Ann Inst Stat Math (2012) 64:107–133
DOI 10.1007/s10463-010-0298-1

Bayesian analysis of conditional autoregressive models

Victor De Oliveira

Received: 18 December 2008 / Revised: 4 January 2010 / Published online: 27 May 2010

Abstract Conditional 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 illustrated by fitting a CAR model to a phosphate dataset from an archaeological region.

Keywords CAR model · Eigenvalues and eigenvectors · Frequentist properties · Integrated likelihood · Maximum likelihood · Spatial data · Weight matrix

1 Introduction

Conditional autoregressive (CAR) models are often 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

V. De Oliveira (✉)
Department of Management Science and Statistics,
The University of Texas at San Antonio, San Antonio, TX 78249, USA
e-mail: victor.deoliveira@utsa.edu

unveil and quantify spatial relations present among the data, in particular, to quantify how quantities of interest vary with explanatory variables and to detect clusters of 'hot spots'. General accounts of CAR models, a class of Gaussian Markov random fields, appear in Cressie (1993), Banerjee et al. (2004) and Rue and Held (2005).

CAR models have been extensively used in spatial statistics to model observed data (Cressie and Chan 1989; Richardson et al. 1992; Bell and Broemeling 2000; Militino et al. 2004; Cressie et al. 2005), as well as (unobserved) latent variables and spatially varying random effects (Clayton and Kaldor 1987; Sun et al. 1999; Pettitt et al. 2002; see Banerjee et al. 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 a default Bayesian analysis of 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 are asymptotic and little is known 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 estimate 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 knowledge about adequate priors for these models and frequentist properties of the resulting Bayesian procedures.

The main goal 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 model 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 to sample from the appropriate posterior distributions. Examples are presented based on simulated data to compare the two Jeffreys priors and their corresponding posterior distributions.

A simulation experiment is performed to compare frequentist properties of inferences about the covariance parameters based on maximum likelihood (ML) with those based on the proposed Jeffreys priors and a uniform prior. It is found that frequentist properties of the above Bayesian procedures are better than those of ML. In addition, frequentist properties of the above Bayesian procedures are adequate and similar to each other in most situations, except when the mean of the observations is not constant or the spatial association is strong. In these cases, inference about the 'spatial parameter' based on the independence Jeffreys prior has better frequentist properties than the procedures based on the other priors. 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 model parameters. Section 4 describes a simple Monte Carlo algorithm to sample from the posterior distribution, 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 archaeological 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

$$N_i = \{j : \text{subregions } i \text{ and } j \text{ share a boundary}\}, \quad i = 1, \ldots, n.$$

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 et al. (1993) for examples.

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 autoregression given by

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

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 $M = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ 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

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

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 appropriate transformations) the matrices M and C satisfy:

(i) $M = \sigma^2 I_n$, with $\sigma^2 > 0$ unknown;
(ii) $C = \phi W$, with ϕ an unknown 'spatial parameter' and $W = (w_{ij})$ 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 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ are the ordered eigenvalues of W, with $\lambda_n < 0 < \lambda_1$ since $\text{tr}(W) = 0$. 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 depends on some aspects of the design (as it depends on W). Finally, the parameter value $\phi = 0$ corresponds to the case when $Y_i - \mathbf{x}_i'\boldsymbol{\beta} \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

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. If M and C satisfy (a) and (b), then $\mathbf{Y} = G^{-\frac{1}{2}}\tilde{\mathbf{Y}}$ satisfies

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

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

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

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

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

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 $0 < \int_\Omega L(\boldsymbol{\eta}; \mathbf{y})\pi(\boldsymbol{\eta})\mathrm{d}\boldsymbol{\eta} < \infty$. A standard calculation with the above likelihood and prior shows that

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

with

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

where

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

$L^I(\phi;\mathbf{y})$ is called the *integrated likelihood* of ϕ. Then the posterior distribution of $\boldsymbol{\eta}$ is proper if and only if

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

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 $\boldsymbol{\eta} \in \Omega$

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

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 "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 $\phi \to \lambda_1^{-1}$*

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

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 Σ_ϕ^{-1}, S_ϕ^2 and the continuity of the determinant function. For any $\phi \in (0, \lambda_1^{-1})$ the

eigenvalues of Σ_ϕ^{-1} are $1-\phi\lambda_1 < 1-\phi\lambda_2 \leq \cdots \leq 1-\phi\lambda_{n-1} < 1-\phi\lambda_n$, 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 $\phi \to \lambda_i^{-1}$, $i = 1, 2$, 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,

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

and

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

where $\nu_1 \geq \cdots \geq \nu_p$ *are the ordered eigenvalues of* $X_o' W X_o$*, and* X_o *is the matrix defined by* (17) *in "Appendix".*

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

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

and

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

where $U_\phi = \left(\frac{\partial}{\partial \phi} \Sigma_\phi \right) \Sigma_\phi^{-1}$, 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

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

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

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

so the first result follows. The second result follows from the first and identity (18) in "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

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

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* $\pi^{J1}(\boldsymbol{\eta}|\mathbf{y})$ *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* $\pi^{J1}(\phi|\mathbf{y})$ *(when exists) has moments of any order* k.
(iv) *The marginal independence Jeffreys posterior* $\pi^{J1}(\sigma^2|\mathbf{y})$ *(when exists) has a finite moment of order* k *if* $n \geq p + 2k + 1$.
(v) *For* $j = 1, \ldots, p$, *the marginal independence Jeffreys posterior* $\pi^{J1}(\beta_j|\mathbf{y})$ *(when exists) has a finite moment of order* k *if* $n \geq p + k + 1$.

