients in the conditional-mean structure as a decreasing function of distance between regions.

Noel Cressie is Professor of Statistics and Distinguished Professor of Mathematical and Physical Sciences, The Ohio State University, Columbus, OH 43210-1247 (E-mail: *ncressie@stat.osu.edu*). Prasenjit Kapat is a Ph.D. student, Department of Statistics, The Ohio State University, Columbus, OH 43210-1247 (E-mail: *pkapat@stat.osu.edu*).

Much effort has been put into constructing, checking, and validating linear models in a spatial setting (e.g., Cressie 1993; Banerjee et al. 2004), but when the spatial error term is a MRF, the neighborhood structure is typically assumed known. In this article, we develop a diagnostic tool for validating and comparing neighborhood structures based on the idea of local diagnostics, which can be found *inter alia* in Getis and Ord (1992), Anselin (1995), and Cressie and Collins (2001). We feature classes of MRFs known as conditional autoregressive (CAR) models, which are MRFs that have well defined joint Gaussian distributions. In specifying the conditional distribution of these CAR models, we use a multiplicative spatial-dependence parameter in the conditional expectation, defined so that all eigenvalues of the precision matrix are positive (i.e. the "propriety parameter" from Carlin and Banerjee 2003). That is, all CAR models considered in this article have a nonsingular variance matrix.

In Section 2, we give a general result that serves as the basis of our new diagnostics for MRFs. In Section 3, we describe the CAR models, which are based on conditional (and hence joint) Gaussian assumptions. We also derive properties of the CAR models that are most pertinent to our MRF diagnostics; details of the derivations are given in the Appendix. In Sections 4 and 5, we apply the diagnostics to two datasets, an archeological dataset of phosphate concentrations on a regular lattice in $\mathbb{R}^2$ (Buck, Cavanagh, and Litton 1988) and a dataset of doctors' prescription amounts on an irregular lattice in $\mathbb{R}^2$ (Cressie, Perrin, and Thomas-Agnan 2005, 2006). Finally, in Section 6, we present discussion and conclusions.

2. GENERAL RESULT

We first establish some notation. The conditional density (or probability mass function) of random variable (r.v.) X given r.v. Y, assuming it exists, is denoted by $p[x|y]$. Let $S \equiv \{1, \ldots, n\}$ be the site indices whose spatial relationships are defined by a set of neighbors $\{N_i : i = 1, \ldots, n\}$, where $N_i \subset S$ is the neighborhood of site i and, by convention, $i \notin N_i$. The data defined on a subset $A \subset S$ are denoted by $\mathbf{Z}(A) \equiv (Z(a) : a \in A)'$. Let $Z(i) \equiv Z(\{i\})$ denote the datum at site i. For a set A, let $\overline{A}$ denote the complement of A. Define $M_i \equiv \overline{N_i \cup \{i\}}$, so that $M_i \cup N_i \cup \{i\} = S$. In what follows, $Z(i)$, $\mathbf{Z}(N_i)$, and $\mathbf{Z}(M_i)$ are featured, along with realized values $z(i)$, $\mathbf{z}(N_i)$, and $\mathbf{z}(M_i)$, respectively.

A MRF on S is defined by its neighborhoods $\{N_i : i = 1, \ldots, n\}$, as follows. Assuming $p[\mathbf{z}(S)]$ exists, then the conditional probabilities of a MRF satisfy

$$p\left[z(i)|\mathbf{z}(S\backslash\{i\})\right] = p\left[z(i)|\mathbf{z}(N_i)\right]; \qquad i = 1, \ldots, n. \tag{2.1}$$

Consequently,

$$p\left[z(i), \mathbf{z}(M_i)|\mathbf{z}(N_i)\right] = p\left[z(i)|\mathbf{z}(M_i), \mathbf{z}(N_i)\right] \ p\left[\mathbf{z}(M_i)|\mathbf{z}(N_i)\right]$$
$$= p\left[z(i)|\mathbf{z}(N_i)\right] \ p\left[\mathbf{z}(M_i)|\mathbf{z}(N_i)\right].$$

That is, given $\mathbf{Z}(N_i)$, $Z(i)$ and $\mathbf{Z}(M_i)$ are conditionally independent. This allows us to prove the following proposition, which we shall see provides a tool to develop local diag-

nostics relevant to each site.

Proposition 1. *For any function g that satisfies* $E(|g(Z(i))|) < \infty$, *for all* $i \in S$,

$$\int g(z(i)) \, p\,[z(i)|\mathbf{z}(M_i)]\,dz(i) = \int m(\mathbf{z}(N_i)) \, p[\mathbf{z}(N_i)|\mathbf{z}(M_i)]d\mathbf{z}(N_i),$$

where

$$m(\mathbf{z}(N_i)) \equiv E\big(g(Z(i))|\mathbf{z}(N_i)\big) = \int g(z(i)) \, p\,[z(i)|\mathbf{z}(N_i)]\,dz(i). \qquad (2.2)$$

Proof: Because $E(|g(Z(i))|) < \infty$, we can use Fubini's theorem, and the conditional independence allows us to write

$$\int g(z(i)) \, p\,[z(i)|\mathbf{z}(M_i)]\,dz(i)$$

$$= \int g(z(i)) \left\{ \int p[z(i)|\mathbf{z}(N_i)] \, p[\mathbf{z}(N_i)|\mathbf{z}(M_i)]d\mathbf{z}(N_i) \right\} dz(i)$$

$$= \int \left\{ \int g(z(i)) \, p[z(i)|\mathbf{z}(N_i)]dz(i) \right\} p[\mathbf{z}(N_i)|\mathbf{z}(M_i)]d\mathbf{z}(N_i)$$

$$= \int m(\mathbf{z}(N_i)) \, p[\mathbf{z}(N_i)|\mathbf{z}(M_i)]d\mathbf{z}(N_i). \qquad \qquad \square$$

Thus, we see that for a suitable function g, and assuming the MRF to be correctly specified, $g(Z(i))$ and $m(\mathbf{Z}(N_i))$ have the same conditional (on $\mathbf{Z}(M_i)$) expectation. The MRF diagnostics that we propose compare these two quantities empirically, site-by-site, looking for departures of the relationship established in the proposition. A departure indicates that the neighborhood-based conditional independence assumed in the proof of the proposition does not hold. More generally, it is an indicator of lack of fit of a candidate MRF; we see in Section 3 how this can be diagnosed locally and globally. In our research, we have considered several "test functions" g, including $g(x) = x$, $g(x) = x^2$, and $g(x) = e^x$, the latter exhibiting greatest diagnostic sensitivity. We apply our diagnostics to the conditional autoregressive (CAR) models, described in the next section.

3. CONDITIONAL AUTOREGRESSIVE (CAR) MODELS

When the conditional distributions in (2.1) are assumed to be Gaussian distributions, then it can be shown that, under symmetry and positive-definiteness conditions, the joint distribution is Gaussian (Besag 1974). The resulting MRFs are known as CAR models. Specifically, we assume that

$$Z(i)|\mathbf{z}(N_i) \sim \text{Gau}\Big(\mu_i + \sum_{j=1}^{n} c_{ij}(z(j) - \mu_j), \, \tau_i^2\Big); \quad i = 1, \ldots, n, \qquad (3.1)$$

where the $\{c_{ij}\}$ satisfy $c_{ij} = 0$, $j \notin N_i$, $c_{ii} = 0$, and $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$; $i, j = 1, \ldots, n$. Also, $\mu_i \in \mathbb{R}$ and $\tau_i^2 > 0$, for all $i \in S$. The choice of neighborhoods $\{N_i\}$ and spatial-

dependence parameters $\{c_{ij}\}$ creates various classes of CAR models.

Using standard notation, define $M \equiv \mathrm{diag}\,(\tau_1^2, \ldots, \tau_n^2)_{n \times n}$, $C \equiv (c_{ij})_{n \times n}$, and $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_n)'$. Further, assume that $\boldsymbol{\mu} = X\boldsymbol{\beta}$, for covariates $X_{n \times p}$ and regression parameters $\boldsymbol{\beta}_{p \times 1}$. Although the diagnostics given here hold for any class of MRF models, in this article we develop them for CAR models and, in particular, for CAR models where $M = \Phi\tau^2$ and $C = \gamma H$, for known $\Phi \equiv \mathrm{diag}\,(\phi_1, \ldots, \phi_n)_{n \times n}$ and known $H \equiv (h_{ij})_{n \times n}$. Clearly, γ is the spatial-dependence parameter.

Specifying the conditional distributions (3.1) does not necessarily imply the existence of a joint distribution. However, provided $\Phi^{-1}(I - \gamma H)$ is symmetric and positive-definite, there is a valid joint distribution given by

$$\mathbf{Z} \sim \mathrm{Gau}\left(X\boldsymbol{\beta}, \ (I - \gamma H)^{-1}\Phi\tau^2\right), \tag{3.2}$$

which depends on unknown parameters $\boldsymbol{\beta}$, τ^2, and γ; see Besag (1974) and Cressie (1993, sec. 6.6).

3.1 THREE CLASSES OF CAR MODELS

From (3.2), the variance matrix of a CAR model is given by $(I - \gamma H)^{-1}\Phi\tau^2$. In this article, we consider three specific types of CAR models. In each one, we make use of the adjacency matrix $A = (a_{ij})_{n \times n}$, where $a_{ij} = a_{ji} = 1$ if and only if i and j are neighbors. Note that because $i \notin N_i$, $a_{ii} = 0$ for all $i = 1, \ldots, n$.

(i) *Homogeneous CAR (HCAR) model:* This is the simplest model we shall consider with $\Phi = I_{n \times n}$ and $H = A$.

(ii) *Weighted (heterogeneous) CAR (WCAR) model:* Following Besag, York, and Mollié (1991), we let

$$\Phi = \mathrm{diag}\left(|N_1|^{-1}, \ldots, |N_n|^{-1}\right),$$

where $|N_i| \equiv \sum_{k=1}^{n} a_{ik}$ is the number of neighbors of site i and $H = (h_{ij})_{n \times n}$ is defined by $h_{ij} \equiv a_{ij}/|N_i|$; $i, j = 1, \ldots, n$.

