THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

WP # 0095MSS-496-2009

Date June 24, 2009

BAYESIAN ANALYSIS OF CONDITIONAL AUTORIEGRESSIVE MODELS

Victor De Oliveira
Department of Management Science and Statistics
University of Texas at San Antonio

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

BAYESIAN ANALYSIS OF CONDITIONAL AUTOREGRESSIVE MODELS

Victor De Oliveira[1]
Department of Management Science and Statistics
The University of Texas at San Antonio
San Antonio, TX 78249, USA.
victor.deoliveira@utsa.edu

January 17, 2009

Abstract

Conditionally autoregressive (CAR) models have been extensively used for the analysis of spatial data in diverse areas, such as demography, economy, epidemiology and geography, as models for both latent and observed variables. In the latter case, the most common inferential method has been maximum likelihood, and the Bayesian approach has not been used much. This work proposes default (automatic) Bayesian analyses of CAR models. Two versions of Jeffreys prior, the independence Jeffreys and Jeffreys-rule priors, are derived for the parameters of CAR models and properties of the priors and resulting posterior distributions are obtained. The two priors and their respective posteriors are compared based on simulated data. Also, frequentist properties of inferences based on maximum likelihood are compared with those based on the Jeffreys priors and the uniform prior. Finally, the proposed Bayesian analysis is illustraded by fitting a CAR model to a phosphate dataset from an archeological region.

Key words: CAR model; Eigenvalues and eigenvectors; Frequentist properties; Integrated likelihood; Maximum likelihood; Spatial data; Weight matrix.

JEL Classifications: C11 and C31.

[1]This work was partially supported by National Science Foundation grant DMS-0719508, and a College of Business (The University of Texas at San Antonio) summer research grant.

1 Introduction

Conditional autoregressive (CAR) models are used to describe the spatial variation of quantities of interest in the form of summaries or aggregates over subregions. These models have been used to analyze data in diverse areas such as demography, economy, epidemiology and geography. The general goal of these spatial models is to unveil and quantify spatial relations present among the data, in particular, to quantify how quantities of interest vary as a function of explanatory variables and detect clusters of 'hot spots'. General accounts of CAR models, a class of Markov random fields, appear in Cressie (1993), Banerjee, Carlin and Gelfand (2004) and Rue and Held (2005).

CAR models have been extensively used in spatial statistics to model observed data (Cressie and Chan, 1989; Richardson, Guihenneuc and Lasserre, 1992; Bell and Broemeling, 2000; Militino, Ugarte and Garcia-Reinaldos, 2004; Cressie, Perrin and Thomas-Agnan, 2005), as well as (unobserved) latent variables and spatially varying random effects (Clayton and Kaldor, 1987; Sun, Tsutakawa and Speckman, 1999; Pettitt, Weir and Hart, 2002; see Banerjee, Carlin and Gelfand, 2004 for further references). In this work I consider the former use of CAR models, but note that the analysis proposed here may serve (or be the base) for the use of CAR models in a default Bayesian analysis for hierarchical models.

The most commonly used method to fit CAR models has been maximum likelihood (Cressie and Chan, 1989; Richardson et al. 1992; Cressie et al. 2005). Most results up to date on the behavior of inferences based on maximum likelihood estimators are asymptotic in nature and little is know about their behavior in small samples. The Bayesian approach, on the other hand, allows 'exact' inference without the need for asymptotic approximations. Although Bayesian analyses of CAR models have been extensively used to describe latent variables and spatially varying random effects in the context of hierarchical models, not much has been done on Bayesian analysis of CAR models to describe the observed data (with only rare exceptions, e.g., Bell and Broemeling, 2000). This may be due to lack of familiarity about what could be adequate priors for these models and lack of knowledge about frequentist properties of the Bayesian procedures.

The goal main of this work is to propose default (automatic) Bayesian analyses for CAR models and study some of their properties. Two versions of Jeffreys prior, called independence Jeffreys and Jeffreys-rule priors, are derived for the parameters of CAR models and results on propriety of the resulting posterior distributions and existence of posterior moments for the covariance parameters are established. It is found that some properties of the posterior distributions based on the proposed Jeffreys priors depend on a certain relation between the column space of the regression design matrix and the extreme eigenspaces of the spatial design matrix.

Simple Monte Carlo algorithms are described for sampling from the appropriate posterior distributions for the cases when the data are complete and when there are missing observations. Examples are presented based on simulated data to compare the two Jeffreys priors and their corresponding posterior distributions.

A computational experiment is performed to compare frequentist properties of inferences about the covariance parameters based on maximum likelihood with those based on the proposed Jeffreys priors and the uniform prior. It is found that frequentist properties of all the above procedures are adequate and similar to each other in most situations, except when the spatial association is strong or the mean of the observations is not constant. In this case inference about the 'spatial parameter' based on the independence Jeffreys prior has better frequentist properties than the procedures based on the other priors or ML. Finally, it is found that the independence Jeffreys prior is not very sensitive to some aspects of the design, such as sample size and regression design matrix, while the Jeffreys-prior displays strong sensitivity to the regression design matrix.

The organization of the paper is as follows. Section 2 describes the CAR model and the behavior of an integrated likelihood. Section 3 derives two versions of Jeffreys prior and provides properties of these priors and their corresponding posterior distributions in terms of propriety and existence of posterior moments of the covariance parameters. Section 4 describes simple Monte Carlo algorithms to sample from the appropriate posterior distributions, and provides some comparisons based on simulated data between the two versions of Jeffreys priors. Section 5 presents a simulation experiment to compare frequentist properties of inferences based on ML with those based on the two versions of Jeffreys priors and the uniform prior, and explores sensitivity of Jeffreys priors to some aspects of the design. The proposed Bayesian methodology is illustrated in Section 6 using a phosphate dataset from an archeological region in Greece. Conclusions are given in Section 7.

2 CAR Models

2.1 Description

Consider a geographic region that is partitioned into subregions indexed by integers $1, 2, \ldots, n$. This collection of subregions (or sites as they are also called) is assumed to be endowed with a neighborhood system, $\{N_i \ : \ i = 1, \ldots, n\}$, where N_i denotes the collection of subregions that, in a well defined sense, are neighbors of subregion i. This neighborhood system, which is key in determining the dependence structure of the CAR model, must satisfy that for any $i, j = 1, \ldots, n$, $j \in N_i$ if and only if $i \in N_j$ and $i \notin N_i$. An emblematic example commonly

used in applications is the neighborhood system defined in terms of geographic adjacency

$$N_i = \{j \ : \ \text{subregion } j \text{ shares a boundary with subregion } i\}, \quad i = 1, \ldots, n.$$

Other examples include neighborhood systems defined based on distance from the centroids of subregions or similarity of an auxiliary variable; see Cressie (1993, p. 554) and Case, Rosen and Hines (1993) for examples. This kind of specification is natural for modeling summary or aggregate data where similarity between subregions often depends on similarity of shared features.

For each subregion it is observed the variable of interest, Y_i, and a set of $p < n$ explanatory variables, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$. The CAR model for the responses, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, is formulated by specifying the set of full conditional distributions satisfying a form of auto-regression given by

$$(Y_i \mid \mathbf{Y}_{(i)}) \sim \mathrm{N}\Big(\mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} c_{ij}(Y_j - \mathbf{x}_j'\boldsymbol{\beta}), \sigma_i^2\Big) \ , \quad i = 1, \ldots, n, \tag{1}$$

where $\mathbf{Y}_{(i)} = \{Y_j \ : \ j \neq i\}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p$ are unknown regression parameters, and $\sigma_i^2 > 0$ and $c_{ij} \geq 0$ are covariance parameters, with $c_{ii} = 0$ for all i. For the set of full conditional distributions (1) to determine a well defined joint distribution for $\mathbf{Y}$, the matrices $M = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and $C = (c_{ij})$ must satisfy the conditions:

(a) $M^{-1}C$ is symmetric, which is equivalent to $c_{ij}\sigma_j^2 = c_{ji}\sigma_i^2$ for all $i, j = 1, \ldots, n$;
(b) $M^{-1}(I_n - C)$ is positive definite;

see Cressie (1993) or Rue and Held (2005) for examples and further details. When (a) and (b) hold, we would have that

$$\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, (I_n - C)^{-1}M),$$

where X is the $n \times p$ matrix with i^{th} row $\mathbf{x}_i'$, assumed to have full rank. This work considers models in which (possibly after a transformation) the matrices M and C satisfy:

(i) $M = \sigma^2 I_n$, where $\sigma^2 > 0$ is unknown;
(ii) $C = \phi W$, where ϕ is a 'spatial parameter' and $W = (w_{ij})$ is a known "weight" ("neighborhood") matrix that is nonnegative ($w_{ij} \geq 0$), symmetric and satisfies that $w_{ij} > 0$ if and only if sites i and j are neighbors (so $w_{ii} = 0$).

To guarantee that $I_n - \phi W$ is positive definite ϕ is required to belong to $(\lambda_n^{-1}, \lambda_1^{-1})$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the ordered eigenvalues of W, with $\lambda_n < 0 < \lambda_1$ since $\mathrm{tr}(W) = 0$. It immediately follows that (i) and (ii) imply that (a) and (b) hold. If $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \phi)$ denote the model parameters, then the parameter space of this model, $\Omega = \mathbb{R}^p \times (0, \infty) \times (\lambda_n^{-1}, \lambda_1^{-1})$, has

the distinctive feature that it depends on some aspects of the design (as it depends on W). Finally, the parameter value $\phi = 0$ corresponds to the case when $Y_i - \mathbf{x}_i'\boldsymbol{\beta} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

Many other CAR models that have been considered in the literature can be reduced to a model where (i) and (ii) hold by the use of an appropriate scaling of the data and covariates (Cressie et al., 2005). Suppose $\tilde{\mathbf{Y}}$ follows a CAR model with mean vector $\tilde{X}\boldsymbol{\beta}$, with $\tilde{X}$ of full rank, and covariance matrix $(I_n - C)^{-1}M$, where $M = \sigma^2 G$, G diagonal with known positive diagonal elements, and $C = \phi\tilde{W}$ with $\tilde{W}$ as in (ii) except that it is not necessarily symmetric, and M and C satisfy (a) and (b). Then, $\mathbf{Y} = G^{-\frac{1}{2}}\tilde{\mathbf{Y}}$ satisfies

$$\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, \sigma^2(I_n - \phi W)^{-1}),$$

where $X = G^{-\frac{1}{2}}\tilde{X}$ has full rank and $W = G^{-\frac{1}{2}}\tilde{W}G^{\frac{1}{2}}$ is nonnegative, symmetric and $w_{ij} > 0$ if and only if sites i and j are neighbors; the symmetry of W follows from condition (a) above. Hence, $\mathbf{Y}$ follows the CAR model satisfying (i) and (ii).