Proof See "Appendix". □

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* $\pi^{J2}(\boldsymbol{\eta}|\mathbf{y})$ *is always proper.*
(iii) *The marginal Jeffreys-rule posterior* $\pi^{J2}(\phi|\mathbf{y})$ *has always moments of any order* k.

(iv) *The marginal Jeffreys posterior* $\pi^{J2}(\sigma^2|\mathbf{y})$ *has a finite moment of order* k *if* $n \geq p + 2k + 1$.
(v) *For* $j = 1, \ldots, p$, *the marginal Jeffreys-rule posterior* $\pi^{J2}(\beta_j|\mathbf{y})$ *has a finite moment of order* k *if* $n \geq p + k + 1$.

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

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

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

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

where $\upsilon_1(\phi), \ldots, \upsilon_{n-p}(\phi)$ are the nonzero eigenvalues of the matrix $V_\phi = \left(\frac{\partial}{\partial \phi} \Sigma_\phi \right) \Sigma_\phi^{-1} P_\phi^W$, 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 Jeffreys-rule prior. Unfortunately, I have not been able to find an explicit expression for the above eigenvalues, and properties

of Bayesian inferences based on this prior remain unknown. In addition, it was found for the data analysis in Sect. 6 that inferences about $(\boldsymbol{\beta}, \sigma^2, \phi)$ based on proper diffuse normal-inverse gamma-uniform priors were similar as those based on prior (3) with $a = 1$ and $\pi^U(\phi)$, so these will not be considered further.

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

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

where from (2) and (3)

$$\pi(\boldsymbol{\beta}|\sigma^2, \phi, \mathbf{y}) = N_p\left(\hat{\boldsymbol{\beta}}_\phi, \sigma^2(X'\Sigma_\phi^{-1}X)^{-1}\right), \tag{9}$$

$$\pi(\sigma^2|\phi, \mathbf{y}) = \text{IG}\left(\frac{n-p}{2} + a - 1, \frac{1}{2}S_\phi^2\right), \tag{10}$$

$$\pi(\phi|\mathbf{y}) \propto \left(|\Sigma_\phi||X'\Sigma_\phi^{-1}X|\right)^{-\frac{1}{2}} \left(S_\phi^2\right)^{-\left(\frac{n-p}{2}+a-1\right)} \pi(\phi). \tag{11}$$

Simulation from (9) and (10) is straightforward, while simulation from (11) would be accomplished using the adaptive rejection Metropolis sampling (ARMS) algorithm proposed by Gilks et al. (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; see Sect. 6.

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 20×20 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-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,

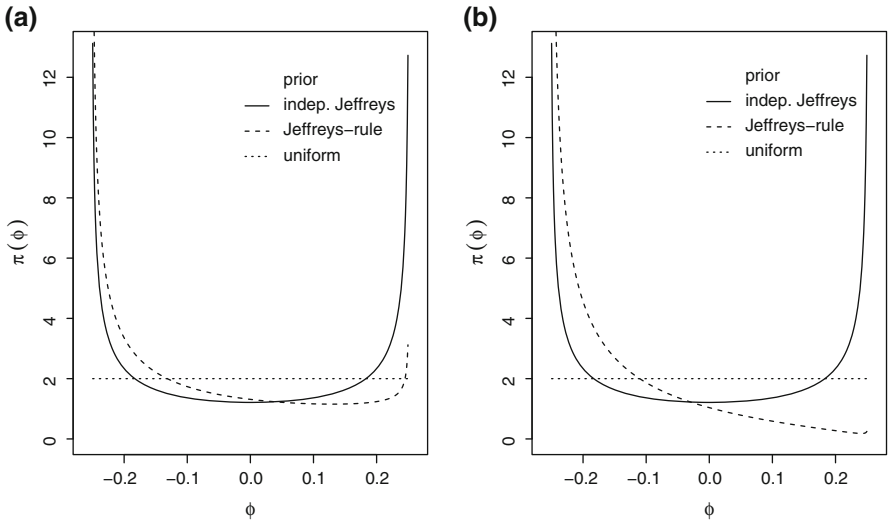

Fig. 1 Marginal independence Jeffreys, Jeffreys-rule and uniform priors of ϕ for models over a 20×20 regular lattice with a first order (or 'rook') neighborhood system when the mean is **a** constant and **b** a degree one polynomial in the site coordinates

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 data. The mean of the observations is 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 $\pi^{J1}(\phi|\mathbf{y})$ being shifted to the right and less disperse when compared to $\pi^{J2}(\phi|\mathbf{y})$, and $\pi^{U}(\phi|\mathbf{y})$ 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 (12). The same patterns were observed for several other simulated datasets (not shown).