(iii) *Autocorrelation (heterogeneous) CAR (ACAR) model:* This is an *adaptation* of the spatial-rates model (Cressie and Chan 1989; Cressie et al. 2005) that has the same autocorrelation for any two neighboring sites. In this case, Φ is the same as in (ii), the WCAR model, and $H = (h_{ij})$ is defined by $h_{ij} \equiv a_{ij}|N_j|^{1/2}/|N_i|^{1/2}$. Then, it is straightforward to show that $\mathrm{corr}\,(Z(i), Z(j)|\mathbf{z}(S\backslash\{i, j\})) = \gamma$, and hence the parameter space of γ is a subset of the open interval $(-1, 1)$.

Notice that for $\gamma = 0$, $\mathbf{Z}$ is made up of independent components, but for the WCAR and ACAR models they are not identically distributed. In particular, their variances depend on the number of neighbors.

3.2 THE GENERAL RESULT APPLIED TO CAR MODELS

Recall from (3.2) that, provided $\Phi^{-1}(I - \gamma H)$ is symmetric and positive-definite, $\mathbf{Z} \sim \mathrm{Gau}\left(X\boldsymbol{\beta}, (I - \gamma H)^{-1}\Phi\tau^2\right)$. Our diagnostic procedure substitutes maximum likeli-

hood estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\tau}^2$ for nuisance parameters $\boldsymbol{\beta}$ and τ^2, and γ takes several possible values including the maximum likelihood estimator $\hat{\gamma}$. In practice, the maximum likelihood estimator of γ is obtained via the profile-likelihood method (Cressie 1993, pp. 465–467).

In what follows, we develop diagnostics for CAR model types (i), (ii), and (iii) and test function $g(x) = e^x$, based on the proposition in Section 2. From (2.2), we have

$$m(\mathbf{Z}(N_i)) = E\left(e^{Z(i)}|\mathbf{Z}(N_i)\right) = \exp\left\{\mu_i + \gamma \sum_{j \in N_i} h_{ij}(Z(j) - \mu_j) + \frac{1}{2}\phi_i\tau^2\right\}, \quad (3.3)$$

where the latter equality is established in Result 1 of the Appendix. Define

$$W_i^* \equiv \frac{e^{Z(i)}}{m(\mathbf{Z}(N_i))}, \quad (3.4)$$

for $m(\mathbf{Z}(N_i))$ given by (3.3); $i = 1, \ldots, n$. From Result 2 of the Appendix, $E(W_i^*|\mathbf{Z}(M_i)) = 1$, and hence $E(W_i^*) = 1$; consequently, we base our diagnostics on deviations of $\{W_i^*\}$ from a target value of 1. Define $\mathbf{W}^* \equiv (W_1^*, \ldots, W_n^*)'$ and $\Sigma^* \equiv \text{var}(\mathbf{W}^*)$. From Result 3 of the Appendix,

$$\Sigma^* = e^B - J,$$

where $B \equiv (I - \gamma H)\Phi\tau^2$, e^B is defined elementwise as $(e^B)_{ij} \equiv e^{B_{ij}}$, and J is the $n \times n$ matrix whose entries are all unity. Our proposed diagnostics actually use a *standardized* set of residuals:

$$\mathbf{W} \equiv (\Sigma^*)^{-1/2}(\mathbf{W}^* - \mathbf{1}), \quad (3.5)$$

and in the following paragraphs we develop diagnostics based on this $\mathbf{W} \equiv (W_1, \ldots, W_n)'$.

Using the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\tau}^2$ plugged into the distribution of $\mathbf{Z}$, we take a parametric-bootstrap approach (e.g., Casella and Berger 2002, p. 480) and simulate $m(= 1000)$ random vectors $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)}$ from the distribution, $\text{Gau}(X\hat{\boldsymbol{\beta}}, (I - \gamma_0 H)^{-1}\Phi\hat{\tau}^2)$, where γ_0 is some plug-in value, such as $\hat{\gamma}$. For each choice of H, we have the following restriction on the possible values of γ (see Cressie 1993, p. 559):

$$\gamma_{\min} \equiv (\min\{\lambda_i\})^{-1} < \gamma < (\max\{\lambda_i\})^{-1} \equiv \gamma_{\max}, \quad (3.6)$$

where $\{\lambda_i\}$ are the eigenvalues of $\Phi^{-1/2}H\Phi^{1/2}$. It should be noted here that $\Phi^{-1/2}H\Phi^{1/2}$ is symmetric, whereas H need not be so, and hence $\{\lambda_i\}$ are real. An easy calculation shows that $\Phi^{-1/2}H\Phi^{1/2}$ is the same (and hence $\gamma_{\min}$ and $\gamma_{\max}$ are the same) for the HCAR and ACAR models.

Four different spatial-dependence-parameter values are compared using our diagnostic approach: $\gamma_0 = \gamma_{\min}+\epsilon$, $\gamma_0 = 0$, $\gamma_0 = \hat{\gamma}$, and $\gamma_0 = \gamma_{\max}-\epsilon$, where ϵ was arbitrarily chosen as 10^{-4} to avoid a singular variance matrix in (3.2). From the discussion in Besag and Kooperberg (1995), we expect to see that $\gamma_0 = \hat{\gamma}$ and $\gamma_0 = \gamma_{\max} - \epsilon$ result in diagnostics that behave similarly.

3.3 DIAGNOSTICS FOR CAR MODELS

The first diagnostic plot, which we call the *MRF-Neighborhoods plot*, is constructed as follows. The standardized residuals, $\{W_1, \ldots, W_n\}$, are depicted on a spatial lattice using *bubbles*. (The area of each bubble was chosen proportional to the absolute value of the residual at that site.) Filled and empty bubbles are used to distinguish positive and negative values of the residuals, respectively. The sites where observations are missing are denoted with an "×."

Recall that we perform a parametric boostrap. For each simulated $\mathbf{Z}^{(j)}{}_{n \times 1}$, we obtain $\mathbf{W}^{(j)}{}_{n \times 1}$ from (3.5); $j = 1, \ldots, m$. We also calculate the lower 2.5 and upper 97.5 percentiles based on $\{W_i^{(1)}, \ldots, W_i^{(m)}\}$, for each site i. If W_i, the residual at site i, is outside the interval defined by these percentiles, then it is marked as an "outlier" by a "dot" inside its bubble on the plot. This dot is white in a filled bubble if W_i is above the 97.5 percentile, and it is black in an empty bubble if W_i is below the 2.5 percentile. Figure 1 (regular lattice) and Figure 8 (irregular lattice) are examples of this diagnostic plot.

While the MRF-Neighborhoods plot contains a lot of information about individual sites and local departures from the model, it is useful to have a single diagnostic statistic that is an overall measure of how well a model fits. To this end, we propose the mean squared error (MSE) based on $\{W_i\}$ for each model, defined as

$$\text{MSE}_W = \frac{1}{n} \sum_{i=1}^{n} W_i^2. \tag{3.7}$$

We could compare MSE_W-values to decide which model we prefer. In what follows, we develop a diagnostic plot for choosing between a pair of models, which we call the *Model-Comparison plot*. The difference between the mean squared errors of two models, A and B, is given by,

$$
\begin{aligned}
\text{MSE}_W^{(A)} - \text{MSE}_W^{(B)} &= \frac{1}{n} \sum_{i=1}^{n} (W_i^{(A)2} - W_i^{(B)2}) \\
&= \frac{1}{n} \sum_{i=1}^{n} (|W_i^{(A)}| + |W_i^{(B)}|)(|W_i^{(A)}| - |W_i^{(B)}|).
\end{aligned}
$$

Now consider a rectangle with vertical side $|W_i^{(A)}| - |W_i^{(B)}|$ and horizontal side $|W_i^{(A)}| + |W_i^{(B)}|$, contributing either a positive area or a negative area. If we sort these rectangles according to

$$W_{i_1}^{(A)2} - W_{i_1}^{(B)2} < \cdots < W_{i_n}^{(A)2} - W_{i_n}^{(B)2},$$

and plot

$$|W_{i_j}^{(A)}| - |W_{i_j}^{(B)}| \text{ against } \frac{1}{n} \sum_{k=1}^{j} (|W_{i_k}^{(A)}| + |W_{i_k}^{(B)}|), \tag{3.8}$$

for $j = 1, \ldots, n$, we obtain the Model-Comparison plot. A rectangle formed above the

horizontal axis adds positive area in favor of Model B, while a rectangle below the horizontal axis adds negative area in favor of Model A. Finally, if there is more positive total area than negative total area, $\text{MSE}_W^{(B)} < \text{MSE}_W^{(A)}$, which means that Model B is preferred to Model A. Figure 3 is an example of this diagnostic plot.

3.4 MODELS APPLIED TO A REGULAR LATTICE IN $\mathbb{R}^2$

In Sections 3.1 and 3.2, we have developed diagnostics for a CAR model on a spatial lattice, regular or irregular, in $\mathbb{R}^d$, with a specified neighborhood structure $\{N_i\}$. We can also compare CAR models, and we are particularly interested in comparing those with different neighborhood structures.

Consider the nearest-neighbor (NN) and the second-nearest-neighbor (2NN) structures over a regular lattice of $n = k \cdot l$ sites, as described in the following. We assume that the regular lattice is defined on $\{(x, y) : x = 1, \ldots, k, \ y = 1, \ldots, l\}$ in $\mathbb{R}^2$. These two neighborhood structures reflect spatial dependence at shorter and longer spatial scales, respectively.

Ignoring for the moment the lattice-boundary constraints, the *NN structure* is defined by the neighborhood,

$$N^{(1)}(x, y) \equiv \{(x - 1, y), (x, y - 1), (x + 1, y), (x, y + 1)\}, \qquad (3.9)$$

and the *2NN structure* is defined by the neighborhood,

$$\begin{aligned}
N^{(2)}(x, y) \equiv\ & N^{(1)}(x, y) \\
& \cup \{(x - 1, y - 1), (x + 1, y - 1), (x + 1, y + 1), (x - 1, y + 1)\} \\
& \cup \{(x - 2, y), (x, y - 2), (x + 2, y), (x, y + 2)\}, \qquad (3.10)
\end{aligned}$$

of (x, y), for $x = 1, \ldots, k$ and $y = 1, \ldots, l$. Clearly, $N^{(1)}(x, y) \subset N^{(2)}(x, y)$. Boundary constraints are introduced by removing any location (x^*, y^*) from a neighborhood if it falls outside the lattice $\{(x, y) : x = 1, \ldots, k, \ y = 1, \ldots, l\}$.