2.2 Integrated Likelihood

The likelihood function of $\boldsymbol{\eta}$ based on the observed data $\mathbf{y}$ is

$$L(\boldsymbol{\eta}; \mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} |\Sigma_\phi^{-1}|^{\frac{1}{2}} \exp\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})'\Sigma_\phi^{-1}(\mathbf{y} - X\boldsymbol{\beta})\}, \tag{2}$$

where $\Sigma_\phi^{-1} = I_n - \phi W$. Similarly to what is often done for Bayesian analysis of ordinary linear models, a sensible class of prior distributions for $\boldsymbol{\eta}$ is given by the family

$$\pi(\boldsymbol{\eta}) \propto \frac{\pi(\phi)}{(\sigma^2)^a}, \qquad \boldsymbol{\eta} \in \Omega, \tag{3}$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(\phi)$ is the 'marginal' prior of ϕ with support $(\lambda_n^{-1}, \lambda_1^{-1})$. The relevance of this class of priors will be apparent when it is shown that the Jeffreys priors derived here belong to this class. An obvious choice, used by Bell and Broemeling (2000), is to set $a = 1$ and $\pi(\phi) = \pi^U(\phi) \propto \mathbf{1}_{(\lambda_n^{-1}, \lambda_1^{-1})}(\phi)$, which I call the uniform prior ($\mathbf{1}_A(\phi)$ denotes the indicator function of the set A). Besides its lack of invariance, the uniform prior may not (arguably) be quite appropriate in some cases. For many datasets found in practice there is strong spatial correlation between observations measured at nearest neighbors, and such strong correlation is reproduced in CAR models only when the spatial parameter ϕ is quite close to one of the boundaries, λ_1^{-1} or λ_n^{-1} (Besag and Kooperberg, 1995). The spatial information contained in the uniform prior is somewhat in conflict with the aforementioned historical information since it assigns too little mass to models with substantial spatial correlation and too much mass to models with weak or no spatial correlation. In contrast, the Jeffreys priors derived here do not have this unappealing feature since, as would be seen, they are unbounded around

λ_1^{-1} and λ_n^{-1}, so they automatically assign substantial mass to spatial parameters near these boundaries. Although using such priors may potentially yield improper posteriors, it would be shown that the propriety of posterior distributions based on these Jeffreys priors depend on a certain relation between the column space of X and the extreme eigenspaces of W (eigenspaces associated with the largest and smallest eigenvalues) which is most likely satisfied in practice. Another alternative, suggested by Banerjee et al. (2004, p. 164), is to use a beta-type prior for ϕ that places substantial prior probability on large values of $|\phi|$, but this would require specifying two hyperparameters.

From Bayes theorem follows that the posterior distribution of $\boldsymbol{\eta}$ is proper if and only if $0 < \int_\Omega L(\boldsymbol{\eta}; \mathbf{y})\pi(\boldsymbol{\eta})d\boldsymbol{\eta} < \infty$. A standard calculation with the above likelihood and prior shows that

$$\int_{\mathbb{R}^p \times (0,\infty)} L(\boldsymbol{\eta}; \mathbf{y})\pi(\boldsymbol{\eta})d\boldsymbol{\beta}d\sigma^2 = L^I(\phi; \mathbf{y})\pi(\phi),$$

with

$$L^I(\phi; \mathbf{y}) \propto |\Sigma_\phi^{-1}|^{\frac{1}{2}}|X'\Sigma_\phi^{-1}X|^{-\frac{1}{2}}(S_\phi^2)^{-(\frac{n-p}{2}+a-1)}, \tag{4}$$

where

$$S_\phi^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}}_\phi)'\Sigma_\phi^{-1}(\mathbf{y} - X\hat{\boldsymbol{\beta}}_\phi),$$

and $\hat{\boldsymbol{\beta}}_\phi = (X'\Sigma_\phi^{-1}X)^{-1}X'\Sigma_\phi^{-1}\mathbf{y}$; $L^I(\phi; \mathbf{y})$ is called the *integrated likelihood* of ϕ. Then, the posterior distribution of $\boldsymbol{\eta}$ is proper if and only if

$$0 < \int_{\lambda_n^{-1}}^{\lambda_1^{-1}} L^I(\phi; \mathbf{y})\pi(\phi)d\phi < \infty, \tag{5}$$

so to determine propriety of posterior distributions based on priors (3) it is necessary to determine the behavior of both the integrated likelihood $L^I(\phi; \mathbf{y})$ and marginal prior $\pi(\phi)$ in the interval $(\lambda_n^{-1}, \lambda_1^{-1})$.

Some notation is now introduced. Let $\mathcal{C}(X)$ denote the subspace of $\mathbb{R}^n$ spanned by the columns of X, and $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be the normalized eigenvectors of W corresponding, respectively, to the eigenvalues $\lambda_1, \ldots, \lambda_n$, and recall that $\lambda_n < 0 < \lambda_1$. Throughout this and the next section $\phi \to \lambda_1^{-1}$ ($\phi \to \lambda_n^{-1}$) is used to denote that ϕ approaches λ_1^{-1} (λ_n^{-1}) from the left (right). Also, it is assumed throughout that $\{\lambda_i\}_{i=1}^n$ are not all equal.

Lemma 1 *Consider the CAR model (2) with $n \geq p + 2$, and suppose λ_1 and λ_n are simple eigenvalues. Then as $\phi \to \lambda_1^{-1}$ we have*

$$|X'\Sigma_\phi^{-1}X| = \begin{cases} O(1 - \phi\lambda_1) & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ O(1) & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases}, \tag{6}$$

and for every $\boldsymbol{\eta} \in \Omega$

$$S_\phi^2 = O(1) \quad \textit{with probability } 1. \tag{7}$$

The same results hold as $\phi \to \lambda_n^{-1}$ *when* λ_1 *and* $\mathbf{u}_1$ *are replaced by, respectively,* λ_n *and* $\mathbf{u}_n$*.*

Proof. See the Appendix

Proposition 1 *Consider the CAR model (2) and the prior distribution (3) with* $n \geq p+2$*, and suppose* λ_1 *and* λ_n *are simple eigenvalues. Then for every* $\boldsymbol{\eta} \in \Omega$ *the integrated likelihood* $L^I(\phi; \mathbf{y})$ *in (4) is with probability* 1 *a continuous function on* $(\lambda_n^{-1}, \lambda_1^{-1})$ *satisfying that as* $\phi \to \lambda_1^{-1}$

$$L^I(\phi; \mathbf{y}) = \begin{cases} O(1) & \textit{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ O((1-\phi\lambda_1)^{\frac{1}{2}}) & \textit{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases}.$$

The same result holds as $\phi \to \lambda_n^{-1}$ *when* λ_1 *and* $\mathbf{u}_1$ *are replaced by, respectively,* λ_n *and* $\mathbf{u}_n$*.*

Proof. The continuity of $L^I(\phi; \mathbf{y})$ on $(\lambda_n^{-1}, \lambda_1^{-1})$ follows from the definitions of $\Sigma_\phi^{-1}, S_\phi^2$ and the continuity of the determinant function. For any $\phi \in (0, \lambda_1^{-1})$ the eigenvalues of Σ_ϕ^{-1} are $1-\phi\lambda_1 < 1-\phi\lambda_2 \leq \ldots \leq 1-\phi\lambda_{n-1} < 1-\phi\lambda_n$, so

$$|\Sigma_\phi^{-1}| = (1-\phi\lambda_1)\prod_{i=2}^{n}(1-\phi\lambda_i),$$

and hence $|\Sigma_\phi^{-1}|^{\frac{1}{2}} = O((1-\phi\lambda_1)^{\frac{1}{2}})$ as $\phi \to \lambda_1^{-1}$. Then the result follows from (6) and (7). The proof on the behavior of $L^I(\phi; \mathbf{y})$ as $\phi \to \lambda_n^{-1}$ follows along the same lines.

Remark 1. Note that the limiting behaviors of $L^I(\phi; \mathbf{y})$ as $\phi \to \lambda_1^{-1}$ and $\phi \to \lambda_n^{-1}$ do not depend on the hyperparameter a.

Remark 2. The neighborhood systems used for modeling most datasets are such that there is a 'path' between any pair of sites. In this case the matrix W is *irreducible*, so λ_1 is guaranteed to be simple by the Perron-Frobenius theorem (Bapat and Raghavan, 1997 p. 17). For all the simulated and real datasets I have looked at λ_n was also simple, but this is not guaranteed to be so. For the case when each subregion is a neighbor of any other subregion, with $w_{ij} = 1$ for all $i \neq j$, it holds that $\lambda_n = -1$ has multiplicity $n-1$. But this kind of neighborhood system is rarely considered in practice.

3 Jeffreys Priors

Default or automatic priors are useful in situations where it is difficult to elicit a prior, either subjectively or from previous data. The most commonly used of such priors is the Jeffreys-rule

prior which is given by $\pi(\boldsymbol{\eta}) \propto (\det[I(\boldsymbol{\eta})])^{\frac{1}{2}}$, where $I(\boldsymbol{\eta})$ is the Fisher information matrix with (i,j) entry

$$[I(\boldsymbol{\eta})]_{ij} = E_\eta \left\{ \left(\frac{\partial}{\partial \eta_i} \log(L(\boldsymbol{\eta}; \mathbf{Y})) \right) \left(\frac{\partial}{\partial \eta_j} \log(L(\boldsymbol{\eta}; \mathbf{Y})) \right) \;\Big|\; \boldsymbol{\eta} \right\}.$$

The Jeffreys-rule prior has several attractive features, such as invariance to one-to-one reparametrizations and restrictions of the parameter space, but it also has some not so attractive features. One of these is the poor frequentist properties that have been noticed for some multiparameter models. This section derives two versions of Jeffreys prior, the Jeffreys-rule prior and the independence Jeffreys prior, where the latter (intended to ameliorate the aforementioned unattractive feature) is obtained by assuming that $\boldsymbol{\beta}$ and (σ^2, ϕ) are 'independent' a priori and computing each marginal prior using Jeffreys-rule when the other parameter is assumed known. Since these Jeffreys priors are improper (as is usually the case) the propriety of the resulting posteriors would need to be checked.

Theorem 1 *Consider the CAR model (2). Then the independence Jeffreys prior and the Jeffreys-rule prior of $\boldsymbol{\eta}$, to be denoted by $\pi^{J1}(\boldsymbol{\eta})$ and $\pi^{J2}(\boldsymbol{\eta})$, are of the form (3) with, respectively,*

$$a = 1 \qquad \text{and} \qquad \pi^{J1}(\phi) \propto \left\{ \sum_{i=1}^{n} \Big(\frac{\lambda_i}{1 - \phi\lambda_i} \Big)^2 - \frac{1}{n} \Big[\sum_{i=1}^{n} \frac{\lambda_i}{1 - \phi\lambda_i} \Big]^2 \right\}^{\frac{1}{2}}, \tag{8}$$

and

$$a = 1 + \frac{p}{2} \qquad \text{and} \qquad \pi^{J2}(\phi) \propto \Big(\prod_{j=1}^{p} (1 - \phi\nu_j) \Big)^{\frac{1}{2}} \pi^{J1}(\phi),$$

where $\nu_1 \geq \ldots \geq \nu_p$ are the ordered eigenvalues of $X_o' W X_o$, and X_o is the matrix defined by (18) in the Appendix.