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

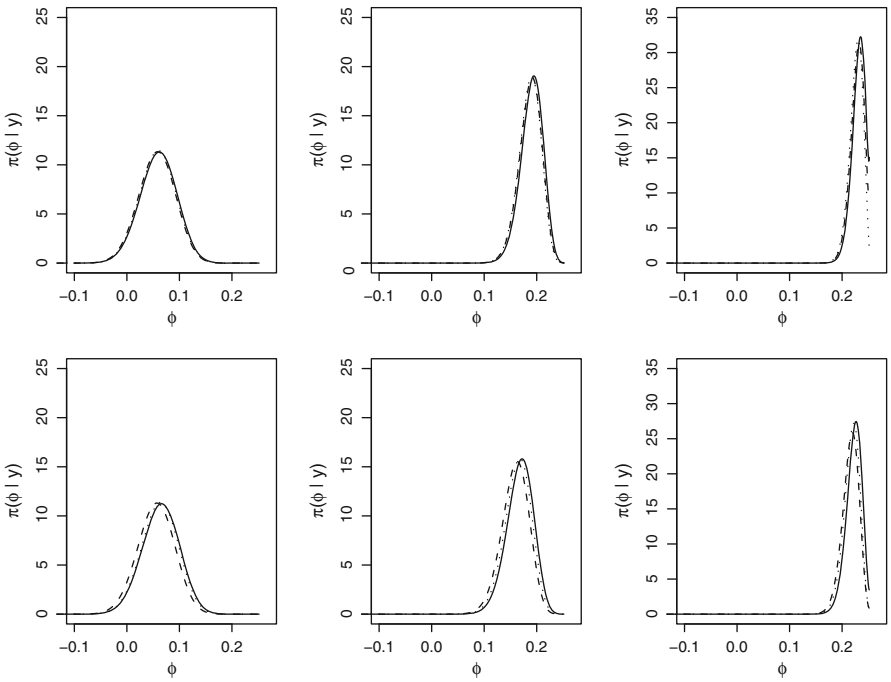

Fig. 2 Marginal independence Jeffreys (*solid*), Jeffreys-rule (*dashed*) and uniform (*dotted*) posteriors of ϕ based on simulated data for models over a 20×20 regular lattice with a first order neighborhood system. $E\{Y_i\}$ is either 10 (*top panels*) or $10 + s_{i1} + s_{i2}$ (*bottom panels*),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

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 squared error of estimators. For the Bayesian procedures, I use the 95% equal-tailed credible intervals for σ^2 and ϕ, and the posterior means as their estimators. 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, $\text{avar}(\cdot)$ denotes asymptotic variance and $\hat{\text{avar}}(\cdot)$ denotes $\text{avar}(\cdot)$ 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

$$\text{avar}(\hat{\sigma}^2) = \frac{2\sigma^4}{n(g(\phi))^2} \sum_{i=1}^{n} \left(\frac{\lambda_i}{1 - \phi\lambda_i} \right)^2 \quad \text{and} \quad \text{avar}(\hat{\phi}) = \frac{2}{(g(\phi))^2}, \tag{12}$$

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

I consider models defined on a 10×10 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

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

	$p = 1$			$p = 6$		
	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$
$\sigma^2 = 0.1$						
Ind. Jeffreys	0.946	0.950	0.948	0.945	0.941	0.949
Jeffreys-rule	0.945	0.947	0.946	0.901	0.915	0.945
Uniform	0.947	0.948	0.943	0.948	0.945	0.946
ML	0.914	0.923	0.927	0.840	0.868	0.928
$\sigma^2 = 2.0$						
Ind. Jeffreys	0.936	0.939	0.952	0.942	0.944	0.949
Jeffreys-rule	0.933	0.940	0.954	0.895	0.916	0.944
Uniform	0.934	0.939	0.952	0.946	0.944	0.946
ML	0.913	0.923	0.943	0.846	0.870	0.942

ϕ. I consider $E\{Y_i\}$ equal to 10 ($p = 1$) or $10 + s_{i1} + s_{i2} + s_{i1}s_{i2} + s_{i1}^2 + s_{i2}^2$ ($p = 6$), 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). For all these scenarios, neither $\mathbf{u}_1$ nor $\mathbf{u}_n$ belong to $\mathcal{C}(X)$, so the posterior based on any of the default priors is proper. 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, 1,500 datasets were simulated and for each dataset a posterior sample of the model parameters of size $m = 3{,}000$ was generated by the algorithm described in Sect. 4.