For brevity, we denote CAR models applied to a regular lattice using the ordered triple: (CAR model type, neighborhood structure, γ_0). For example, (ACAR, 2NN, $\hat{\gamma}$) represents the ACAR model with 2NN structure and $\gamma_0 = \hat{\gamma}$.

3.5 MODELS APPLIED TO AN IRREGULAR LATTICE IN $\mathbb{R}^2$

For an irregular lattice, we consider distance-based neighborhoods. Recall that $S = \{1, \ldots, n\}$ is the set of all site indices. Suppose ξ_{ij} is the distance between sites i and j, based on well defined geographical co-ordinates for each of the n sites. For a given $d > 0$, we construct neighborhoods as follows:

$$N_i^{(d)} = \{j \in S : 0 < \xi_{ij} < d\}; \qquad i = 1, \ldots, n. \qquad (3.11)$$

Then, for any $d_1 < d_2$, $N_i^{(d_1)} \subset N_i^{(d_2)}$, for all $i = 1, \ldots, n$, thereby creating nested

neighborhoods analogous to the NN and 2NN structures of the regular lattices given in Section 3.4. A benefit of our analysis on these types of neighborhood structures is that we can do a comparison of the model fit for different values of d. Just as for the NN and 2NN structures, we now have a way to determine a preferred spatial scale of association from the data, rather than fixing it a priori.

For brevity, we denote CAR models applied to an irregular lattice using the ordered triple: (CAR model type, neighborhood structure, γ_0). For example, (WCAR, $N^{(30)}$, 0) represents the WCAR model with neighborhood structure $N^{(30)} \equiv \{N_i^{(30)}\}$ and $\gamma_0 = 0$.

4. ARCHEOLOGICAL DATASET OF PHOSPHATE CONCENTRATIONS

These spatial data are on a regular lattice in $\mathbb{R}^2$.

4.1 DESCRIPTION OF THE DATA

Between 1983 and 1989, a joint team from the British School in Athens, Greece, and the Universities of Amsterdam and Nottingham carried out an intensive survey of a 70 sq km area of Laconia across the Evrotas (ancient Eurotas) river, east from the ancient site of Sparta, Greece. These data consist of raw phosphate concentration readings (in mg P/100 g of soil) taken 10 m apart, from the site LS 165 of the Laconia Survey; the observations are distributed over a regular 16×16 grid (Buck et al. 1988). We denote these raw data by $D(x, y)$, where (x, y) represents a location of a datum on the grid; $x, y = 1, \ldots, 16$. Based on exploratory data analysis, we symmetrized the histogram using the fourth-root transformation; that is, $Z(x, y) \equiv [D(x, y)]^{0.25}$. The raw data are given in Table 1; we see that data are unavailable at the following sites:

$$(5, 11), (5, 12), (5, 15), (6, 11), (6, 12), (6, 15), (12, 4), (16, 13), (16, 14),$$

which is denoted by a "×" in the table.

4.2 DIAGNOSTICS APPLIED TO THE DATA

In this section, we apply the diagnostic methodology developed in Section 3 to the archeological data $\{Z(x, y) : x, y = 1, \ldots, 16\}$. For the mean structure of the linear model, we use

$$E(Z(x, y)) = (1, \ x, \ y)\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$. We obtain parameter estimates for every combination of models (the three CAR models described in Section 3.1) and neighborhood structures (the NN and 2NN structures described in Section 3.4). Table 2 gives $\hat{\gamma}$, the bounds $\gamma_{\min}$, $\gamma_{\max}$ for γ as obtained from (3.6), and the values $\gamma_{\min} + \epsilon$, $\gamma_{\max} - \epsilon$, where $\epsilon = 10^{-4}$, for each of the six model combinations. We observe that, for most of the model-neighborhood combinations, the estimated value $\hat{\gamma}$ of γ was close to the maximum possible value, $\gamma_{\max}$, for that model.

Table 1. Archeological dataset: Phosphate concentrations at locations $\{(x, y) : x, y = 1, \ldots, 16\}$ in Laconia, Greece. Missing values are indicated by "×."

16	121	112	108	91	68	59	294	50	101	27	71	48	36	71	66	83
15	108	101	75	83	×	×	52	55	50	41	30	47	47	55	75	108
14	62	80	50	88	77	77	73	50	50	59	57	55	57	38	71	×
13	17	52	60	91	166	68	60	32	47	45	34	57	60	64	68	×
12	32	48	27	88	×	×	116	66	34	62	77	41	23	38	68	68
11	73	33	60	66	×	×	62	143	60	62	80	59	75	57	27	57
10	55	53	80	80	62	91	71	68	77	104	75	41	33	131	41	37
9	64	45	62	21	60	38	47	77	73	62	27	44	53	53	52	36
8	64	28	44	45	60	62	34	47	75	83	71	77	83	73	77	59
7	59	38	32	55	60	30	41	59	57	71	66	83	85	85	77	83
6	45	47	48	68	80	44	64	64	68	68	88	116	108	85	91	73
5	37	41	38	36	19	57	47	131	80	83	80	88	73	73	97	62
4	31	45	34	66	71	85	80	121	91	136	108	×	108	80	80	73
3	55	34	62	41	80	75	101	50	71	91	94	94	91	75	68	59
2	57	55	66	40	57	68	73	80	71	125	83	66	77	71	47	55
1	77	59	45	55	59	60	48	68	71	57	60	55	53	57	62	64
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

(with *y* labeling the rows on the left, and *x* labeling the columns at the bottom.)

(See Besag and Kooperberg 1995, for an explanation of why this might be expected.) Hence, we expect that the diagnostics we obtain should be similar when we plug in $\gamma_0 = \hat{\gamma}$ or $\gamma_0 = \gamma_{\max} - \epsilon$, and we expect dissimilarities between the models with $\gamma_0 = \hat{\gamma}$ and $\gamma_0 = \gamma_{\min} + \epsilon$.

In what follows, we use diagnostics based on $g(x) = e^x$. Figure 1 shows a MRF-Neighborhoods plot, which was described in Section 3.2; this plot is for the model (ACAR, 2NN, $\hat{\gamma}$), a model that fitted the data comparatively well. We see both positive and negative outliers, with a few noticeably high residuals. Apart from these spatial outliers, we do not observe any obvious indications of model misfit. On the other hand, in Figure 2, which shows the MRF-Neighborhoods plot for the the model (ACAR, 2NN, $\gamma_{\min} + \epsilon$),

Table 2. Smallest ($\gamma_{\min}$) and largest ($\gamma_{\max}$) possible values for γ for the various CAR-model combinations. The parameter space for γ is ($\gamma_{\min}, \gamma_{\max}$); the values of γ_0 used near the end points of the parameter space, $\gamma_{\min} + \epsilon$ and $\gamma_{\max} - \epsilon$, were obtained by choosing $\epsilon = 10^{-4}$. Also shown are the maximum likelihood estimates $\hat{\gamma}$ for the archeological dataset.

Models		$\gamma_{\min}$	$\gamma_{\max}$	γ_0		
				$\gamma_{\min} + \epsilon$	$\hat{\gamma}$	$\gamma_{\max} - \epsilon$
HCAR	NN	−0.2565	0.2565	−0.2564	0.2321	0.2564
	2NN	−0.2453	0.0880	−0.2452	0.0866	0.0879
WCAR	NN	−1.0000	1.0000	−0.9999	0.8222	0.9999
	2NN	−2.4290	1.0000	−2.4289	0.9359	0.9999
ACAR	NN	−0.2565	0.2565	−0.2564	0.2320	0.2564
	2NN	−0.2453	0.0880	−0.2452	0.0863	0.0879

NOTE: From Section 3.2, $\gamma_{\min}$ and $\gamma_{\max}$ are the same for the HCAR and ACAR models. All the values shown are rounded to the fourth decimal place.
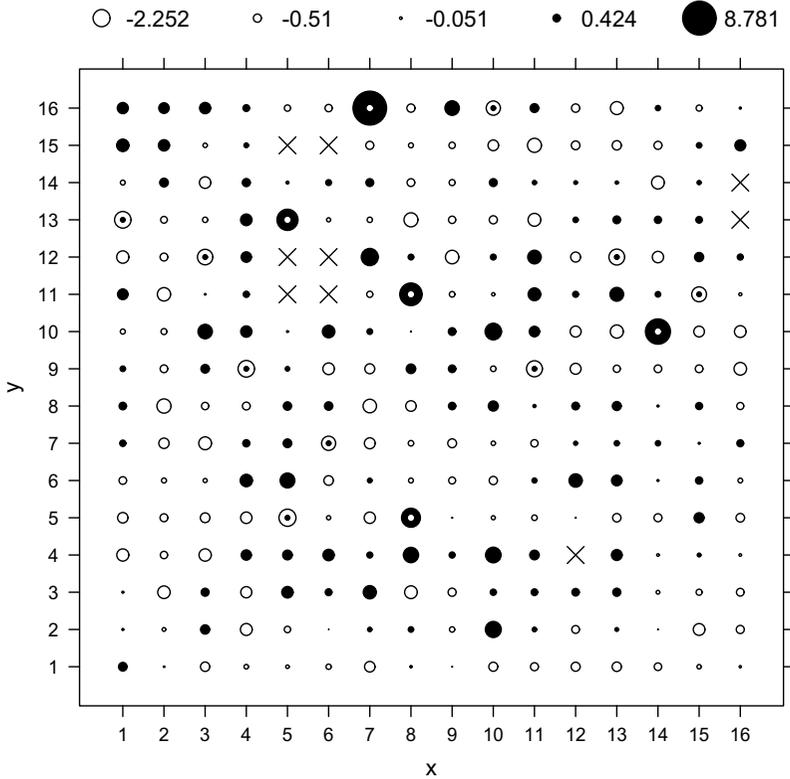
Figure 1. MRF-Neighborhoods plot based on (ACAR, 2NN, $\hat{y} = 0.0863$), for the archeological dataset. Solid black bubbles represent positive residuals, and empty white bubbles represent negative residuals. A white dot inside a solid black bubble flags that site as being above the 97.5 percentile of the simulated residual distribution, and a black dot inside an empty white bubble flags that site as being below the 2.5 percentile. A scale showing the size of the bubbles for the five quantiles of the residuals is shown at the top. Sites with missing data are marked by "×."

we observe quite obvious nonrandom patterns in the residuals; the positive residuals seem to be clustered in two distinct groups. Moreover, all the outliers are positive. This plot indicates a model that does not fit the data nearly as well as the model that yields Figure 1.