Proof. From Theorem 5 in Berger et al. (2001) follows that for the spatial model $\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, \sigma^2\Sigma_\phi)$, the independence Jeffreys prior and Jeffreys-rule prior are both of the form (3) with, respectively,

$$a = 1 \qquad \text{and} \qquad \pi^{J1}(\phi) \propto \left\{ \mathrm{tr}[U_\phi^2] - \frac{1}{n} (\mathrm{tr}[U_\phi])^2 \right\}^{\frac{1}{2}},$$

and

$$a = 1 + \frac{p}{2} \qquad \text{and} \qquad \pi^{J2}(\phi) \propto |X' \Sigma_\phi^{-1} X|^{\frac{1}{2}} \pi^{J1}(\phi),$$

where $U_\phi = (\frac{\partial}{\partial \phi} \Sigma_\phi) \Sigma_\phi^{-1}$, and $\frac{\partial}{\partial \phi} \Sigma_\phi$ denotes the matrix obtained by differentiating Σ_ϕ element by element. For the CAR model $\Sigma_\phi^{-1} = I_n - \phi W$, so

$$U_\phi = -\Sigma_\phi \Big(\frac{\partial}{\partial \phi} \Sigma_\phi^{-1} \Big) = (I_n - \phi W)^{-1} W.$$

Noting now that $\{\frac{\lambda_i}{1-\phi\lambda_i}\}_{i=1}^n$ are the eigenvalues of U_ϕ, it follows that

$$\mathrm{tr}[U_\phi^2] - \frac{1}{n}\Big(\mathrm{tr}[U_\phi]\Big)^2 = \sum_{i=1}^n \Big(\frac{\lambda_i}{1-\phi\lambda_i}\Big)^2 - \frac{1}{n}\Big[\sum_{i=1}^n \frac{\lambda_i}{1-\phi\lambda_i}\Big]^2,$$

so the first result follows. The second result follows from the first and identity (19) in the Appendix.

Lemma 2 *Suppose λ_1 and λ_n are simple eigenvalues. Then as $\phi \to \lambda_1^{-1}$ it holds that*

$$\pi^{J1}(\phi) = O((1-\phi\lambda_1)^{-1}),$$

and

$$\pi^{J2}(\phi) = \begin{cases} O((1-\phi\lambda_1)^{-\frac{1}{2}}) & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ O((1-\phi\lambda_1)^{-1}) & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases}.$$

The same results hold as $\phi \to \lambda_n^{-1}$ when λ_1 and $\mathbf{u}_1$ are replaced by, respectively, λ_n and $\mathbf{u}_n$.

Proof. From (8) and after some algebraic manipulation follow that

$$\begin{aligned}
\big(\pi^{J1}(\phi)\big)^2 &\propto \Big(\frac{\lambda_1}{1-\phi\lambda_1}\Big)^2 + \sum_{i=2}^n \Big(\frac{\lambda_i}{1-\phi\lambda_i}\Big)^2 - \frac{1}{n}\Big[\frac{\lambda_1}{1-\phi\lambda_1} + \sum_{i=2}^n \frac{\lambda_i}{1-\phi\lambda_i}\Big]^2 \\
&= \Big(\frac{\lambda_1}{1-\phi\lambda_1}\Big)^2\Big(1 - \frac{1}{n} - \frac{2(1-\phi\lambda_1)}{n\lambda_1}\sum_{i=2}^n \frac{\lambda_i}{1-\phi\lambda_i} \\
&\qquad + \Big(\frac{1-\phi\lambda_1}{\lambda_1}\Big)^2\Big(\sum_{i=2}^n \Big(\frac{\lambda_i}{1-\phi\lambda_i}\Big)^2 - \frac{1}{n}\Big[\sum_{i=2}^n \frac{\lambda_i}{1-\phi\lambda_i}\Big]^2\Big)\Big) \\
&= O((1-\phi\lambda_1)^{-2}) \quad \text{as} \quad \phi \to \lambda_1^{-1},
\end{aligned}$$

since $\lambda_1 > \lambda_i$ for $i = 2, \ldots, n$. The behavior as $\phi \to \lambda_n^{-1}$ is established in the same way, and the second result follows from the first and (6).

Corollary 1 *Consider the CAR model (2) and let $k \in \mathbb{N}$. Then:*
(i) The marginal independence Jeffreys prior $\pi^{J1}(\phi)$ is unbounded and not integrable.
(ii) The joint independence Jeffreys posterior $\pi^{J1}(\boldsymbol{\eta} \mid \mathbf{y})$ is proper when neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ are in $\mathcal{C}(X)$, while it is improper when either $\mathbf{u}_1$ or $\mathbf{u}_n$ are in $\mathcal{C}(X)$.
(iii) The marginal independence Jeffreys posterior $\pi^{J1}(\phi \mid \mathbf{y})$ has moments of any order k when neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ are in $\mathcal{C}(X)$.
(iv) The marginal independence Jeffreys posterior $\pi^{J1}(\sigma^2 \mid \mathbf{y})$ has a finite moment of order k if $n \geq p + 2k + 1$.

Proof. (i) From Lemma 2 follows that for $i = 1$ or n, $\lim_{\phi \to \lambda_i^{-1}} \pi^{J1}(\phi) = \infty$. Also

$$\int_0^{\lambda_1^{-1}} \pi^{J1}(\phi) d\phi \propto \int_0^{\lambda_1^{-1}} (1 - \phi\lambda_1)^{-1} h(\phi) d\phi = \lambda_1^{-1} \int_0^1 t^{-1} \tilde{h}(t) dt,$$

where the last identity is obtained by the change of variable $t = 1 - \phi\lambda_1$, $h(\phi)$ is an (unspecified) function that is continuous on $(0, \lambda_1^{-1})$ and $O(1)$ as $\phi \to \lambda_1^{-1}$, and $\tilde{h}(t)$ is an (unspecified) function that is continuous on $(0, 1)$ and $O(1)$ as $t \to 0$; a similar identity holds for $\int_{\lambda_n^{-1}}^0 \pi^{J1}(\phi) d\phi$. The result now follows since t^{-1} is not integrable around 0.

(ii) From Proposition 1 and Lemma 2, and noting that $\pi^{J1}(\phi \mid \mathbf{y}) \propto L^I(\phi; \mathbf{y}) \pi^{J1}(\phi)$

$$\begin{aligned} \int_0^{\lambda_1^{-1}} \pi^{J1}(\phi \mid \mathbf{y}) d\phi \quad &\propto \quad \begin{cases} \int_0^{\lambda_1^{-1}} (1 - \phi\lambda_1)^{-1} h(\phi) d\phi & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ \int_0^{\lambda_1^{-1}} (1 - \phi\lambda_1)^{-\frac{1}{2}} h(\phi) d\phi & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases} \\ &= \quad \begin{cases} \lambda_1^{-1} \int_0^1 t^{-1} \tilde{h}(t) dt & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ \lambda_1^{-1} \int_0^1 t^{-\frac{1}{2}} \tilde{h}(t) dt & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases}, \end{aligned} \tag{9}$$

where $h(\phi)$ and $\tilde{h}(t)$ are as in (i); a similar identity holds for $\int_{\lambda_n^{-1}}^0 \pi^{J1}(\phi \mid \mathbf{y}) d\phi$ with λ_1 and $\mathbf{u}_1$ replaced by, respectively, λ_n and $\mathbf{u}_n$. The result then follows by the same argument as in (i).

(iii) By a similar calculation as in (i) and (ii) and using the binomial expansion of $(1 - t)^k$

$$\int_0^{\lambda_1^{-1}} \phi^k \pi^{J1}(\phi \mid \mathbf{y}) d\phi \propto \begin{cases} \lambda_1^{-(k+1)} \int_0^1 \left(t^{-1} + \sum_{j=1}^{k-1} (-1)^{k-j} \binom{k}{j} t^{k-1-j} \right) \tilde{h}(t) dt & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ \lambda_1^{-(k+1)} \int_0^1 \left(t^{-\frac{1}{2}} + \sum_{j=1}^{k-1} (-1)^{k-j} \binom{k}{j} t^{k-\frac{1}{2}-j} \right) \tilde{h}(t) dt & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases},$$

where $\tilde{h}(t)$ is as in (i); a similar identity holds for $\int_{\lambda_n^{-1}}^0 \phi^k \pi^{J1}(\phi \mid \mathbf{y}) d\phi$ with λ_1 and $\mathbf{u}_1$ replaced by, respectively, λ_n and $\mathbf{u}_n$. The result follows since t^{-1} is not integrable on $(0, 1)$, while $t^{-\frac{1}{2}}$, t^{k-1+j} and $t^{k-\frac{1}{2}+j}$ are, $j = 0, 1, \ldots, k-1$.

(iv) Note that $E\{(\sigma^2)^k \mid \mathbf{y}\}$ exists if $E\{(\sigma^2)^k \mid \phi, \mathbf{y}\}$ exists and is integrable with respect to $\pi^{J1}(\phi \mid \mathbf{y})$. A standard calculation shows that for any prior in (3) with $a = 1$, $\pi^{J1}(\sigma^2 \mid \phi, \mathbf{y}) = \mathrm{IG}\left(\frac{n-p}{2}, \frac{1}{2} S_\phi^2\right)$, where $\mathrm{IG}(a, b)$ denotes the inverse gamma distribution with mean $b/(a-1)$. Again by direct calculation, $E\{(\sigma^2)^k \mid \phi, \mathbf{y}\} = c(S_\phi^2)^k < \infty$, provided $n \geq p + 2k + 1$, where $c > 0$ does not depend on $(\phi, \mathbf{y})$. The result then follows from (7), (9) and the relation analogous to (9) for $\int_{\lambda_n^{-1}}^0 \pi^{J1}(\phi \mid \mathbf{y}) d\phi$.

Corollary 2 *Consider the CAR model (2) and let $k \in \mathbb{N}$. Then:*
(i) The marginal Jeffreys-rule prior $\pi^{J2}(\phi)$ is unbounded. Also, it is integrable when both $\mathbf{u}_1$ and $\mathbf{u}_n$ are in $\mathcal{C}(X)$, while it is not integrable when either $\mathbf{u}_1$ or $\mathbf{u}_n$ is not in $\mathcal{C}(X)$.
(ii) The joint Jeffreys-rule posterior $\pi^{J2}(\boldsymbol{\eta} \mid \mathbf{y})$ is always proper.

(iii) The marginal Jeffreys-rule posterior $\pi^{J2}(\phi \mid \mathbf{y})$ has always moments of any order k.
(iv) The marginal Jeffreys posterior $\pi^{J2}(\sigma^2 \mid \mathbf{y})$ has a finite moment of order k if $n \geq p+2k+1$.

Proof. These results are proved similarly as their counterparts in Corollary 1.