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 ML confidence intervals are below nominal, while the coverage of the credible intervals based on the independence Jeffreys and uniform priors are similar to each other and reasonably close to the nominal 0.95. On the other hand, the coverage of the credible intervals based on the Jeffreys-rule prior are below nominal when the mean of the observations is not constant.

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 and the same hold (substantially so) for the credible intervals based on the Jeffreys-rule prior when the mean of the observations is not constant. On the other hand, the coverage of the credible intervals based on the independence Jeffreys and uniform priors are similar to each other and reasonably close to the nominal 0.95 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.

Table 3 shows (empirical) mean squared error (MSE) of the posterior mean of σ^2 corresponding to the three default priors and the MSE of the ML estimator of σ^2.

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

	$p = 1$			$p = 6$		
	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$
$\sigma^2 = 0.1$						
Ind. Jeffreys	0.954	0.954	0.972	0.940	0.938	0.978
Jeffreys-rule	0.948	0.946	0.949	0.897	0.853	0.794
Uniform	0.962	0.964	0.883	0.967	0.968	0.885
ML	0.930	0.932	0.904	0.858	0.844	0.827
$\sigma^2 = 2.0$						
Ind. Jeffreys	0.952	0.950	0.966	0.944	0.938	0.982
Jeffreys-rule	0.948	0.946	0.951	0.891	0.860	0.791
Uniform	0.962	0.956	0.883	0.963	0.965	0.865
ML	0.929	0.931	0.926	0.838	0.849	0.828

Table 3 Mean squared error $\times 10^2$ of the posterior means and ML estimate of σ^2

	$p = 1$			$p = 6$		
	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$
$\sigma^2 = 0.1$						
Ind. Jeffreys	0.0213	0.0213	0.0239	0.0221	0.0222	0.0245
Jeffreys-rule	0.0208	0.0209	0.0240	0.0233	0.0221	0.0232
Uniform	0.0212	0.0213	0.0282	0.0221	0.0223	0.0295
ML	0.0207	0.0205	0.0226	0.0250	0.0233	0.0225
$\sigma^2 = 2.0$						
Ind. Jeffreys	8.734	9.362	9.262	9.124	8.719	9.916
Jeffreys-rule	8.560	9.991	9.298	9.330	8.771	9.391
Uniform	8.738	9.378	10.925	9.105	8.693	11.650
ML	8.482	8.907	8.785	9.970	9.285	9.112

The MSEs of the Bayesian estimators based on the three default priors and the ML estimator are close to each other under all scenarios.

Table 4 shows (empirical) MSE of the posterior mean of ϕ corresponding to the three default priors and the ML estimator of ϕ. For small or moderate values of ϕ the MSEs of the three Bayesian estimators and the ML estimator are close to each other, with the MSE 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 MSE of the Bayesian estimator based on the independence Jeffreys prior is substantially smaller than the MSEs of the other three estimators. Also, when the mean of the observations is not constant the MSE of the estimator based on the independence Jeffreys prior is smaller than the MSE of the estimator based on the Jeffreys-rule prior.

Table 4 Mean squared error $\times 10^2$ of the posterior means and ML estimate of ϕ

	$p = 1$			$p = 6$		
	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$	$\phi = 0.05$	$\phi = 0.12$	$\phi = 0.25$
$\sigma^2 = 0.1$						
Ind. Jeffreys	0.472	0.421	0.084	0.608	0.528	0.135
Jeffreys-rule	0.471	0.457	0.122	0.794	0.984	0.671
Uniform	0.377	0.380	0.167	0.422	0.418	0.329
ML	0.476	0.453	0.125	0.807	0.992	0.691
$\sigma^2 = 2.0$						
Ind. Jeffreys	0.487	0.397	0.096	0.621	0.549	0.150
Jeffreys-rule	0.486	0.428	0.135	0.845	0.967	0.690
Uniform	0.390	0.356	0.182	0.439	0.426	0.343
ML	0.492	0.425	0.139	0.859	0.975	0.708

Table 5 Frequentist coverage of Bayesian equal-tailed 95% credible intervals and large sample 95% confidence interval for ϕ (top), and mean squared error $\times 10^2$ of the posterior means and ML estimate of ϕ (bottom) (both based on a 20×20 regular lattice)

	$p = 1$		$p = 6$	
	$\phi = 0.12$	$\phi = 0.24$	$\phi = 0.12$	$\phi = 0.24$
Frequentist coverage				
Ind. Jeffreys	0.957	0.941	0.943	0.943
Jeffreys-rule	0.950	0.940	0.925	0.843
Uniform	0.953	0.932	0.949	0.940
ML	0.957	0.931	0.933	0.899
Mean squared error				
Ind. Jeffreys	0.0950	0.0117	0.0963	0.0147
Jeffreys-rule	0.0980	0.0140	0.1328	0.0378
Uniform	0.0950	0.0158	0.0947	0.0218
ML	0.0979	0.0139	0.1311	0.0376