Next, we give the diagnostic statistic MSE_W defined by (3.7), which recall represents the global behavior of different models. Table 3 shows MSE_W for different CAR model types and for different choices of γ_0. From this table, we observe that for all types of CAR models (except where $\gamma_0 = \gamma_{min} + \epsilon$) considered, the NN structures give higher MSE_W-values (i.e., fit less well) than the 2NN structures. Furthermore, for both neighborhood structures, the ACAR model appears to do better than the others. So, if we were to choose one model for these data from those presented, the model (ACAR, 2NN, $\hat{y}$) is the one suggested by the diagnostic statistic MSE_W. Another interesting observation is that for the NN structure and $\gamma_0 = 0$, the values of MSE_W are all smaller than the respective values of MSE_W for the NN structure and the three other values of γ_0. This indicates that the choice of neighborhood structure can play an important role in comparing models with
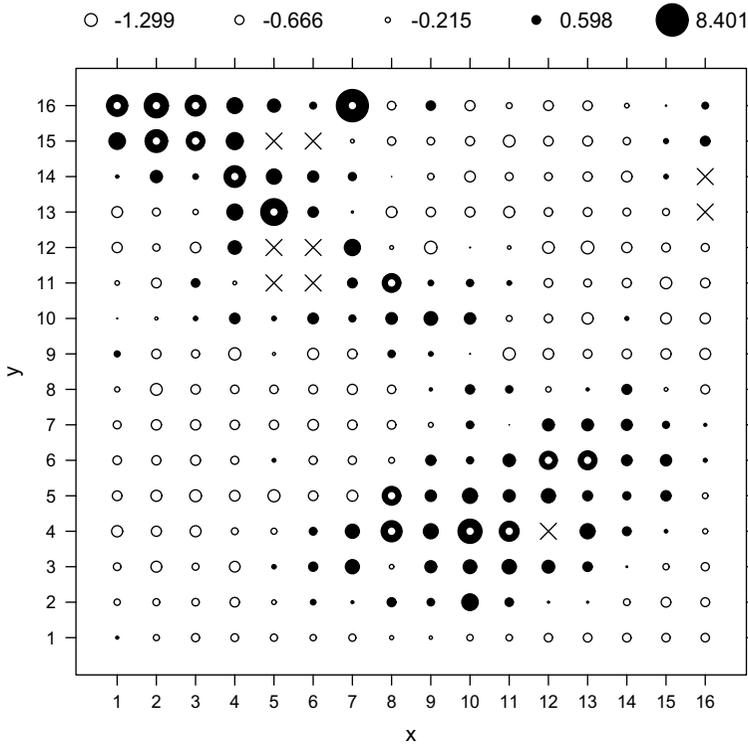
Figure 2.　MRF-Neighborhoods plot based on (ACAR, 2NN, $\gamma_{\min} + \epsilon = -0.2452$), for the archeological dataset. The caption for Figure 1 explains the symbols used in the figure.

and without spatial dependence. For this archeological dataset, the NN structure has to be augmented to a 2NN structure in order for the fitted CAR model to take spatial variability out from the mean and put it into the spatial dependence.

　　Figure 3 shows the Model-Comparison plot, which was described in Section 3.3. The two models compared here are Model A: (ACAR, NN, $\hat{\gamma}$), and Model B: (ACAR, 2NN, $\hat{\gamma}$). The dotted horizontal line represents the difference in the mean squared errors for the two models, namely, $\mathrm{MSE}_W^{(A)} - \mathrm{MSE}_W^{(B)}$. The plot in Figure 3 is targeted at a comparison of the NN and the 2NN structures. Clearly, there is one site where Model A performs poorly as compared to Model B. The site in question is (7, 16) in Table 1. We looked at many other such Model-Comparison plots and found that the NN structure was generally not able to capture the "influential" nature of this site as well as the 2NN structure. Returning to Figure 1, the MRF-Neighborhoods plot also shows that site (7, 16) is unusual. Therefore, in Section 4.3, we removed the datum at (7, 16) and did a full reanalysis.

　　For the complete data, using our notation of representing the CAR models by an ordered triple, we see from Table 3 that the best (smallest) four values of MSE_W yield the

Table 3. Values of MSE_W for the various CAR-model combinations, fitted to the archeological dataset. (The smallest MSE_W-value is shown in bold.)

			γ_0		
Models		$\gamma_{\min} + \epsilon$	0	$\hat{\gamma}$	$\gamma_{\max} - \epsilon$
HCAR	NN	1.3318	1.2891	1.4681	1.5236
	2NN	1.7186	1.2891	1.2725	1.2714
WCAR	NN	1.3501	1.2565	1.5082	1.9086
	2NN	1.7805	1.2128	1.2523	1.2816
ACAR	NN	1.3269	1.2565	1.3915	1.4304
	2NN	1.7178	1.2128	**1.1936**	1.1943

NOTE: The values shown are rounded to the fourth decimal place.

following MSE_W-inequalities:

$$(\text{ACAR, 2NN}, \hat{\gamma}) < (\text{ACAR, 2NN}, \gamma_{\max} - \epsilon) < (\text{ACAR, 2NN}, 0)$$
$$< (\text{WCAR, 2NN}, \hat{\gamma}).$$

Of course, with $\gamma_0 = 0$, the ACAR and WCAR models are the same.

4.3 MODIFIED ARCHEOLOGICAL DATASET

We removed the datum at site (7, 16) (see Table 1) that had an unusually high response, and we repeated the analyses given in Section 4.2. Table 4 gives analogous values of γ_0 obtained from the modified archeological dataset, and Figure 4 shows the analogous MRF-Neighborhoods plot for the model (ACAR, 2NN, $\hat{\gamma}$). Comparing Figure 4 with Figure 1, we can see that the general pattern of positive and negative residuals is the same in both cases.

The values of MSE_W for all the refitted models are provided in Table 5. Again, for all choices of γ_0, except $\gamma_{\min} + \epsilon$, the 2NN structure is preferred over the corresponding NN structure. We see from Table 5 that the best (smallest) four values of MSE_W yield the following MSE_W-inequalities

$$(\text{HCAR, 2NN}, \hat{\gamma}) < (\text{HCAR, 2NN}, \gamma_{\max} - \epsilon) < (\text{ACAR, 2NN}, \hat{\gamma})$$
$$< (\text{ACAR, 2NN}, \gamma_{\max} - \epsilon).$$

The model (HCAR, 2NN, $\hat{\gamma}$) has the lowest MSE_W-value, but it is not that much smaller than that for (ACAR, 2NN, $\hat{\gamma}$). Figure 5 shows the Model-Comparison plot between the two models. Notice that while HCAR is barely better overall, there are a few sites where the areas are large and positive and hence HCAR does not fit as well as ACAR at these sites. Figure 6 shows the Model-Comparison plot between the models (ACAR, NN, $\hat{\gamma}$) and (ACAR, 2NN, $\hat{\gamma}$); clearly, the 2NN structure gives a better model, but not because of the presence of any unusual site, in contrast to Figure 3. The model (ACAR, 2NN, $\hat{\gamma}$) performed well whether site (7, 16) was included or not, and hence we conclude from our diagnostics that this would be a good model choice.
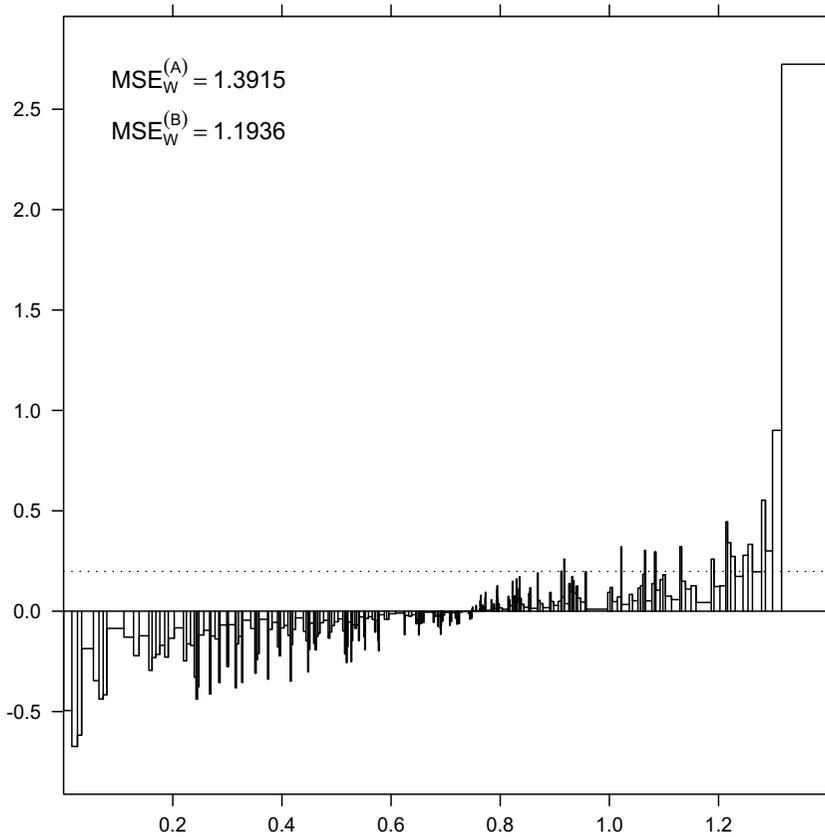
Figure 3.   Model-Comparison plot based on Model A: (ACAR, NN, $\hat{\gamma}$ = 0.232) and Model B: (ACAR, 2NN, $\hat{\gamma}$ = 0.0863), for the archeological dataset. The plot is described in Section 3.3. The horizontal dotted line represents $\text{MSE}_W^{(A)} - \text{MSE}_W^{(B)}$.