Establishing some of the properties of posterior distributions based on the Jeffreys priors requires numerical computation of $\mathbf{u}_1$ and $\mathbf{u}_n$, and determining whether or not these eigenvectors belong to $\mathcal{C}(X)$. The latter can be done by computing the rank of matrices $A_i = (X \vdots \mathbf{u}_i)$, since $\mathbf{u}_i \notin \mathcal{C}(X)$ if and only if $\text{rank}(A_i) = p+1$, $i = 1, n$ (recall that X has full rank). This rank can be computed from the QR decomposition of A_i (Schott, 2005)

$$A_i = Q_i \begin{pmatrix} R_i \\ \mathbf{0} \end{pmatrix},$$

where Q_i is an $n \times n$ orthogonal matrix, $Q_i'Q_i = Q_iQ_i' = I_n$, and R_i is a $(p+1) \times (p+1)$ upper triangular matrix with non-negative diagonal elements. Then

$$\text{rank}(A_i) = \text{rank}(R_i) = \#\text{ of non-zero diagonal elements in } R_i.$$

Remark 3. The independence Jeffreys prior yields a proper posterior when neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ are in $\mathcal{C}(X)$. For all the simulated and real datasets I have looked at, neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ were in $\mathcal{C}(X)$, and it seems unlikely to encounter in practice a situation where either $\mathbf{u}_1$ or $\mathbf{u}_n$ are in $\mathcal{C}(X)$. Nevertheless, posterior impropriety is a potential problem when the independence Jeffreys prior is used. On the other hand, the Jeffreys-rule prior always yields a proper posterior but, as will be seen later, frequentist properties of Bayesian inferences based on Jeffreys-rule priors are somewhat inferior to those based on independence Jeffreys priors.

Remark 4. Another commonly used default prior is the reference prior proposed by Bernardo (1979) and Berger and Bernardo (1992). It can be shown from a result in Berger et al. (2001) that a reference prior for the parameters of model (2) is also of the form (3), with

$$a = 1 \qquad \text{and} \qquad \pi^R(\phi) \propto \left\{ \sum_{i=1}^{n-p} \upsilon_i^2(\phi) - \frac{1}{n-p} \Big[\sum_{i=1}^{n-p} \upsilon_i(\phi) \Big]^2 \right\}^{\frac{1}{2}}$$

where $\upsilon_1(\phi), \ldots, \upsilon_{n-p}(\phi)$ are the nonzero eigenvalues of the matrix $V_\phi = (\frac{\partial}{\partial \phi}\Sigma_\phi)\Sigma_\phi^{-1}P_\phi^W$, with $P_\phi^W = I_n - X(X'\Sigma_\phi^{-1}X)^{-1}X'\Sigma_\phi^{-1}$. It was shown in Berger et al. (2001) that for some geostatistical models inferences based on this prior have similar or better properties than those based on the independence Jeffreys prior. Unfortunately, it does not seem feasible to find an explicit expression for the above eigenvalues, and properties of Bayesian inferences based on this prior remain unknown.

4 Inference and Comparison

4.1 Inference

Posterior inference about the unknown quantities would be based on a sample from their posterior distribution. When the observed data are complete, a sample from the posterior distribution of the model parameters is simulated using a noniterative Monte Carlo algorithm based on the factorization

$$\pi(\boldsymbol{\beta}, \sigma^2, \phi \mid \mathbf{y}) = \pi(\boldsymbol{\beta} \mid \sigma^2, \phi, \mathbf{y})\pi(\sigma^2 \mid \phi, \mathbf{y})\pi(\phi \mid \mathbf{y}),$$

where from (2) and (3)

$$\begin{aligned}
\pi(\boldsymbol{\beta} \mid \sigma^2, \phi, \mathbf{y}) &= \mathrm{N}_p\big(\hat{\boldsymbol{\beta}}_\phi, \sigma^2(X'\Sigma_\phi^{-1}X)^{-1}\big), && (10)\\
\pi(\sigma^2 \mid \phi, \mathbf{y}) &= \mathrm{IG}\Big(\frac{n-p}{2} + a - 1, \frac{1}{2}S_\phi^2\Big), && (11)\\
\pi(\phi \mid \mathbf{y}) &\propto \big(|\Sigma_\phi||X'\Sigma_\phi^{-1}X|\big)^{-\frac{1}{2}}\big(S_\phi^2\big)^{-(\frac{n-p}{2}+a-1)}\pi(\phi). && (12)
\end{aligned}$$

Simulation from (10) and (11) is straightforward, while simulation from (12) would be accomplished using the adaptive rejection Metropolis sampling (ARMS) algorithm proposed by Gilks, Best and Tan (1995). The ARMS algorithm requires no tuning and works very well for this model. It was found that produces well mixed chains with very low autocorrelations, so long runs are not required for precise inference.

The above algorithm cannot be easily modified to handle the case when there are missing values at some of the sites. In this case, sampling from the posterior of the model parameters and missing values is done using a Gibbs sampling algorithm. Let $\mathbf{y} = (\mathbf{y}_J, \mathbf{y}_I)$ be the "complete data", where J and I are the sites that correspond to, respectively, the observed and missing values, and $\mathbf{y}_J = \{y_j : j \in J\}$ are the observed values. Again from (2) and (3) follow that the full conditional distribution of $\boldsymbol{\beta}$ is given by (10), while the full conditionals of the other components of $(\mathbf{y}_I, \boldsymbol{\beta}, \sigma^2, \phi)$ are given by

$$\begin{aligned}
\pi(y_i \mid \mathbf{y}_{(i)}, \boldsymbol{\beta}, \sigma^2, \phi) &= \mathrm{N}\Big(\mathbf{x}_i'\boldsymbol{\beta} + \phi\sum_{j=1}^{n} w_{ij}(y_j - \mathbf{x}_j'\boldsymbol{\beta}), \sigma^2\Big)\ , \quad i \in I, && (13)\\
\pi(\sigma^2 \mid \boldsymbol{\beta}, \phi, \mathbf{y}) &= \mathrm{IG}\Big(\frac{n}{2} + a - 1, \frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})'\Sigma_\phi^{-1}(\mathbf{y} - X\boldsymbol{\beta})\Big), && (14)\\
\pi(\phi \mid \boldsymbol{\beta}, \sigma^2, \mathbf{y}) &\propto |\Sigma_\phi^{-1}|^{\frac{1}{2}} \exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})'\Sigma_\phi^{-1}(\mathbf{y} - X\boldsymbol{\beta})\Big\}\pi(\phi). && (15)
\end{aligned}$$

Simulation from (13) and (14) is straightforward, while simulation from (15) would be accomplished using the ARMS algorithm. This Gibbs sampling algorithm also works very well for this model, as it was found that produces well mixed chains with low autocorrelations. Finally,

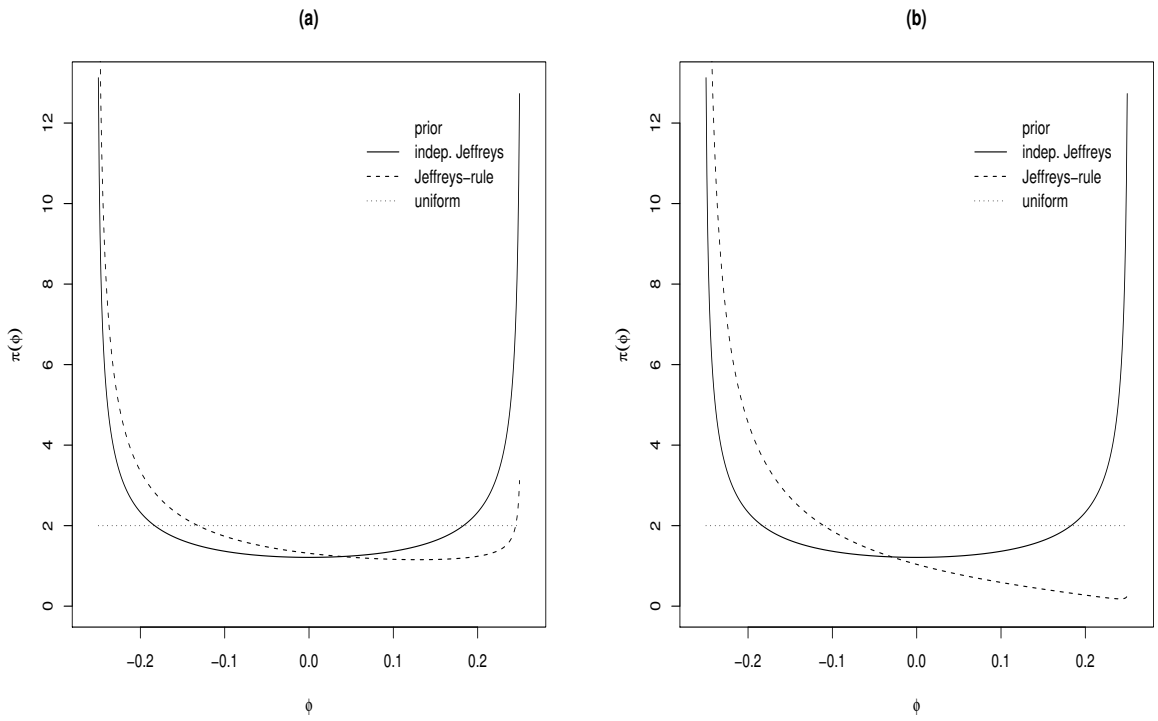

Figure 1: Marginal independence Jeffreys, Jeffreys-rule and uniform priors of ϕ for models over a 20 by 20 regular lattice with a first order (or 'rook') neighborhood system when the mean is (a) constant, and (b) a degree one polynomial in the site coordinates.

based on initial estimates of the parameters and missing values, sucessive sampling is done from (10), (13), (14) and (15).

4.2 Comparison

This section presents comparisons between the two versions of Jeffreys prior and the uniform prior, as well as their corresponding posteriors distributions. For this I consider models defined on a 20 by 20 regular lattice with a first order (or 'rook') neighborhood system (the neighbors of a site are the sites adjacent to the north, south, east and west), with $w_{ij} = 1$ if sites i and j are neighbors, and $w_{ij} = 0$ otherwise; the resulting W matrix is often called the *adjacency matrix*. In this case ϕ must belong to the interval $(-0.252823, 0.252823)$.