To gain some insight on the behavior of Bayesian and ML inferences for larger sample sizes, a more limited simulation was run for models defined on a 20×20 regular lattice with first order neighborhood system and W the adjacency matrix. For this case $E\{Y_i\}$ is the same as in the previous simulation, $\sigma^2 = 2$ and ϕ is either 0.12 or 0.24. Table 5 shows (empirical) frequentist coverage of Bayesian equal-tailed 95% credible intervals for ϕ and large sample 95% confidence intervals for ϕ (top), as well as (empirical) MSE of the posterior mean of ϕ and the ML estimator of ϕ (bottom). The frequentist coverage of credible intervals based on the three default priors and ML are similar under most scenarios, except when the mean is not constant or ϕ is large. In these cases coverage of intervals based on the Jeffreys-rule prior and ML are below nominal, although to a lesser extent than for the 10×10 regular lattice. The MSEs of all estimators are similar under most scenarios, except when the mean is not constant in which case estimators based on the Jeffreys-rule prior and ML have larger MSE.

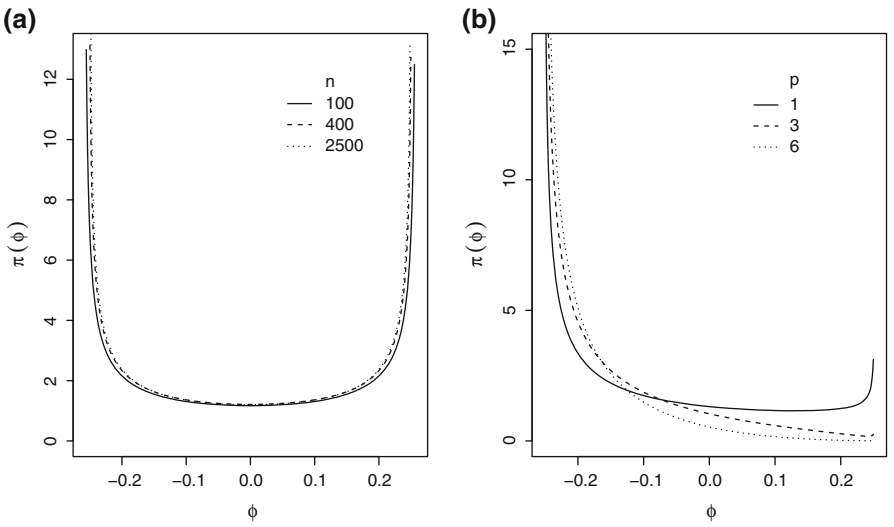

Fig. 3 **a** Marginal independence Jeffreys priors of ϕ defined for models over 10×10, 20×20 and 50×50 regular lattices with first order neighborhood system. **b** Marginal Jeffreys-rule priors of ϕ for models defined over a 20×20 regular lattices with the same neighborhood system as in **a**, and with mean constant, a degree one polynomial and a degree two polynomial in the site coordinates

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 mean of the observations is not constant or the spatial parameter ϕ is large. In these cases, frequentist properties of Bayesian inferences based on the independence Jeffreys prior are similar or better than those based on any of the other two default priors, specially 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 the default priors are to these features.

Sample size. Consider models defined over 10×10, 20×20 and 50×50 regular lattices with first order neighborhood system and W the adjacency matrix. Figure 3a 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 sample size was displayed by $\pi^{J2}(\phi)$, provided the models have the same mean structure (not shown).

Regression matrix. Consider models defined over a 20×20 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 3b 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 a 20×20 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 10×10 regular lattice with first order neighborhood system ϕ must belong to $(-0.260554, 0.260554)$, while for the irregular lattice formed by the 100 counties of the state of North Carolina in the United States with neighborhood system defined in terms of geographic adjacency ϕ must belong to $(-0.327373, 0.189774)$.

6 Example

To illustrate the Bayesian analysis of CAR models I use a spatial dataset initially analyzed by Buck et al. (1988), and more recently reanalyzed by Cressie and Kapat (2008) (from now on referred to as CK). The dataset consists of raw phosphate concentration readings (in mg P/100 g of soil) collected over several years in an archaeological region of Laconia across the Evrotas river in Greece. The original observations were collected 10 m apart over a 16×16 regular lattice; they are denoted by $\{D_i : i = 1, \ldots, 256\}$ and displayed in Fig. 4. A particular feature of this dataset is that there are missing observations at nine sites (marked with '$\times$' in Fig. 4). In their analysis, CK did not mention how these missing observations were dealt with when fitting the model, although presumably they were imputed with summaries of observed values at neighboring sites. Initially, I follow this imputation approach, but an alternative (fully) Bayesian analysis is also provided that accounts for the uncertainty of the missing values.

CK built a model for this dataset based on exploratory data analysis and graphical diagnostics developed in their paper. I mostly use the model selected by them, except for one difference indicated below. The original phosphate concentration readings were transformed as $\tilde{Y}_i = D_i^{\frac{1}{4}}$, $i = 1, \ldots, 256$, to obtain a response 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 (linear) relation between the phosphate concentration readings and the sites coordinates, as seen from Fig. 5, so I assume $E\{\tilde{\mathbf{Y}}\} = \beta_1 \mathbf{1}$ ($\mathbf{1}$ is the vector of ones).