5. DATASET OF DOCTORS' PRESCRIPTION AMOUNTS

These spatial data are on an irregular lattice in $\mathbb{R}^2$, where the sites are in fact small areas (cantons) in the southwest of France.

5.1 DESCRIPTION OF THE DATA

The data consist of average prescription amounts per doctor consultation, during the period January 1, 1999–December 31, 1999, in the region in southwest France known as the Midi-Pyrénées, and they have been analyzed previously by Cressie, Perrin, and Thomas-Agnan (2005, 2006). The Midi-Pyrénées is made up of contiguous *cantons*, for which we have data for 268 of them. Each canton, identified by the X and Y coordinates of its centroid (in meters), provides one record. The dataset also provides the percentage of patients 70 or older, the per-capita income, and the number of consultations during 1999 for each canton. The general idea is to look at the spatial dependence of patients' average
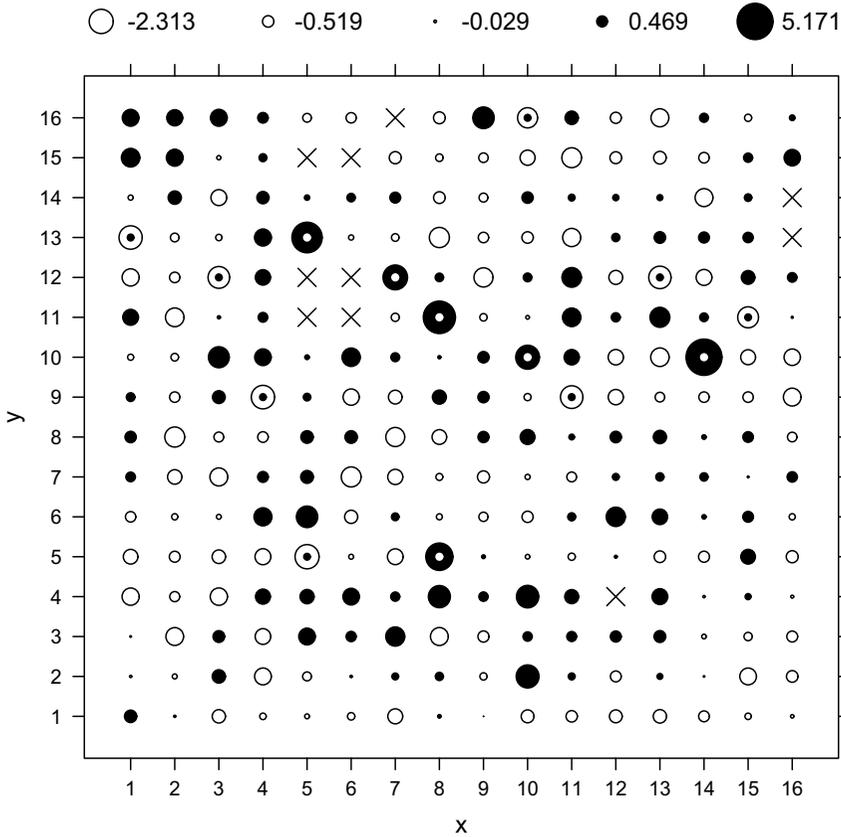
Figure 4.   MRF-Neighborhoods plot based on (ACAR, 2NN, $\hat{y} = 0.0863$), refitted to the modified archeological dataset after removing site $(7, 16)$. The caption for Figure 1 explains the symbols used in the figure.

prescription amounts per doctor consultation. Figure 1 in Cressie, Perrin, and Thomas-Agnan (2005) shows the Lambert projection of the cantons.

Let D_i denote the average prescription amount for the ith canton; $i = 1, \ldots, 268$. Based on some initial exploratory data analyses, it was decided to take logs of the data, which was not done in the analyses of Cressie, Perrin, and Thomas-Agnan (2005, 2006). We chose $Z_i = k \cdot \log_{10} D_i$, where k is an undisclosed constant meant to preserve confidentiality of the original data. Our analysis also differs from theirs in the type of CAR models chosen. In this article, the average prescription amounts are used to illustrate our diagnostics for the HCAR, WCAR, and ACAR models (Section 3.1) fitted to irregular lattice data. Cressie, Perrin, and Thomas-Agnan's analyses effectively weight the untransformed data according to the square roots of the number of consultations in each of the 268 cantons.

Table 4.   Values of γ_0 used in the refitted CAR models for the modified archeological dataset after removing site (7, 16).

Models		γ_{min}	γ_{max}	γ_0		
				$\gamma_{min} + \epsilon$	$\hat{\gamma}$	$\gamma_{max} - \epsilon$
HCAR	NN	−0.2565	0.2565	−0.2564	0.2403	0.2564
	2NN	−0.2454	0.0880	−0.2453	0.0866	0.0879
WCAR	NN	−1.0000	1.0000	−0.9999	0.8634	0.9999
	2NN	−2.2836	1.0000	−2.2835	0.9423	0.9999
ACAR	NN	−0.2565	0.2565	−0.2564	0.2391	0.2564
	2NN	−0.2454	0.0880	−0.2453	0.0863	0.0879

NOTE: The caption for Table 2 gives the necessary explanations for table entries. The values shown are rounded to the fourth decimal place.

5.2   DIAGNOSTICS APPLIED TO THE DATA

In this section, we apply the diagnostic methodology developed earlier, to the prescription data $\{Z_i : i = 1, \ldots, 268\}$. For the mean structure of the linear model, we use

$$E(Z_i) = (1, \; u_i)\boldsymbol{\beta},$$

where u_i is the percentage of patients over the age of 70 in the ith canton and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Like Cressie, Perrin, and Thomas-Agnan (2005, 2006), we found that $\{u_i\}$ captures a lot of the variability in $\{Z_i\}$. The distance between two cantons is obtained by calculating the Euclidean distance between their centroids (distances are small enough to ignore the Earth's curvature). As described in Section 3.5, we constructed the neighborhood structure for these data using different values of $d \in \{20, 30, 40, \ldots, 180\}$ (in units of km). For each choice of d, (3.11) provides a neighborhood structure $N^{(d)} = \{N_i^{(d)} : i = 1, \ldots, 268\}$, which is used to fit the three types of CAR models given in Section 3.1. As for the archeological dataset, we use four different choices for γ_0, namely, $\gamma_{min} - \epsilon$, 0, $\hat{\gamma}$, and, $\gamma_{max} + \epsilon$, where $\epsilon = 10^{-4}$.

As mentioned in Section 3.5, one strength of our diagnostic methodology is being able

Table 5.   Values of MSE_W for the various CAR-model combinations, fitted to the modified archeological dataset after removing site (7, 16). (The smallest MSE_W-value is shown in bold.)

Models		γ_0			
		$\gamma_{min} + \epsilon$	0	$\hat{\gamma}$	$\gamma_{max} - \epsilon$
HCAR	NN	1.1797	1.0164	1.0022	1.0361
	2NN	1.5885	1.0164	**0.9769**	0.9789
WCAR	NN	1.2216	1.0131	1.0102	1.1029
	2NN	1.6200	1.0099	0.9905	1.0042
ACAR	NN	1.1852	1.0131	1.0020	1.0465
	2NN	1.6152	1.0099	0.9826	0.9875

NOTE: The values shown are rounded to the fourth decimal place.
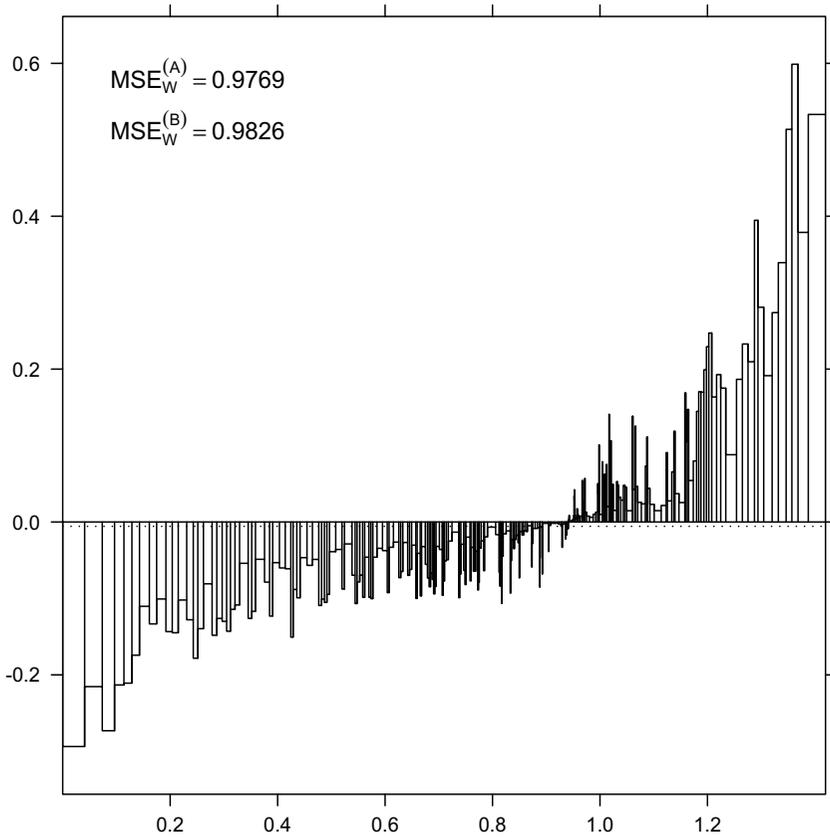
Figure 5. Model-Comparison plot based on Model A: (HCAR, 2NN, $\hat{\gamma} = 0.0866$) and Model B: (ACAR, 2NN, $\hat{\gamma} = 0.0863$), refitted to the modified archeological dataset after removing site (7, 16). The horizontal dotted line represents $\mathrm{MSE}_W^{(A)} - \mathrm{MSE}_W^{(B)}$.

to visualize the effect on the spatial model of changing the neighborhood structure. Figure 7 illustrates how the MSE_W, given by (3.7), changes with d for all combinations of model types and $\gamma_0 = \hat{\gamma}$ and $\gamma_{\max} - \epsilon$. The plot based on the maximum likelihood estimator, $\gamma_0 = \hat{\gamma}$, deserves particular attention. Clearly, $d = 60$ km provides the best neighborhood structure for most of the models. It is interesting to note that for large d, MSE_W-values increase; that is, there is an inherent penalization for overly complex neighborhood structures.