Figure 1 displays the independence Jeffreys, Jeffreys-rule and uniform priors of ϕ for models where $E\{Y_i\}$ is (a) constant and (b) a degree one polynomial in the site coordinates. For both models neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ are in $\mathcal{C}(X)$. The graphs of $\pi^{J1}(\phi)$ and $\pi^{J2}(\phi)$ are both 'bathtub-

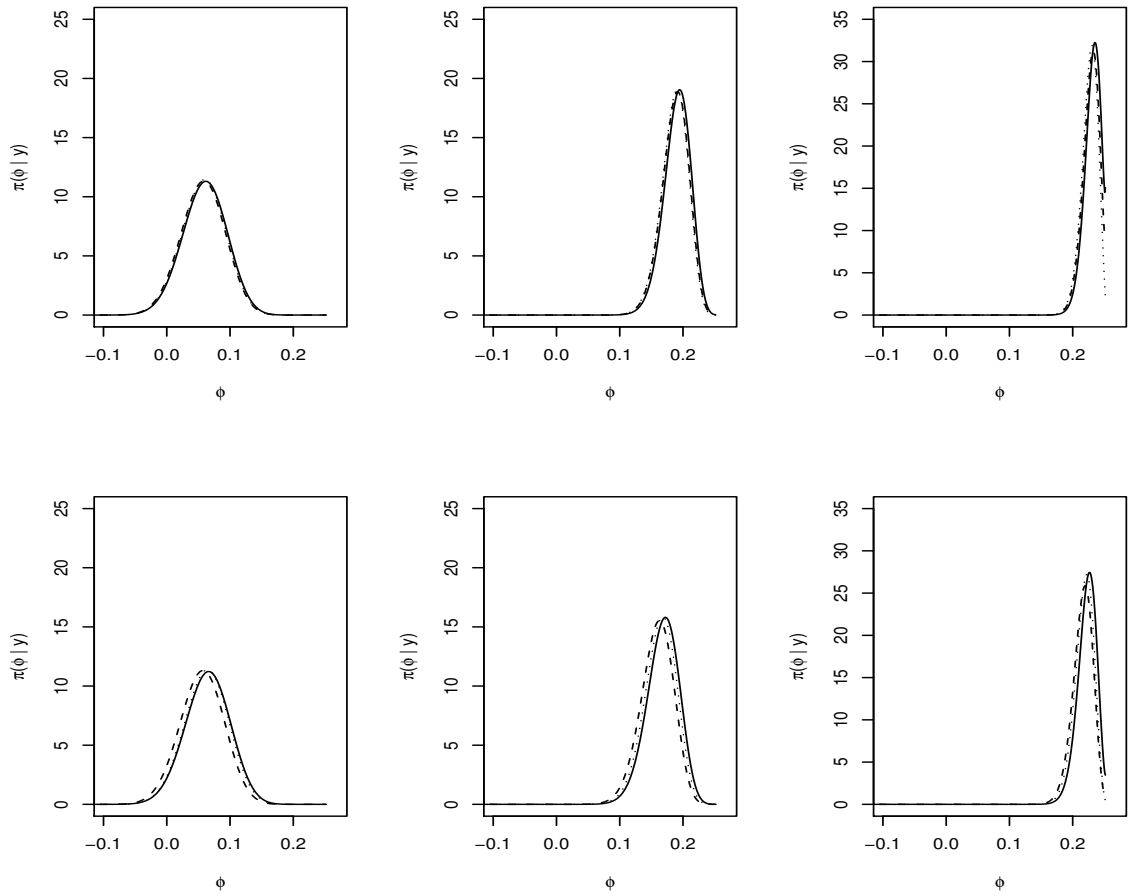

Figure 2: Marginal independence Jeffreys (solid), Jeffreys-rule (dashed) and uniform (doted) posteriors of ϕ based on simulated *complete* data for models over a 20 by 20 regular lattice with a first order neighborhood system. $E\{Y_i\}$ is either 10 (top panels) or $10 + s_{i1} + s_{i2}$ (bottom panels), $\sigma^2 = 2$ and ϕ is 0.1 (left panels), 0.2 (middle panels) or 0.24 (right panels).

shaped', in great contrast with the graph of $\pi^U(\phi)$. In particular $\pi^{J1}(\phi)$ assigns substantial mass to values of ϕ close to the boundaries, while $\pi^{J2}(\phi)$ does the same for values of ϕ close to the left boundary but not so much for values of ϕ close to the right boundary, specially when the mean is not constant. The reason for this asymmetry is unclear, but it makes the Jeffreys-rule prior somewhat unappealing since most fits of the CAR model to datasets reported in the literature have yielded large positive estimates for ϕ.

Figure 2 displays the independence Jeffreys (solid), Jeffreys-rule (dashed) and uniform (dotted) posteriors of ϕ based on simulated *complete* data. The mean of the observations is

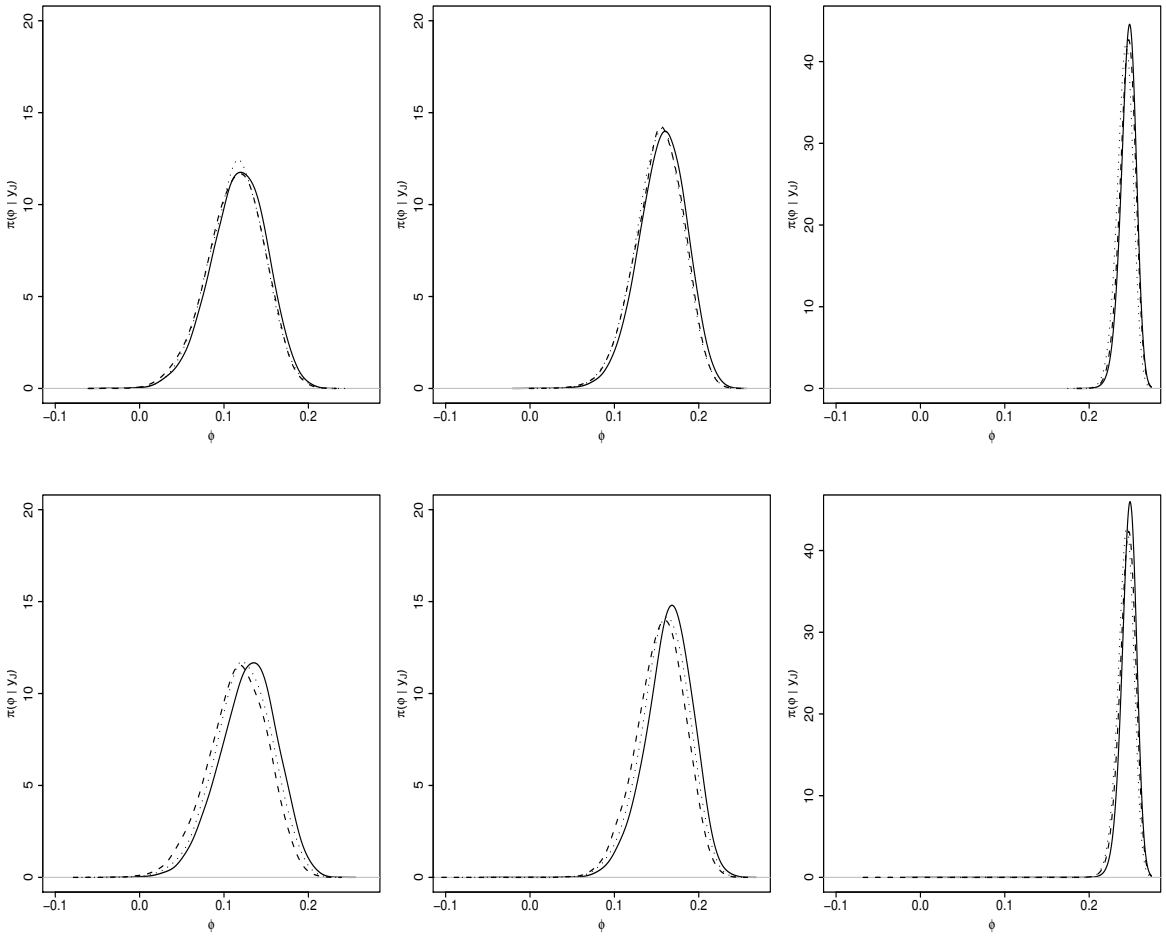

Figure 3: Marginal independence Jeffreys (solid), Jeffreys-rule (dashed) and uniform (doted) posteriors of ϕ based on simulated data with *missing* values for models over a 20×20 regular lattice with a first order neighborhood system. $E\{Y_i\}$ is either 10 (top panels) or $10 + s_{i1} + s_{i2}$ (bottom panels), $\sigma^2 = 2$ and ϕ is 0.1 (left panels), 0.2 (middle panels) or 0.24 (right panels).

either 10 (top panels) or $10 + s_{i1} + s_{i2}$ (bottom panels), with (s_{i1}, s_{i2}) the coordinates of site i, $\sigma^2 = 2$ and ϕ is 0.1 (left panels), 0.2 (middle panels) or 0.24 (right panels). The three default posteriors are usually close to each other when the mean is constant. When the mean is not constant the three default posteriors of ϕ differ somewhat, with $\pi^{J1}(\phi \mid \mathbf{y})$ being shifted to the right and less disperse when compared to $\pi^{J2}(\phi \mid \mathbf{y})$, and $\pi^{U}(\phi \mid \mathbf{y})$ located and shaped somewhere 'between' the other two. Also, as ϕ gets large the posterior distributions become more concentrated around the true value, which is consistent with the asymptotic result in (16). The same patters were observed for several other simulated datasets (not shown).

Figure 3 displays the independence Jeffreys (solid), Jeffreys-rule (dashed) and uniform (dotted) posteriors of ϕ based on simulated data with *missing* values (obtained by smoothing the histogram of the posteriors). The data were simulated from the same models as in Figure 2 and later the observations from 40 sites (10%) selected at random were removed. The main properties and comparative behavior of these posteriors are the same as for the case of complete data.

5 Further Properties

5.1 Frequentist Properties

This section presents results of a simulation experiment to study some of the frequentist properties of Bayesian inferences based on the independence Jeffreys, Jeffreys-rule and uniform priors, as well as those based on maximum likelihood (ML). These properties are often proposed as a way to evaluate and compare default priors. The focus of interest is on the covariance parameters, and the frequentist properties to be considered are frequentist coverage of credible and confidence intervals, and mean absolute error of estimators. For the Bayesian procedures I use the 95% equal-tailed credible intervals for σ^2 and ϕ, and the posterior medians as their estimators, all of which are readily obtained from the Monte Carlo output. For the ML procedure I use the large sample (approximate) 95% confidence intervals given by $\hat{\sigma}^2 \pm 1.96(\hat{\text{avar}}(\hat{\sigma}^2))^{1/2}$ and $\hat{\phi} \pm 1.96(\hat{\text{avar}}(\hat{\phi}))^{1/2}$, where $\hat{\sigma}^2$ and $\hat{\phi}$ are the ML estimators of σ^2 and ϕ, avar$(\cdot)$ denotes asymptotic variance and $\hat{\text{avar}}(\cdot)$ indicates that avar$(\cdot)$ is evaluated at the ML estimators. Using the result on the asymptotic distribution of ML estimators in Mardia and Marshall (1984) and after some algebra, it follows that the above asymptotic variances are given by

$$\text{avar}(\hat{\sigma}^2) = \frac{2\sigma^4}{n(g(\phi))^2} \sum_{i=1}^{n} \Big(\frac{\lambda_i}{1-\phi\lambda_i} \Big)^2 \qquad \text{and} \qquad \text{avar}(\hat{\phi}) = \frac{2}{(g(\phi))^2}, \tag{16}$$

where $g(\phi)$ is given by the right hand side of (8).

I consider models defined on a 10 by 10 regular lattice with first order neighborhood system, and W the adjacency matrix. Then ϕ must belong to the interval $(-0.260554, 0.260554)$. The factors to be varied in the experiment are $E\{Y_i\}$, σ^2 and ϕ. I consider $E\{Y_i\}$ equal to 10 ($p = 1$) or $10+s_{i1}+s_{i2}$ ($p = 3$), σ^2 equal to 0.1 or 2, and ϕ equal to $0.05, 0.12$ or 0.25 (negative estimates of the spatial parameter are rare in practice, if they appear at all, so only positive values of ϕ are considered). This setup provides a range of different scenarios in terms of trend, variability and spatial association. For each of the 12 ($2 \times 2 \times 3$) possible scenarios, 1500 datasets were simulated and for each dataset a posterior sample of the model parameters of size $m = 3000$ was generated by the algorithm described in Section 4.

Table 1: Frequentist coverage of Bayesian equal-tailed 95% credible intervals, and large sample 95% confidence interval for σ^2.