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). Let $a_{ij} = 1$ if sites i and j are neighbors, and $a_{ij} = 0$ otherwise. For the spatial association structure, it is assumed that

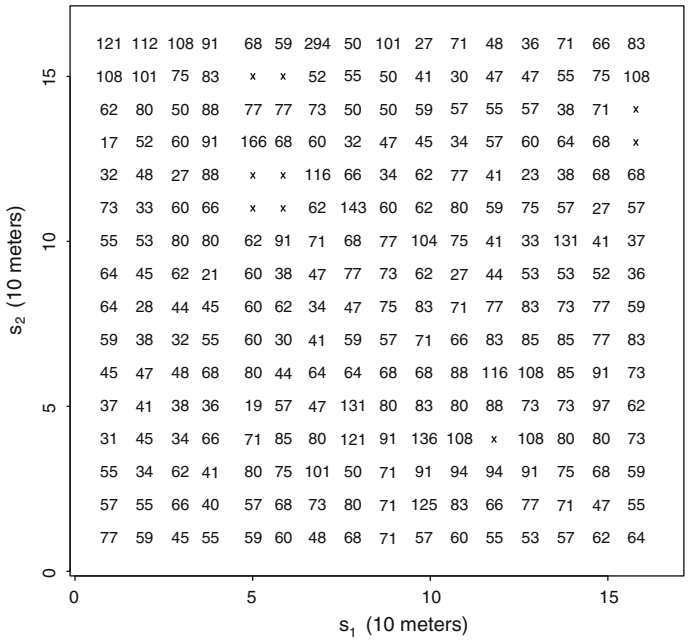

Fig. 4 Phosphate concentration readings (in mg P/100 g of soil) measured over a 16×16 regular lattice. Locations where values are missing are indicated with *x symbol*

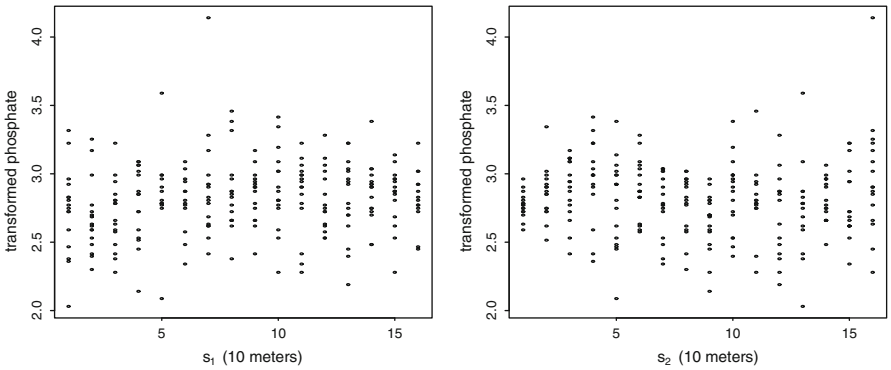

Fig. 5 Plots of phosphate concentration readings versus sites coordinates

$\text{var}\{\tilde{\mathbf{Y}}\} = \sigma^2(I_{256} - \phi\tilde{W})^{-1}G$, with

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

where $|N_i|$ is the number of neighbors of site i, varying between 5 and 12 in this lattice. This model is called by CK the autocorrelation (homogeneous) CAR model.