To make model comparisons, we use two neighborhood structures, $N^{(30)}$ and $N^{(60)}$; Cressie, Perrin, and Thomas-Agnan (2005, 2006) used $N^{(30)}$. Table 6 gives the values of $\hat{\gamma}$, $\gamma_{\min}$, $\gamma_{\max}$, from which the values of γ_0 used in the model comparisons are obtained. As for the archeological dataset, $\hat{\gamma}$ here is close to $\gamma_{\max}$, and hence we expect similar performances for the three CAR models with $\gamma_0 = \hat{\gamma}$ and $\gamma_0 = \gamma_{\max} - \epsilon$.

Figure 8 shows the MRF-Neighborhoods plot for the model (WCAR, $N^{(60)}, \hat{\gamma}$), as described in Section 3.2. This model seems to fit very well with only a few positive and negative outliers. On the other hand, the MRF-Neighborhoods plot for the model
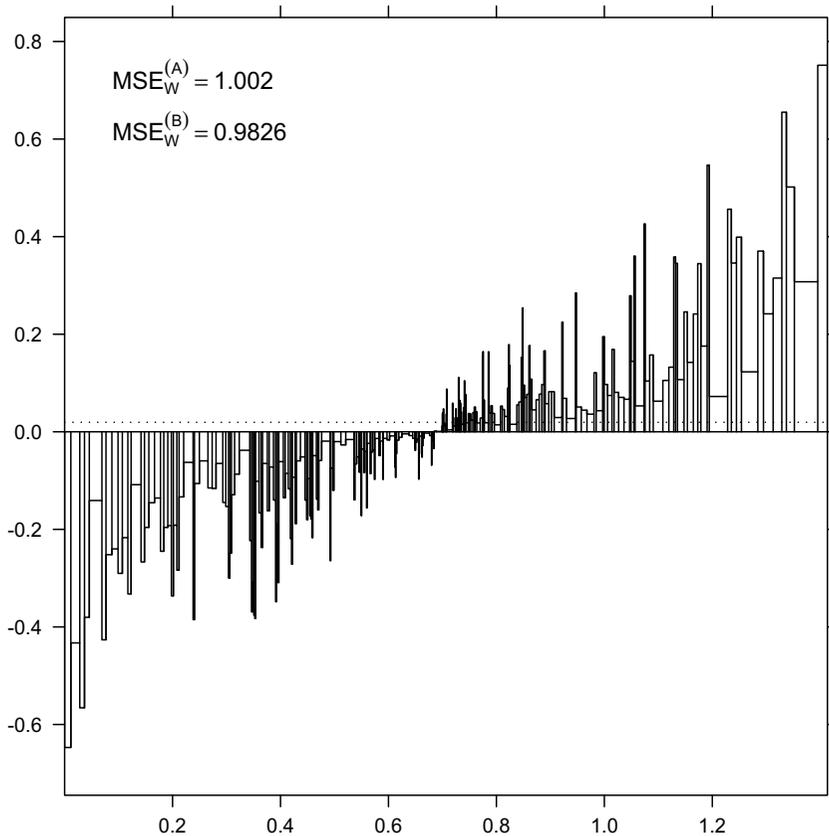
Figure 6.   Model-Comparison plot based on Model A: (ACAR, NN, $\hat{\gamma}$ = 0.2391) and Model B: (ACAR, 2NN, $\hat{\gamma} = 0.0863$), refitted to the modified archeological dataset after removing site (7, 16). The horizontal dotted line represents $\mathrm{MSE}_W^{(A)} - \mathrm{MSE}_W^{(B)}$.

(WCAR, $N^{(60)}$, $\gamma_{\min}+\epsilon$) given in Figure 9 shows spatial clustering of positive and negative residuals. Such a plot indicates that the model does not fit the data very well.

Next, we look at the diagnostic statistic, MSE_W, defined by (3.7). Table 7 shows MSE_W for different models and for different choices of γ_0. We observe that for the three types of CAR models (except where $\gamma_0 = \gamma_{\min} + \epsilon$) considered, the neighborhood structure $N^{(30)}$ gives higher MSE_W-values (i.e., fits less well) than the corresponding neighborhood structure $N^{(60)}$. Furthermore, for $N^{(60)}$ (again, except where $\gamma_0 = \gamma_{\min} + \epsilon$), WCAR models and ACAR models perform comparably.

In Figure 10, we give a Model-Comparison plot between two models that have large MSE_W differences, in an attempt to see if there are any unusual sites. We compare Model A: (WCAR, $N^{(60)}$, $\gamma_{\min}+\epsilon$), and Model B: (WCAR, $N^{(60)}$, $\hat{\gamma}$); clearly, Figure 10 does not show any unusual areas contributing to the difference. We conclude that Model B, based on $\gamma_0 = \hat{\gamma}$, performs consistently better than Model A, based on $\gamma_0 = \gamma_{\min} + \epsilon$.

Based on the tables and figures given, we see that for the dataset of doctors' prescription
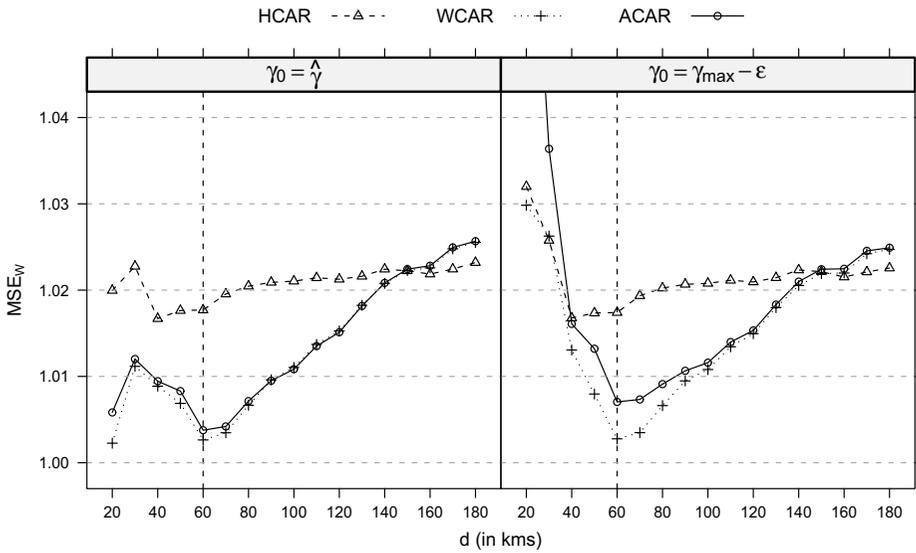
Figure 7. Plot of MSE_W-values vs. d (in km), for the dataset of doctors' prescription amounts. The three types of CAR models, grouped by γ_0, are shown: '$\triangle$' joined by a dashed line (– – –) represents the HCAR models; "+" joined by dotted a line ($\cdots$) represents the WCAR models; and '$\circ$' joined by solid a line (—) represents the ACAR models. The choice of γ_0 is specified at the top of each panel.

amounts, ACAR models and WCAR models perform very similarly for most of the model combinations and are superior to the HCAR models. The neighborhood structure $N^{(60)}$ outperforms the neighborhood structure $N^{(30)}$ for $\gamma_0 = \hat{\gamma}$ and $\gamma_0 = \gamma_{\max} - \epsilon$. We see from Table 7 that the best (smallest) four values of MSE_W yield the following MSE_W-inequalities:

$$(\text{WCAR}, \ N^{(60)}, \hat{\gamma}) < (\text{WCAR}, \ N^{(60)}, \gamma_{\max} - \epsilon) < (\text{ACAR}, \ N^{(60)}, \hat{\gamma})$$
$$< (\text{ACAR}, \ N^{(60)}, \gamma_{\max} - \epsilon).$$

Table 6. Values of γ_0 used in the fitted CAR models for the dataset of doctors' prescription amounts.

| | | | | | γ_0 | |
Models		$\gamma_{\min}$	$\gamma_{\max}$	$\gamma_{\min} + \epsilon$	$\hat{\gamma}$	$\gamma_{\max} - \epsilon$
HCAR	$N^{(30)}$	−0.2130	0.0552	−0.2129	0.0474	0.0551
	$N^{(60)}$	−0.0920	0.0166	−0.0919	0.0153	0.0165
WCAR	$N^{(30)}$	−2.6562	1.0000	−2.6561	0.7442	0.9999
	$N^{(60)}$	−5.4895	1.0000	−5.4894	0.8976	0.9999
ACAR	$N^{(30)}$	−0.2130	0.0552	−0.2129	0.0461	0.0551
	$N^{(60)}$	−0.0920	0.0166	−0.0919	0.0158	0.0165