	$p=1$			$p=3$		
ϕ	0.05	0.12	0.25	0.05	0.12	0.25
			$\sigma^2=0.1$			
Ind. Jeffreys	.957	.950	.950	.948	.939	.950
Jeffreys-rule	.958	.946	.951	.933	.932	.950
Uniform	.961	.956	.944	.949	.940	.943
ML	.934	.935	.946	.940	.896	.932
			$\sigma^2=2$			
Ind. Jeffreys	.937	.947	.948	.944	.942	.955
Jeffreys-rule	.934	.948	.950	.936	.936	.953
Uniform	.938	.950	.944	.950	.947	.943
ML	.912	.932	.943	.898	.897	.940

Table 1 shows (empirical) frequentist coverage of Bayesian equal-tailed 95% credible intervals for σ^2 corresponding to three default priors, and large sample 95% confidence intervals for σ^2. The coverage of the three credible intervals are similar to each other and reasonably

Table 2: Frequentist coverage of Bayesian equal-tailed 95% credible intervals, and large sample 95% confidence interval for ϕ.

	$p=1$			$p=3$		
ϕ	0.05	0.12	0.25	0.05	0.12	0.25
			$\sigma^2=0.1$			
Ind. Jeffreys	.952	.952	.968	.948	.954	.978
Jeffreys-rule	.952	.947	.955	.939	.929	.931
Uniform	.965	.964	.878	.966	.969	.884
ML	.935	.932	.917	.918	.920	.906
			$\sigma^2=2$			
Ind. Jeffreys	.948	.948	.974	.949	.947	.968
Jeffreys-rule	.950	.949	.962	.926	.925	.922
Uniform	.961	.960	.893	.963	.958	.886
ML	.932	.927	.920	.906	.908	.893

Table 3: Mean absolute error $\times 10^2$ of the posterior median and ML estimate of σ^2.

	$p=1$			$p=3$		
ϕ	0.05	0.12	0.25	0.05	0.12	0.25
			$\sigma^2 = 0.1$			
Ind. Jeffreys	1.130	1.136	1.140	1.144	1.183	1.172
Jeffreys-rule	1.132	1.132	1.140	1.156	1.187	1.173
Uniform	1.126	1.134	1.217	1.140	1.177	1.237
ML	1.134	1.130	1.140	1.164	1.194	1.176
			$\sigma^2 = 2$			
Ind. Jeffreys	23.737	23.292	23.282	23.043	23.636	23.524
Jeffreys-rule	23.678	23.160	23.310	23.404	23.520	23.461
Uniform	23.666	23.166	24.953	22.950	23.462	25.093
ML	23.702	23.147	23.234	23.610	23.602	23.570

close to the nominal .95, while the coverage of the ML confidence intervals are slightly below nominal.

Table 2 shows (empirical) frequentist coverage of Bayesian equal-tailed 95% credible intervals for ϕ corresponding to the three default priors, and large sample 95% confidence intervals for ϕ. The coverage of the ML confidence intervals are below nominal, while the coverage of

Table 4: Mean absolute error $\times 10^2$ of the posterior median and ML estimate of ϕ.

	$p=1$			$p=3$		
ϕ	0.05	0.12	0.25	0.05	0.12	0.25
			$\sigma^2 = 0.1$			
Ind. Jeffreys	5.750	5.410	1.486	6.012	5.756	1.525
Jeffreys-rule	5.632	5.421	1.906	6.134	6.361	2.850
Uniform	5.090	4.856	2.562	5.205	5.120	2.930
ML	5.473	5.277	2.263	5.991	6.310	9.713
			$\sigma^2 = 2$			
Ind. Jeffreys	5.761	5.307	1.375	6.125	5.872	1.604
Jeffreys-rule	5.727	5.372	1.760	6.268	6.232	2.961
Uniform	5.127	4.865	2.415	5.316	5.110	2.965
ML	5.570	5.246	2.110	6.127	6.144	3.462

the three credible intervals are similar to each other and reasonably close to the nominal .95 under most scenarios, except when ϕ is large. In this case the coverage of credible intervals based on the uniform prior are well below nominal.

Tables 3 shows (empirical) mean absolute error (MAE) of the posterior median of σ^2 corresponding to the three default priors and the MAE of the ML estimator of σ^2. The MAEs of the Bayesian estimators based on the three default priors and the ML estimator are close to each other under all scenarios.

Tables 4 shows (empirical) MAE of the posterior median of ϕ corresponding to the three default priors and the ML estimator of ϕ. For small or moderate values of ϕ the MAEs of the three Bayesian estimators and the ML estimator are close to each other, with the MAE of the Bayesian estimator based on the uniform prior being slightly smaller than the other three. On the other hand, for large values of ϕ the MAE of the Bayesian estimator based on the independence Jeffreys prior is substantially smaller than the MAEs of the other three estimators. Also, when the the mean of the observations is not constant the MAE of the estimator based on the independence Jeffreys prior is smaller than the MAE of the estimator based on the Jeffreys-rule prior.

In summary, frequentist properties of ML estimators are inferior than those of Bayesian inferences based on any of the three default priors. More notably, Bayesian inferences based on the three default priors are reasonably good and similar to each other under most scenarios, except when the spatial parameter ϕ is large or the mean of the observations is not constant. In these cases frequentist properties of Bayesian inferences based on the independence Jeffreys prior no worse or better than those based on any of the other two default priors in regard to inference about ϕ.

5.2 Sensitivity to Design

The proposed Jeffreys priors depend on several features of the selected design, such as sample size and regression matrix. This section explores how sensitive these default priors are to the above features.

Sample Size. Consider the models defined over 10 by 10, 20 by 20 and 50 by 50 regular lattices with first order neighborhood system and W the adjacency matrix. Figure 4(a) displays the marginal independence Jeffreys priors $\pi^{J1}(\phi)$ corresponding to the three sample sizes, showing that they are very close to each other. It should be noted that the domains of $\pi^{J1}(\phi)$ for the above three models are not exactly the same, but are quite close. The priors were plotted over the interval $(-0.25, 0.25)$, the limit of $(\lambda_n^{-1}, \lambda_1^{-1})$ as $n \to \infty$. The same lack of sensitivity to

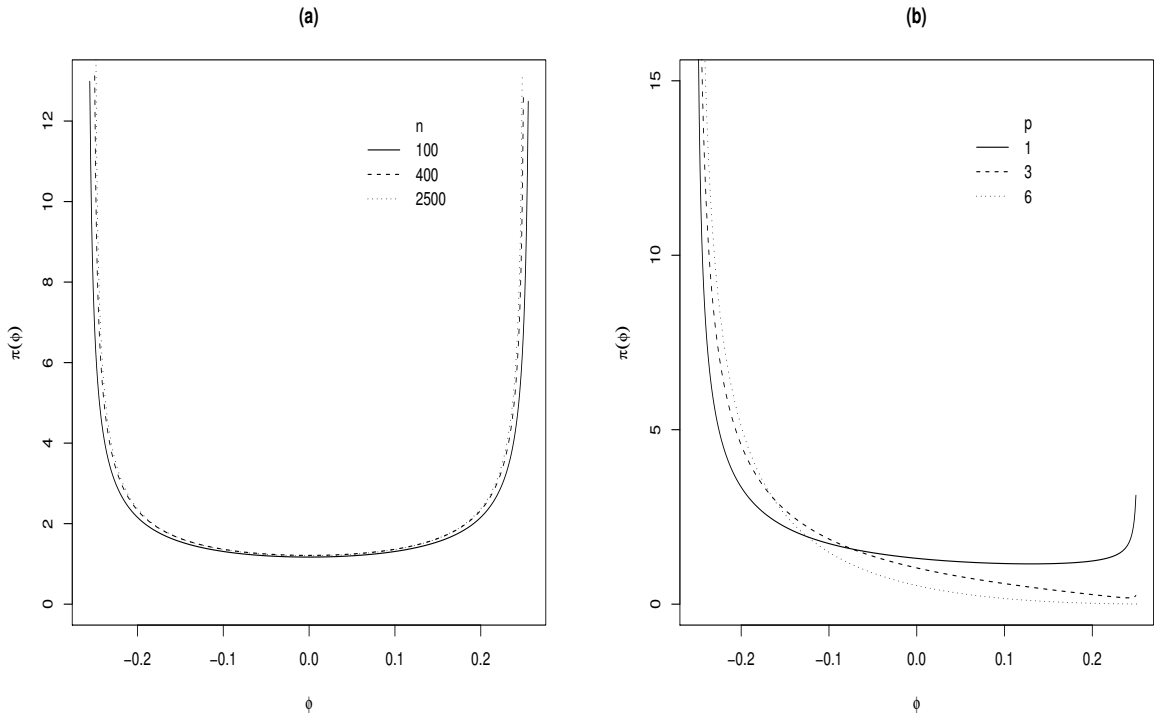

Figure 4: (a) Marginal independence Jeffreys priors of ϕ defined for models over 10 by 10, 20 by 20 and 50 by 50 regular lattices with first order neighborhood system. (b) Marginal Jeffreys-rule priors of ϕ for models defined over a 20 by 20 regular lattices with the same neighborhood system as in (a), and with mean constant, a degree one polynomial and a degree two polynomial in the site coordinates.

sample size was displayed by $\pi^{J2}(\phi)$, provided the models have the same mean structure (not shown).

Regression Matrix. Consider the models defined over a 20 by 20 regular lattices with the same neighborhood system as in the previous comparison, and mean a constant ($p = 1$), a degree one polynomial in the site coordinates ($p = 3$), and a degree two polynomial in the site coordinates ($p = 6$). By construction the marginal prior $\pi^{J1}(\phi)$ does not depend on the mean structure of the model. Figure 4(b) displays the marginal Jeffreys-rule priors $\pi^{J2}(\phi)$ corresponding to the three models, showing that these do depend substantially on the mean structure.

It could also be considered studying the sensitivity of the proposed default priors to other features of the design, such as neighborhood system or type of lattice, but these may not be sensible for CAR models since the parameter space depends substantially on these features. For

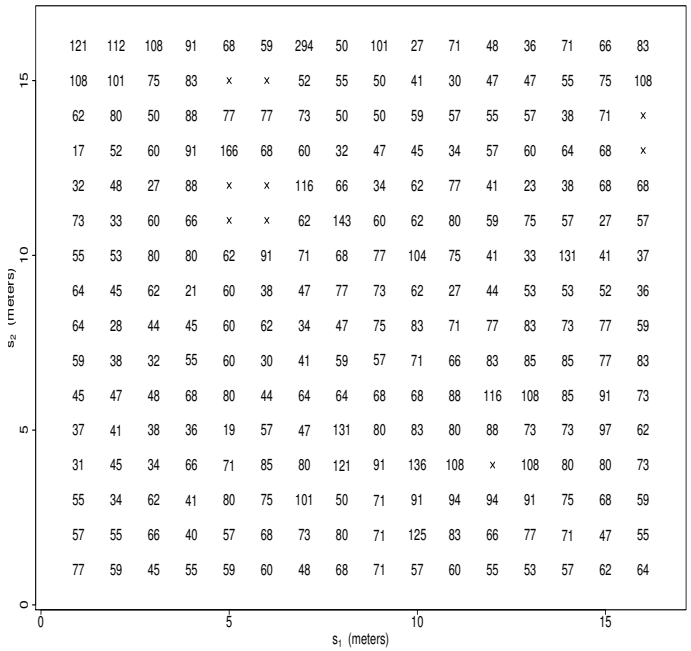

Figure 5: Phosphate concentration readings (in mg P/100 gr of soil) measured over a 16 by 16 regular lattice. Locations where values are missing are indicated with an '×'.

a 20 by 20 regular lattice, a valid CAR model requires that ϕ belongs to $(-0.252823, 0.252823)$ when the lattice is endowed with a first order neighborhood system, while ϕ must belong to $(-0.255679, 0.127121)$ when the lattice is endowed with a 'queen' neighborhood system (first order neighbors plus their adjacent sites to the northeast, northwest, southeast and southwest). Similarly, for a 10 by 10 regular lattice with first order neighborhood system ϕ must belong to $(-0.260554, 0.260554)$, while for the irregular lattice with first order neighborhood system formed by the 100 counties of the state of North Carolina in the United States ϕ must belong to $(-0.327373, 0.189774)$.