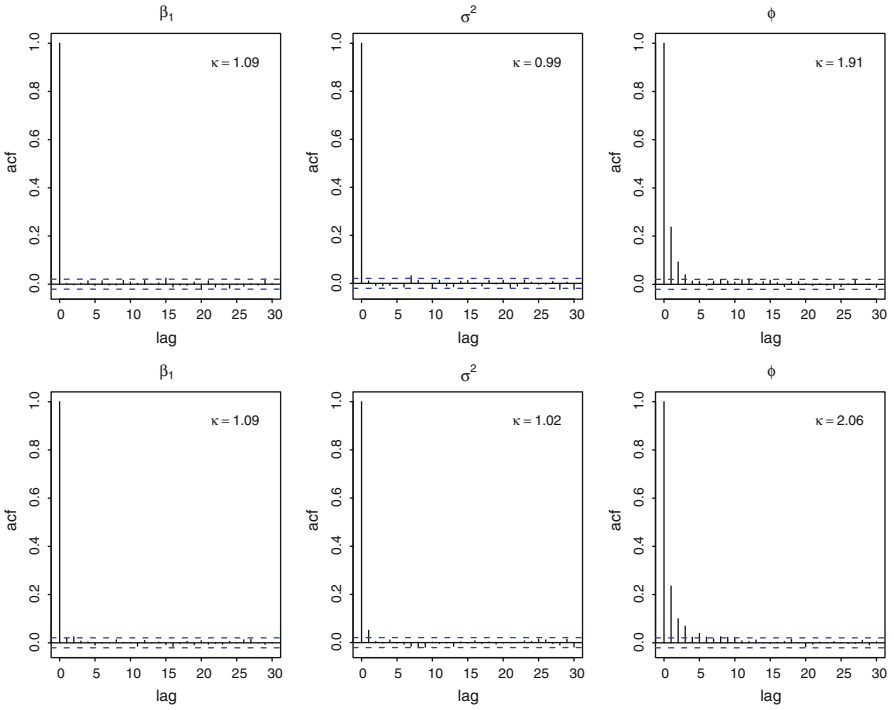

Fig. 6 Sample autocorrelations of the simulated values of β_1, σ^2 and ϕ using the algorithm described in Sect. 4.1 based on completed data $\mathbf{y}_C$ and the independence Jeffreys prior (*top panels*), and using the algorithm described in Sect. 6.1 based on the observed data $\mathbf{y}_J$ and the independence Jeffreys prior (*bottom panels*). The values κ are inefficiency factors

Finally, following the discussion at the end of Sect. 2.1, I work with the scaled data $\mathbf{Y} = G^{-\frac{1}{2}}\tilde{\mathbf{Y}}$ so the model to be fit is

$$\mathbf{Y} \sim N_{256}(\beta_1 \mathbf{z}, \sigma^2 (I_{256} - \phi W)^{-1}), \tag{13}$$

where $\mathbf{z} = G^{-\frac{1}{2}}\mathbf{1}$, $W = G^{-\frac{1}{2}}\tilde{W}G^{\frac{1}{2}}$ 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 $\mathcal{C}(\mathbf{z})$.

Let $\mathbf{y} = (\mathbf{y}_J, \mathbf{y}_I)$ denote the 'complete data', where J and I are the sites that correspond, respectively, to the observed and missing values, and $\mathbf{y}_J = \{y_j : j \in J\}$ are the observed values. Based on the independence Jeffreys, Jeffreys-rule and uniform priors, model (13) was fit to the 'completed data' $\mathbf{y}_C = (\mathbf{y}_J, \hat{\mathbf{y}}_I)$, where the components of $\hat{\mathbf{y}}_I$ are the medians of the respective neighboring observed values. The Monte Carlo algorithm described in Sect. 4.1 was run to obtain a sample of size 10,000 from the posterior $\pi(\beta_1, \sigma^2, \phi|\mathbf{y}_C)$. Figure 6 (top panels) displays sample autocorrelations, $\hat{\rho}(h)$, $h = 0, 1, \ldots, 30$, of the simulated values of β_1, σ^2 and ϕ based on the independence Jeffreys prior, as well as the inefficiency factor $\kappa = 1 + 2\sum_{h=1}^{30}\left(1 - \frac{h}{30}\right)\hat{\rho}(h)$. The cross-correlations are all small, with the largest being that between σ^2 and ϕ

Table 6 Summaries of the marginal posterior distributions (2.5% quantile, mean, 97.5% quantile) using the three default priors, based on the *completed* phosphate data (top part) and the *observed* phosphate data (bottom part)

	Independence Jeffreys			Jeffreys-rule			Uniform		
	2.5%	Mean	97.5%	2.5%	Mean	97.5%	2.5%	Mean	97.5%
From completed data $\mathbf{y}_C$									
β_1	2.6933	2.8136	2.9335	2.6962	2.8139	2.9312	2.7150	2.8162	2.9171
σ^2	0.4757	0.5654	0.6735	0.4766	0.5669	0.6766	0.4781	0.5716	0.6792
ϕ	0.0783	0.0848	0.0866	0.0767	0.0843	0.0866	0.0741	0.0824	0.0863
From observed data $\mathbf{y}_J$									
β_1	2.6742	2.7973	2.9160	2.6788	2.7985	2.9128	2.7007	2.8049	2.9047
σ^2	0.4779	0.5701	0.6825	0.4757	0.5702	0.6830	0.4756	0.5782	0.6830
ϕ	0.0785	0.0849	0.0866	0.0765	0.0843	0.0866	0.0737	0.0823	0.0862

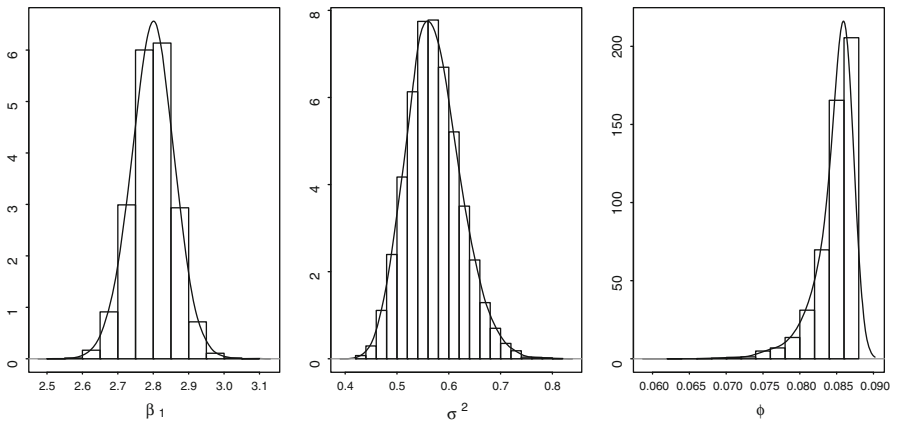

Fig. 7 Posterior distribution of the model parameters based on the observed phosphate data and the independence Jeffreys prior