NOTE: The caption for Table 2 gives the necessary explanations for table entries. The values shown are rounded to the fourth decimal place.
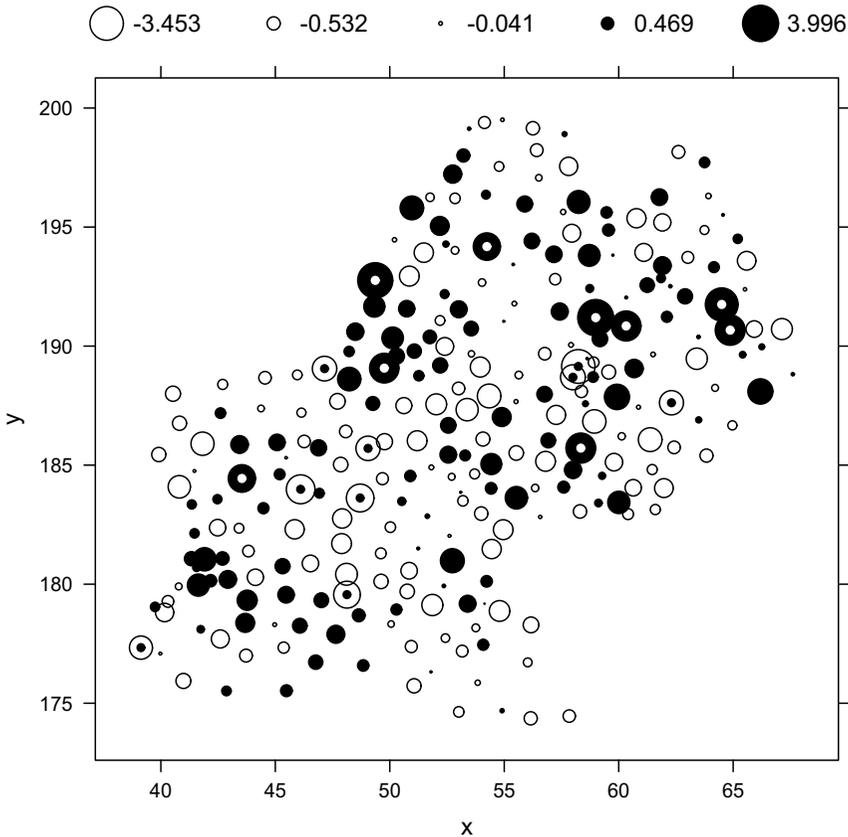
Figure 8.   MRF-Neighborhoods plot based on (WCAR, $N^{(60)}$, $\hat{\gamma} = 0.8976$), for the dataset of doctors' pre-scription amounts. The caption for Figure 1 explains the symbols used in the figure.

6. DISCUSSION AND CONCLUSIONS

In this article, we prove a general result (not just for CAR models but for any MRF) that depends on an assumed neighborhood structure, and we use it to define diagnostics for departures from the assumed structure. The result also encompasses the generality of the choice of a test function; for the examples considered, we use $g(x) = e^x$ for its sensitivity in detecting local departures. We use two examples, one on a regular lattice and one on an irregular lattice, to illustrate our diagnostic methodology for CAR models. Based on local residuals $\mathbf{W}^*$ and their standardized version $\mathbf{W}$, we develop the MRF-Neighborhoods plot and the Model-Comparison plot, along with the diagnostic statistic MSE_W. In the second example, we show graphically the effect of increasing neighborhood complexity on how well the CAR models fit. Typically, one expects such behavior but was hitherto unable to explore it, and we see this as an important contribution of our article. In summary, our diagnostics show that the neighborhood structure plays a crucial role in determining whether a given MRF model is appropriate or not.

From the MRF-Neighborhood plot (e.g., Figures 1, 2, 4, 8, and 9), we are able to
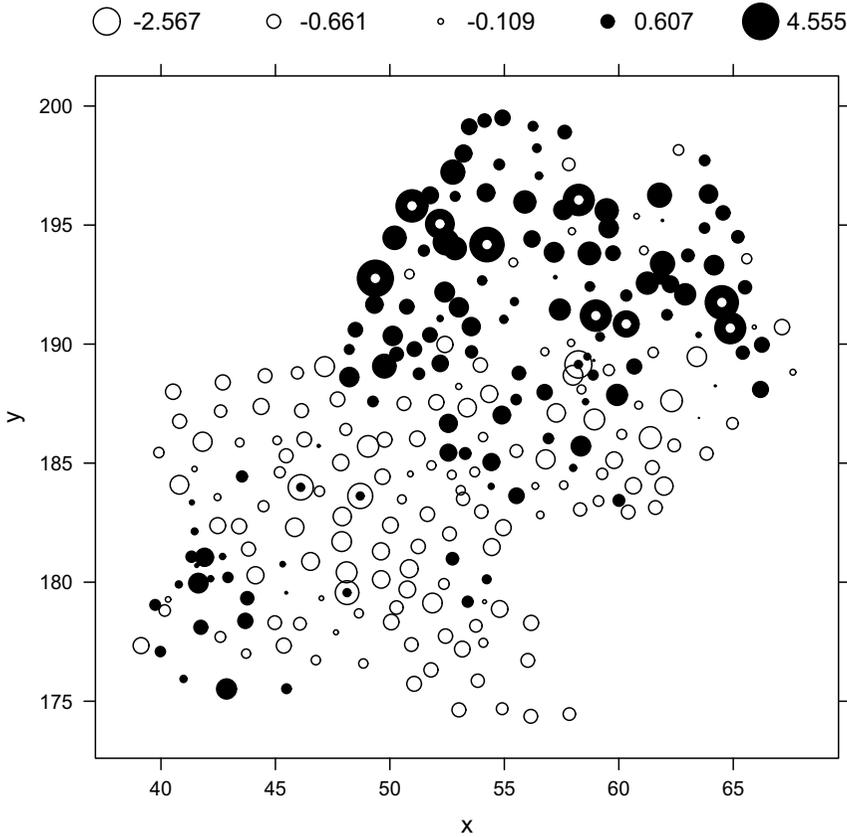
Figure 9. MRF-Neighborhoods plot based on (WCAR, $N^{(60)}$, $\gamma_{\min}+\epsilon = -5.4894$), for the dataset of doctors' prescription amounts. The caption for Figure 1 explains the symbols used in the figure.

determine departures from an imposed neighborhood structure, locally, site-by-site. We are also able to identify unusual sites where there are spatial "outliers." Next, through the global diagnostic statistic, MSE_W, we can compare directly all plausible models (e.g., Tables 3, 5, and 7). Once the class of plausible models is reduced, we can use the Model-Comparison plots (e.g., Figures 3, 5, 6, and 10) to compare competing models in a local manner.

Suggestions for future work include extending our methodology to the Intrinsic Autoregressive (IAR) processes, developing a global model-selection statistic based on **W** that includes a penalty for the number of model parameters (see the AIC and BIC model-selection criteria), and studying the effect of using conditional variances (from Result 2) at the sites to standardize the components of $\mathbf{W}^*$ (rather than using the joint variance matrix).

APPENDIX: RESULTS WITH PROOFS

Before establishing our results, some notation and a few basic properties are needed. The results presented here are for general choices of $\boldsymbol{\mu}$, M, and C in $\text{Gau}\left(\boldsymbol{\mu},\ (I - C)^{-1}M\right)$,

Table 7. Values of MSE_W for the various CAR-model combinations, fitted to the dataset of doctors' prescription amounts. (The smallest MSE_W-value is shown in bold.)

Models		γ_0			
		$\gamma_{\min} + \epsilon$	0	$\hat{\gamma}$	$\gamma_{\max} - \epsilon$
HCAR	$N^{(30)}$	1.1037	1.0241	1.0228	1.0258
	$N^{(60)}$	1.1216	1.0241	1.0177	1.0174
WCAR	$N^{(30)}$	1.0535	1.0122	1.0112	1.0263
	$N^{(60)}$	1.1446	1.0090	**1.0026**	1.0028
ACAR	$N^{(30)}$	1.1017	1.0122	1.0120	1.0364
	$N^{(60)}$	1.1092	1.0090	1.0038	1.0070

NOTE: The values shown are rounded to the fourth decimal place.

where $M^{-1}(I - C)$ is symmetric and positive-definite for $M = \text{diag}(\tau_1^2, \ldots, \tau_n^2)$ and $C = (c_{ij})_{n \times n}$.

(a) A Gaussian *random variable X* is denoted as: $X \sim \text{Gau}(\nu, \sigma^2)$. Its moment generating function is given by $E\left(e^{tX}\right) = \exp\left\{\nu t + \sigma^2 t^2/2\right\}$, which implies that

$$E\left(e^X\right) = \exp\{\nu + \sigma^2/2\}.$$

(b) In the CAR model (3.1), notate $I - C = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n]'$; that is, the ith row of $I - C$ is $\boldsymbol{\delta}_i' \equiv (\delta_{i1}, \ldots, \delta_{in})$. Then,

$$\delta_{ii} = 1 - c_{ii} = 1, \quad \delta_{ij} = -c_{ij} \quad \text{for} \quad i \neq j.$$

Note that C need not be symmetric, but $M^{-1/2}CM^{1/2}$ must be.

(c) Define $\mathbf{e}_i \equiv (0, \ldots, 0, 1, 0, \ldots, 0)'_{n \times 1}$, an n-dimensional vector with 1 in the ith position and zeros elsewhere.

Then using (b) we obtain,

$$\boldsymbol{\delta}_i'(I - C)^{-1}M\boldsymbol{\delta}_i = \mathbf{e}_i' M\boldsymbol{\delta}_i = \tau_i^2 \mathbf{e}_i'\boldsymbol{\delta}_i = \tau_i^2 \delta_{ii} = \tau_i^2; \quad i = 1, \ldots, n,$$

and for $i \neq j \in \{1, \ldots, n\}$,

$$\boldsymbol{\delta}_i'(I - C)^{-1}M\boldsymbol{\delta}_j = \mathbf{e}_i' M\boldsymbol{\delta}_j = \tau_i^2 \mathbf{e}_i'\boldsymbol{\delta}_j = \tau_i^2 \delta_{ji} = -\tau_i^2 c_{ji}.$$