6 Example

To illustrate the Bayesian analysis of CAR models, I use a spatial dataset initially analized by Buck, Cavanagh and Litton (1988), and more recently reanalyzed by Cressie and Kapat (2008) [for now on refered to as CK]. The dataset consists of raw phosphate concentration readings (in mg P/100 g of soil) collected over several years in an archeological region of Laconia across the Evrotas river in Greece. The original observations were collected 10 meters apart over a 16 by 16 regular latice; they are denoted by $\{D_i \ : \ i = 1, \ldots, 256\}$ and displayed in Figure 5.

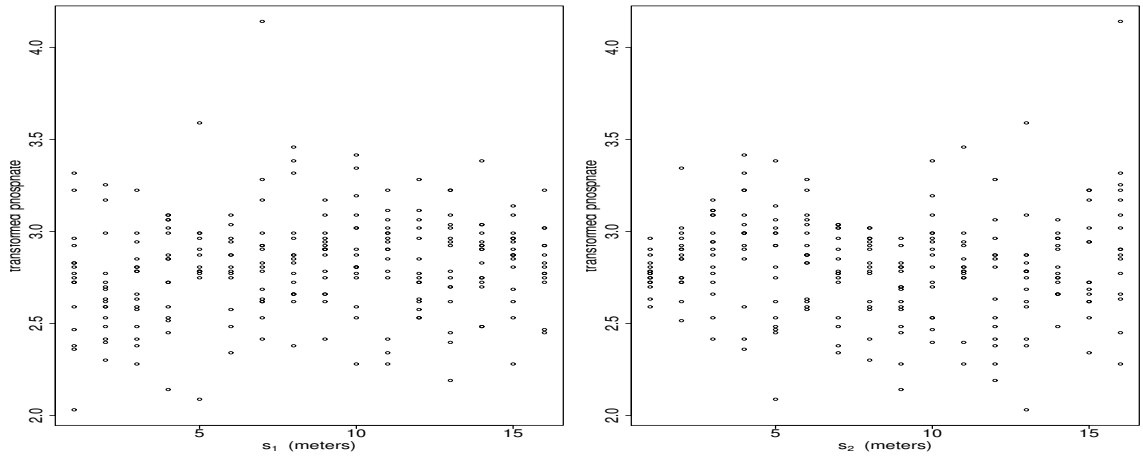

Figure 6: Plots of phosphate concentration readings versus sites coordinates.

A particular feature of this dataset is that there are missing observations at nine sites (marked with an '×' in Figure 5). In their analysis, CK did not mention how these missing observations were dealt with when fitting the model, although presumably they were inputed with the mean (or median) of the observed values at the neighboring sites. The Bayesian analysis below fully accounts for the uncertainty of the missing values.

CK built a model for this dataset based on exploratory data analysis and some numerical and graphical diagnostics developed in that paper. I mostly use the model selected by these authors, except for one difference. The original phosphate concentration readings were transformed as $\tilde{Y}_i = D_i^{\frac{1}{4}}$, $i = 1, \ldots, 256$, to obtain a reponse with distribution close to Gaussian. CK assumed that $E\{\tilde{Y}_i\} = \beta_1 + \beta_2 s_{i1} + \beta_3 s_{i2}$, with (s_{i1}, s_{i2}) the coordinates of site i, but I find no basis for this choice. There seems to be no apparent relation between the phosphate concentration readings and the sites coordinates, as indicated in Figure 6, so I assume $E\{\tilde{\mathbf{Y}}\} = \beta_1 \mathbf{1}$ ($\mathbf{1}$ is a vector of ones).

As for the spatial dependence structure, CK modeled these (transformed) data using a second order neighborhood system, meaning that the neighbors of site i are its first order neighbors and their first order neighbors (except for site i); the number of neighbors varies between 5 and 12. Let $a_{ij} = 1$ if sites i and j are neighbors, and $a_{ij} = 0$ otherwise. For the spatial dependence structure I use the model selected by CK, which they call the autocorrelation (homogeneous) CAR model. It is assumed that $\text{var}\{\tilde{\mathbf{Y}}\} = \sigma^2 (I_{256} - \phi \tilde{W})^{-1} G$, with

$$G = \text{diag}(|N_1|^{-1}, \ldots, |N_{256}|^{-1}) \quad \text{and} \quad \tilde{W}_{ij} = a_{ij} \big(|N_j|/|N_i|\big)^{\frac{1}{2}},$$

where $|N_i|$ the number of neighbors of site i. Finally following the discussion in Section 2.1, I

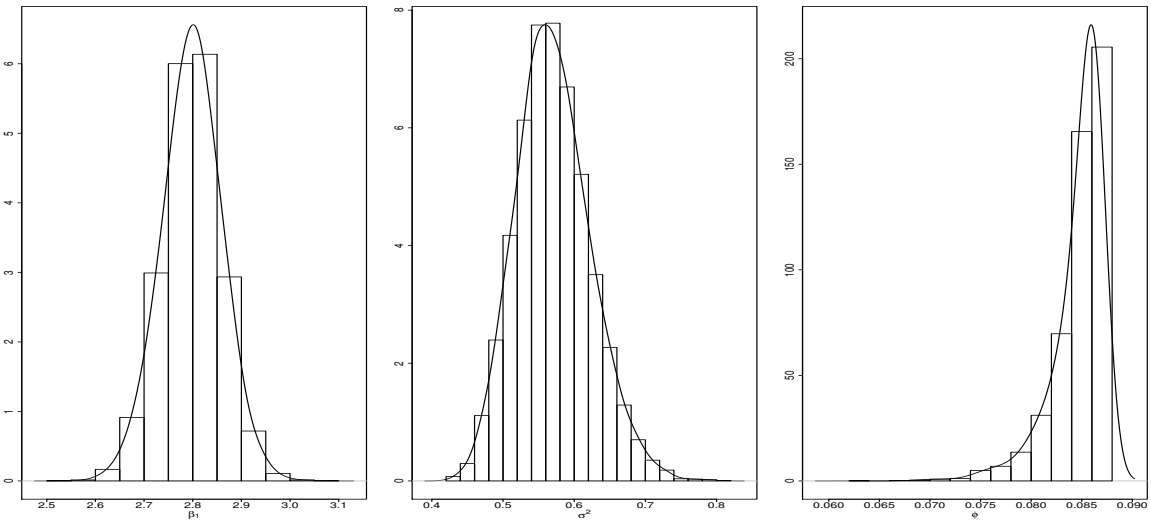

Figure 7: Posterior distribution of the model parameters based on the observed phosphate data.

work with the scaled data $\mathbf{Y} = G^{-\frac{1}{2}}\tilde{\mathbf{Y}}$, so the model to be fit is

$$\mathbf{Y} \sim \mathrm{N}_{256}(\beta_1 \mathbf{z}, \sigma^2 (I_{256} - \phi W)^{-1}), \tag{17}$$

where $\mathbf{z} = G^{-\frac{1}{2}}\mathbf{1}$ and $W = G^{-\frac{1}{2}}\tilde{W}G^{\frac{1}{2}}$, and the unknown parameters are $\beta_1 \in \mathbb{R}$, $\sigma^2 > 0$ and $\phi \in (-0.243062, 0.086614)$. It holds that neither $\mathbf{u}_1$ nor $\mathbf{u}_{256}$ belongs to $\mathcal{C}(\mathbf{z})$. As the original data contain nine missing values, we have that $\mathbf{y} = (\mathbf{y}_J, \mathbf{y}_I)$ where $\mathbf{y}_J$ denote the observed values and $\mathbf{y}_I$ the missing values (both displayed in Figure 5).

Model (17) was fit to the observed data $\mathbf{y}_J$ using the independence Jeffreys prior. The Gibbs sampling algorithm described in Section 4.1 was run to obtain a sample of size 10000 from the posterior $\pi(\beta_1, \sigma^2, \phi, \mathbf{y}_I \mid \mathbf{y}_J)$. Figure 7 displays the marginal posteriors of the model parameters. As is typical when fitting CAR models, the posterior of ϕ is highly concentrated around the right boundary of the parameter space. Summaries of the posterior distributions of the model parameters (posterior medians and 95% credible intervals) are given in Table 5.

Table 5: Summaries of the posterior distributions of the model parameters.

	β_1	σ^2	ϕ
estimate	2.7993	0.5676	0.0856
credible interval	(2.6763,2.9161)	(0.4768,0.6824)	(0.0774,0.0866)

Remark 5. A possible caveat in the above analysis is in order. If model (17) is assumed for $\mathbf{Y}$, then the form of the joint distribution of $\mathbf{Y}_J$ (the observed values) is unknown, and in particular it does not follow a CAR model. As a result Proposition 1 and Corollary 1 do not apply in this case and, strictly speaking, propriety of the posterior of the model parameters is not guaranteed. Nevertheless, the Monte Carlo output of the above analysis based on the independence Jeffreys prior was very close to that based on vague proper priors (normal–inverse gamma–uniform), so the possibility of an improper posterior in the above analysis seems remote.

7 Conclusions

This work derives two versions of Jeffreys priors for CAR models, which provide default (automatic) Bayesian analyses for these models, and obtains properties of Bayesian inferences based on them. It was found that inferences based on the Jeffreys priors and the uniform prior have similar frequentist properties under most scenarios, except when strong spatial association is present. In this case the independence Jeffreys prior displayed a superior performance. So this prior is the one I recommend, but note that the uniform prior is almost as good and may be preferred by some due to its simplicity. It was also found that inferences based on ML have inferior frequentist properties than inferences based on any of the three default priors.

Modeling of non-Gaussian (e.g. count) spatial data is often based on hierarchical models where CAR models are used to describe (unobserved) latent processes or spatially varying random effects. In this case choice of prior for the CAR model parameters has been done more or less ad hoc, and problems have been reported for their estimation. A tentative possibility to deal with this issue is to use one of the default priors proposed here, having in mind that this is not a Jeffreys prior in the hierarchical model context but a reasonably proxy at best. For this to be feasible, further research is needed to establish propriety of the relevant posterior distribution in the hierarchical model context and to determine inferential properties of procedures based on such default prior.

Appendix

Proof of Lemma 1.
Let $X'X = VTV'$ be the spectral decomposition of $X'X$, with V orthogonal, $V'V = VV' = I_p$, and T diagonal with positive diagonal elements (since $X'X$ is positive definite). Then

$$T^{-\frac{1}{2}}V'(X'\Sigma_\phi^{-1}X)VT^{-\frac{1}{2}} = X_o'\Sigma_\phi^{-1}X_o = I_p - \phi X_o'WX_o,$$

where

$$X_o = XVT^{-\frac{1}{2}}. \tag{18}$$

It then holds that $X_o'X_o = I_p$,

$$|X_o'\Sigma_\phi^{-1}X_o| = |VT^{-\frac{1}{2}}|^2|X'\Sigma_\phi^{-1}X|, \tag{19}$$

and $\text{rank}(X_o'\Sigma_\phi^{-1}X_o) = \text{rank}(X'\Sigma_\phi^{-1}X)$ since $|VT^{-\frac{1}{2}}| \neq 0$. The cases when $\mathbf{u}_1 \in \mathcal{C}(X)$ and when $\mathbf{u}_1 \notin \mathcal{C}(X)$ are now considered separately.

Suppose $\mathbf{u}_1 \in \mathcal{C}(X)$. In this case $\mathbf{u}_1 = X_o\mathbf{a}$ for some $\mathbf{a} \neq \mathbf{0}$ (since $\mathcal{C}(X) = \mathcal{C}(X_o)$), and then

$$(X_o'\Sigma_\phi^{-1}X_o)\mathbf{a} = X_o'(I_n - \phi W)\mathbf{u}_1 = (1-\phi\lambda_1)X_o'\mathbf{u}_1 = (1-\phi\lambda_n)\mathbf{a},$$

so $(1-\phi\lambda_1)$ is an eigenvalue of $X_o'\Sigma_\phi^{-1}X_o$. Now, for any $\mathbf{c} \in \mathbb{R}^p$, with $||\mathbf{c}|| = 1$, and $\phi \in (0, \lambda_1^{-1})$

$$\mathbf{c}'X_o'\Sigma_\phi^{-1}X_o\mathbf{c} = 1 - \phi\mathbf{c}_o'W\mathbf{c}_o \geq 1 - \phi\lambda_1, \quad \text{with} \quad \mathbf{c}_o = X_o\mathbf{c},$$

where the inequality holds by the extremal property of Rayleigh quotient (Schott, 2005 p. 105)

$$\lambda_1 = \max_{||\mathbf{c}||=1} \mathbf{c}'W\mathbf{c} = \mathbf{u}_1'W\mathbf{u}_1.$$

Hence $1-\phi\lambda_1$ is the smallest eigenvalue of $X_o'\Sigma_\phi^{-1}X_o$. In addition, $1-\phi\lambda_1$ must be simple. Otherwise there would exits at least two orthonormal eigenvectors associated to $1-\phi\lambda_1$, say $\mathbf{c}_1$ and $\mathbf{c}_2$, satisfying

$$1 - \phi\lambda_1 = \mathbf{c}_i'(X_o'\Sigma_\phi^{-1}X_o)\mathbf{c}_i = 1 - \phi\mathbf{c}_{oi}'W\mathbf{c}_{oi}, \quad \text{with} \quad \mathbf{c}_{oi} = X_o\mathbf{c}_i, \quad i = 1, 2,$$

which implies that $\lambda_1 = \mathbf{c}_{oi}'W\mathbf{c}_{oi}$, so $\mathbf{c}_{o1}$ and $\mathbf{c}_{o2}$ are two orthonormal eigenvectors of W associated with λ_1; but this contradicts the fact that λ_1 is a simple eigenvalue of W. Finally, if $v_1(\phi) \geq \ldots \geq v_{p-1}(\phi) > 1 - \phi\lambda_1 > 0$ are the eigenvalues of $X_o'\Sigma_\phi^{-1}X_o$, it follows from (19) that for all $\phi \in (0, \lambda_1^{-1})$

$$\begin{aligned} |X'\Sigma_\phi^{-1}X| &\propto |X_o'\Sigma_\phi^{-1}X_o| \\ &= (1-\phi\lambda_1)\prod_{i=1}^{p-1} v_i(\phi) = O(1-\phi\lambda_1) \quad \text{as} \quad \phi \to \lambda_1^{-1}. \end{aligned} \tag{20}$$

Suppose now $\mathbf{u}_1 \notin \mathcal{C}(X)$. In this case it must holds that $X_o'\Sigma^{-1}_{\lambda_1^{-1}}X_o$ is nonsingular. Otherwise, if $X_o'\Sigma^{-1}_{\lambda_1^{-1}}X_o$ were singular, there is $\mathbf{b} \neq \mathbf{0}$ with $||\mathbf{b}|| = 1$ for which $X_o'WX_o\mathbf{b} = \lambda_1\mathbf{b}$, so λ_1 is an eigenvalue of $X_o'WX_o$ with $\mathbf{b}$ as its associated eigenvector. Then

$$\lambda_1 = \mathbf{b}'X_o'WX_o\mathbf{b} = \mathbf{b}_o'W\mathbf{b}_o, \quad \text{with} \quad \mathbf{b}_o = X_o\mathbf{b}.$$

By the extremal property of Rayleigh quotient $\mathbf{b}_o$ is also an eigenvector of W associated with the eigenvalue λ_1. Since λ_1 is simple, there is $t \neq 0$ for which $\mathbf{u}_1 = t\mathbf{b}_o = X(tVT^{-\frac{1}{2}}\mathbf{b})$, with $tVT^{-\frac{1}{2}}\mathbf{b} \neq \mathbf{0}$, implying that $\mathbf{u}_1 \in \mathcal{C}(X)$; but this is a contradiction. From (19) it follows that $X'\Sigma^{-1}_{\lambda_1^{-1}}X$ is non-singular and since $|X'\Sigma_\phi^{-1}X|$ is a continuous function of ϕ, $|X'\Sigma_\phi^{-1}X|^{-\frac{1}{2}} = O(1)$ as $\phi \to \lambda_1^{-1}$. This and (20) prove (6).

It is now shown that for every $\boldsymbol{\eta} \in \Omega$, $S^2_{\lambda_1^{-1}} > 0$ with probability 1. Let $\mathcal{C}(X, \mathbf{u}_1)$ denote the subspace of $\mathbb{R}^n$ spanned by the columns of X and $\mathbf{u}_1$. If $\mathbf{y} \in \mathcal{C}(X, \mathbf{u}_1)$, then $\mathbf{y} = X\mathbf{a} + t\mathbf{u}_1$ for some $\mathbf{a} \in \mathbb{R}^p$ and $t \in \mathbb{R}$, and hence $X'\Sigma^{-1}_{\lambda_1^{-1}}X\mathbf{a} = X'\Sigma^{-1}_{\lambda_1^{-1}}\mathbf{y}$. This means that $\mathbf{a}$ is a solution to the (generalized) normal equations, so $X\mathbf{a} = X\hat{\boldsymbol{\beta}}_{\lambda_1^{-1}}$ and $\mathbf{y} - X\hat{\boldsymbol{\beta}}_{\lambda_1^{-1}} = t\mathbf{u}_1$, which implies that $S^2_{\lambda_1^{-1}} = 0$.
Suppose now that $S^2_{\lambda_1^{-1}} = 0$. Since 0 is the smallest eigenvalue of $\Sigma^{-1}_{\lambda_1^{-1}}$, and is simple with $\mathbf{u}_1$ as its associated eigenvector, it follows by the extremal property of the Rayleigh quotient that $\mathbf{y} - X\hat{\boldsymbol{\beta}}_{\lambda_1^{-1}} = t\mathbf{u}_1$ for some $t \in \mathbb{R}$, which implies that $\mathbf{y} \in \mathcal{C}(X, \mathbf{u}_1)$. Since $n \geq p + 2$, $\mathcal{C}(X, \mathbf{u}_1)$ is a proper subspace of $\mathbb{R}^n$ and because $\mathbf{Y}$ has an absolutely continuous distribution

$$P(S^2_{\lambda_1^{-1}} > 0 \mid \boldsymbol{\eta}) = P(\mathbf{Y} \notin \mathcal{C}(X, \mathbf{u}_1) \mid \boldsymbol{\eta}) = 1 \qquad \text{for every } \boldsymbol{\eta} \in \Omega.$$

Since S^2_ϕ is a continuous function of ϕ, it holds with probability 1 that $S^2_\phi = O(1)$ as $\phi \to \lambda_1^{-1}$. The proofs of the results as $\phi \to \lambda_n^{-1}$ follow along the same lines.

References

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

Bapat, R.B. and Raghavan, T.E.S. (1997). *Nonnegative Matrices and Applications*. Cambridge University Press.

Bell, B.S. and Broemeling, L.D. (2000). A Bayesian Analysis of Spatial Processes With Application to Disease Mapping. *Statistics in Medicine*, 19, 957-974.

Berger, J.O., De Oliveira, V. and Sansó, B. (2001). Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association*, 96, 1361-1374.

Berger, J.O. and Bernardo, J.M. (1992). On the Development of the Reference Prior Method. In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford University Press, pp. 35-60.

Bernardo, J.M (1979). Reference Posterior Distributions for Bayes Inference (with discussion). *Journal of the Royal Statistical Society Ser. B*, 41, 113-147.

Besag, J. and Kooperberg, C. (1995). On Conditional and Intrinsic Autoregressions. *Biometrika*, 82, 733-746.

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.

Case, A.C., Rosen, H.S. and Hines, J.R. (1993). Budget Spillovers and Fiscal Policy Interdependence. *Journal of Public Economics*, 52, 285-307.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes Estimates of Age-standardized Relative Risks for use in Disease Mapping. *Biometrics*, 43, 671-681.

Cressie, N.A.C. (1993). *Statistics for Spatial Data* (rev. ed.). Wiley.

Cressie, N. and Kapat, P. (2008). Some Diagnostics for Markov Random Fields. *Journal of Computational and Graphical Statistics*, 17, 726-749.

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

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

Gilks, W.R., Best, N.G. and Tan, K.K.C. (1995). Adaptive Rejection Metropolis Sampling Within Gibbs Sampling. *Applied Statistics*, 44, 455–472.

Mardia, K.V. and R.J. Marshall (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, 71, 135-146.

Militino, A.F., Ugarte, M.D. and Garcia-Reinaldos, L. (2004). Alternative Models for Describing Spatial Dependence Among Dwelling Selling Prices. *Journal of Real Estate Finance and Economics*, 29, 193-209.

Pettitt, A.N., Weir, I.S. and Hart, A.G. (2002). A Conditional Autoregressive Gaussian Process for Irregularly Spaced Multivariate Data with Application to Modelling Large Sets of Binary Data. *Statistics and Computing*, 12, 353-367.

Richardson, S., Guihenneuc, C. and Lasserre, V. (1992). Spatial Linear Models with Autocorrelated Error Structure. *The Statistician*, 41, 539-557.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC.

Schott, J.R. (2005). *Matrix Analysis for Statistics*, 2nd. ed. Wiley.

Sun, D, Tsutakawa, R.K. and Speckman, P.L. (1999). Posterior Distribution of Hierarchical Models Using CAR(1) Distributions. *Biometrika*, 86, 341-350.