Result 1. *For $g(x) = e^x$ in (2.2), the CAR model (3.1) implies that*

$$m(\mathbf{Z}(N_i)) = \exp\left\{\mu_i + \sum_{j \in N_i} c_{ij}(Z(j) - \mu_j) + \frac{1}{2}\tau_i^2\right\}; \quad i = 1, \ldots, n.$$
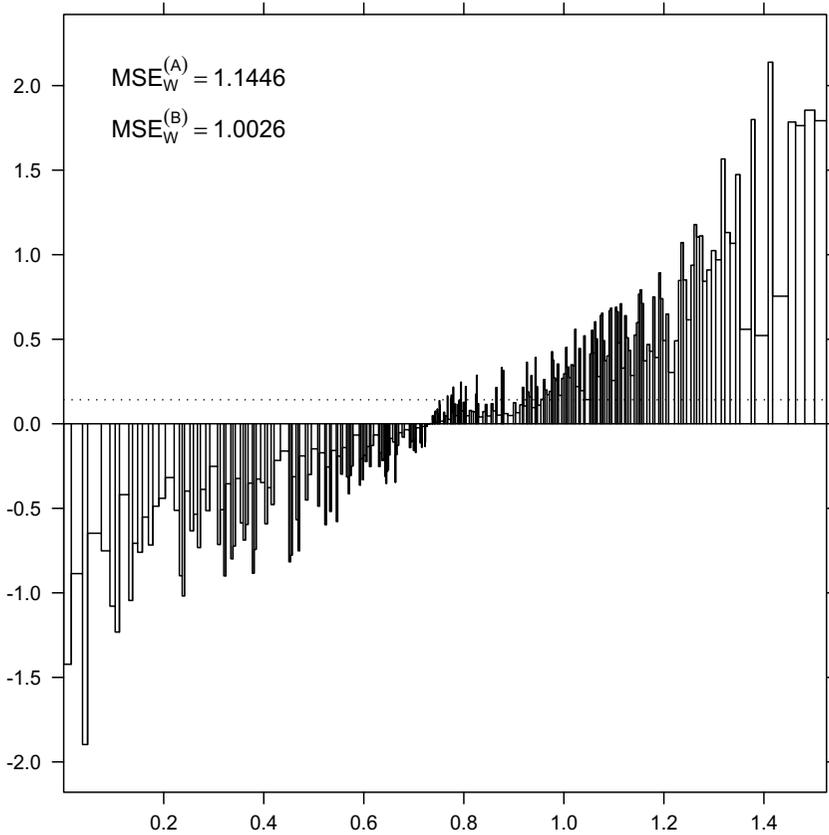
Figure 10. Model-Comparison plot based on Model A: (WCAR, $N^{(60)}$, $\gamma_{\min} + \epsilon = -5.4894$) and Model B: (WCAR, $N^{(60)}$, $\hat{\gamma} = 0.8976$), for the dataset of doctors' prescription amounts. The horizontal dotted line represents $\text{MSE}_W^{(A)} - \text{MSE}_W^{(B)}$.

Proof: The result follows trivially from (a) above by noting that

$$Z(i)|\mathbf{Z}(N_i) \sim \text{Gau}\Big(\mu_i + \sum_{j \in N_i} c_{ij}(Z(j) - \mu_j), \tau_i^2\Big); \quad i = 1, \ldots, n.$$

$\square$

Result 2. *For W_i^* defined in (3.4), $E(W_i^*|\mathbf{Z}(M_i)) = 1$, where recall from Section 2 that $M_i \equiv \overline{N_i \cup \{i\}}$; $i = 1, \ldots, n$. Furthermore, for the CAR model (3.1), $\text{var}(W_i^*|\mathbf{Z}(M_i)) = \exp\{\tau_i^2\} - 1$; $i = 1, \ldots, n$.*

Proof: From (2.1) and (2.2),

$$\begin{aligned}
E\big(W_i^*|\mathbf{Z}(M_i)\big) &= E\big(E\big(W_i^*|\mathbf{Z}(N_i), \mathbf{Z}(M_i)\big)\big) \\
&= E\big(E\big(\exp\{Z(i)\}|\mathbf{Z}(N_i), \mathbf{Z}(M_i)\big)/m(\mathbf{Z}(N_i))\big) \\
&= 1.
\end{aligned}$$

For the conditional variance, consider the following equality from Result 1:

$$E\big(\exp\{2\,Z(i)\}|\mathbf{Z}(N_i)\big) = \exp\Big\{2\,\mu_i + 2\sum_{j\in N_i} c_{ij}(Z(j)-\mu_j) + 2\,\tau_i^2\Big\}$$

$$= \exp\{\tau_i^2\}\,m(\mathbf{Z}(N_i))^2.$$

Hence,

$$\begin{aligned}
\mathrm{var}\big(W_i^*|\mathbf{Z}(M_i)\big) &= E\big(W_i^{*2}|\mathbf{Z}(M_i)\big) - 1\\
&= E\big(E\big(\exp\{2\,Z(i)\}|\mathbf{Z}(N_i),\mathbf{Z}(M_i)\big)/m(\mathbf{Z}(N_i))^2\big) - 1\\
&= \exp\{\tau_i^2\} - 1.
\end{aligned}$$

$\square$

Result 3. *From (3.4),* $\mathrm{var}(\mathbf{W}^*) \equiv \Sigma^* = e^B - J$, *where* $B = (I - C)M$.
Proof: From Result 2,

$$\begin{aligned}
\mathrm{var}(W_i^*) &= E\big(\mathrm{var}\big(W_i^*|\mathbf{Z}(M_i)\big)\big) + \mathrm{var}\big(E\big(W_i^*|\mathbf{Z}(M_i)\big)\big)\\
&= \exp\{\tau_i^2\} - 1 = \exp\{B_{ii}\} - 1.
\end{aligned}$$

Fix $i \neq j \in \{1,\ldots,n\}$.

$$\begin{aligned}
W_i^* &= \exp\Big\{Z(i) - \mu_i - \sum_{j\in N_i} c_{ij}(Z(j)-\mu_j) - \tau_i^2/2\Big\}\\
&= \exp\{\boldsymbol{\delta}_i'(\mathbf{Z}-\boldsymbol{\mu}) - \tau_i^2/2\};
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}(W_i^*, W_j^*) &= E\big((W_i^*-1)(W_j^*-1)\big) = E(W_i^* W_j^*) - 1\\
&= E\big(\exp\{\boldsymbol{\delta}_i'(\mathbf{Z}-\boldsymbol{\mu}) - \tau_i^2/2\}\exp\{\boldsymbol{\delta}_j'(\mathbf{Z}-\boldsymbol{\mu}) - \tau_j^2/2\}\big) - 1\\
&= \exp\{-(\tau_i^2+\tau_j^2)/2\}E\big(\exp\{(\boldsymbol{\delta}_i+\boldsymbol{\delta}_j)'(\mathbf{Z}-\boldsymbol{\mu})\}\big) - 1.
\end{aligned}$$

From (3.2), we have,

$$\mathbf{Z} - \boldsymbol{\mu} \sim \mathrm{Gau}\big(\mathbf{0}, (I-C)^{-1}M\big).$$

Thus,

$$(\boldsymbol{\delta}_i+\boldsymbol{\delta}_j)'(\mathbf{Z}-\boldsymbol{\mu}) \sim \mathrm{Gau}\big(0, (\boldsymbol{\delta}_i+\boldsymbol{\delta}_j)'(I-C)^{-1}M(\boldsymbol{\delta}_i+\boldsymbol{\delta}_j)\big).$$

Using (c), the variance simplifies to:

$$(\boldsymbol{\delta}_i+\boldsymbol{\delta}_j)'(I-C)^{-1}M(\boldsymbol{\delta}_i+\boldsymbol{\delta}_j) = \tau_i^2 - \tau_i^2 c_{ji} - \tau_j^2 c_{ij} + \tau_j^2.$$

Hence,

$$\begin{aligned}
\mathrm{cov}(W_i^*, W_j^*) &= \exp\{-(\tau_i^2+\tau_j^2)/2\}\exp\{(\tau_i^2 - \tau_i^2 c_{ji} - \tau_j^2 c_{ij} + \tau_j^2)/2\} - 1\\
&= \exp\{-(\tau_i^2 c_{ji} + \tau_j^2 c_{ij})/2\} - 1.
\end{aligned}$$

Because $\tau_i^2 c_{ji} = \tau_j^2 c_{ij}$ (e.g., Cressie 1993, p. 407), B is symmetric. Hence,

$$\mathrm{cov}(W_i^*, W_j^*) = \exp\{B_{ij}\} - 1,$$

where B_{ij}, $i \neq j$, is an off-diagonal element of B. □

ACKNOWLEDGMENTS

REFERENCES

Anselin, L. (1995), "Local Indicators of Spatial Association-LISA," *Geographical Analysis*, 27, 93–115.

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

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 36, 192–236.

Besag, J., and Kooperberg, C. (1995), "On Conditional and Intrinsic Autoregression," *Biometrika*, 82, 733–746.

Besag, J., York, J. C., and Mollié, A. (1991), "Bayesian Image Restoration, With Two Applications in Spatial Statistics" (with discussion), *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Buck, C. E., Cavanagh, W. G., and Litton, C. D. (1988), "The Spatial Analysis of Site Phosphate Data," in *Computer Applications and Quantitative Methods in Archeology*, ed. S. P. Q. Rahtz, British Archeological Reports, International Series, vol. 446, 151–160.

Carlin, B. P., and Banerjee, S. (2003), "Hierarchical Multivariate CAR Models for Spatio-Temporally Correlated Survival Data," in *Bayesian Statistics*, eds. J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, Oxford: Oxford University Press, vol. 7, 45–63.

Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Duxbury.

Conlon, E., and Waller, L. (2000), "Flexible Neighborhood Structures in Hierarchical Models for Disease Mapping," in *American Statistical Association Proceedings of the Section on Statistics and the Environment*, Alexandria, VA: American Statistical Association, 82–87.

Cressie, N. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

Cressie, N., and Chan, N. (1989), "Spatial Modeling of Regional Variables," *Journal of the American Statistical Association*, 84, 393–401.

Cressie, N., and Collins, L. B. (2001), "Patterns in Spatial Point Locations: Local Indicators of Spatial Association in a Minefield with Clutter," *Naval Research Logistics*, 48, 333–347.

Cressie, N., Perrin, O., and Thomas-Agnan, C. (2005), "Likelihood-Based Estimation for Gaussian MRFs," *Statistical Methodology*, 2, 1–16.

——— (2006), "Doctors' Prescribing Patterns in the Midi-Pyrénées Region of France: Point-process Aggregation," in *Case Studies in Spatial Point Process Modelling*, eds. A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, New York: Springer, vol. 185 of *Springer Lecture Notes in Statistics*, pp. 183–195.

Getis, A., and Ord, J. (1992), "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24, 473–477.

Hrafnkelsson, B., and Cressie, N. (2003), "Hierarchical Modelling of Count Data With Application to Nuclear Fall-Out." *Environmental and Ecological Statistics*, 10, 179–200.

Waller, L. A., and Gotway, C. A. (2004), *Applied Spatial Statistics for Public Health Data*, New York: Wiley-Interscience.