(= 0.18). These summaries show the Markov chain mixes well and converges fast to the posterior distribution.

After discarding the initial burn-in section of the first 1,000 iterates, summaries of the marginal posterior distributions (2.5% quantile, mean, 97.5% quantile) are given in Table 6 (top part), where it is seen that the three analyses produce essentially the same results. Figure 7 displays the marginal posterior densities 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.

6.1 An alternative analysis

I now describe results from a Bayesian analysis that accounts for the uncertainty in $\mathbf{y}_I$. The algorithm in Sect. 4.1 can not be modified to handle the case of missing data, so sampling from the posterior of model parameters and missing values is done using

a Gibbs sampling algorithm. As before the full conditional distribution of $\boldsymbol{\beta}$ is given by (9), while the full conditionals of the other components of $(\mathbf{y}_I, \boldsymbol{\beta}, \sigma^2, \phi)$ are given by

$$\pi(y_i|\boldsymbol{\beta}, \sigma^2, \phi, \mathbf{y}_{(i)}) = N\left(\mathbf{x}_i'\boldsymbol{\beta} + \phi \sum_{j=1}^{n} w_{ij}(y_j - \mathbf{x}_j'\boldsymbol{\beta}), \sigma^2\right), \quad i \in I, \tag{14}$$

$$\pi(\sigma^2|\boldsymbol{\beta}, \phi, \mathbf{y}) = \text{IG}\left(\frac{n}{2} + a - 1, \frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})'\Sigma_\phi^{-1}(\mathbf{y} - X\boldsymbol{\beta})\right), \tag{15}$$

$$\pi(\phi|\boldsymbol{\beta}, \sigma^2, \mathbf{y}) \propto |\Sigma_\phi^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-X\boldsymbol{\beta})'\Sigma_\phi^{-1}(\mathbf{y}-X\boldsymbol{\beta})\right\}\pi(\phi). \tag{16}$$

Simulation from (14) and (15) is straightforward, while simulation from (16) 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, based on initial estimates of the parameters and missing values, successive sampling is done from (9), (14), (15) and (16).

Using this Gibbs sampling algorithm model (13) was fit to the observed data $\mathbf{y}_J$ based on the independence Jeffreys, Jeffreys-rule and uniform priors. The algorithm was run to obtain a sample of size 10,000 from the posterior $\pi(\beta_1, \sigma^2, \phi, \mathbf{y}_I|\mathbf{y}_J)$, and the first 1,000 iterates were discarded. Figure 6 (bottom panels) displays sample autocorrelations of the simulated values of β_1, σ^2 and ϕ based on the independence Jeffreys prior and the inefficiency factors, showing that this Markov chain also mixes well. Summaries of the marginal posterior distributions are given in Table 6 (bottom part). The three analyses produce again essentially the same results. For this dataset the analyses based on the completed and observed data also produce essentially the same results due to the small fraction of missing values (9 out of 256).

Remark 5 A caveat in the above analysis is in order. If model (13) 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 outputs of the above analyses based on the three default priors were very close to that based on a vague proper prior (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 the mean of the observations is not constant or strong spatial association is present. In this case, the independence Jeffreys prior displays 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. 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

$$X_o = XVT^{-\frac{1}{2}}. \tag{17}$$

Then it holds that $X_o'X_o = I_p$,

$$|X_o'\Sigma_\phi^{-1}X_o| = |VT^{-\frac{1}{2}}|^2|X'\Sigma_\phi^{-1}X|, \tag{18}$$

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

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

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 exist at least two orthonormal eigenvectors associated with $1-\phi\lambda_1$, say $\mathbf{c}_1$ and $\mathbf{c}_2$, satisfying

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

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 assumption that λ_1 is a simple eigenvalue of W. Finally, if $r_1(\phi) \geq \cdots \geq r_{p-1}(\phi) > 1-\phi\lambda_1 > 0$ are the eigenvalues of $X_o' \Sigma_\phi^{-1} X_o$, it follows from (18) that for all $\phi \in (0, \lambda_1^{-1})$

$$\begin{aligned} |X' \Sigma_\phi^{-1} X| &\propto |X_o' \Sigma_\phi^{-1} X_o| \\ &= (1-\phi\lambda_1) \prod_{i=1}^{p-1} v_i(\phi) \\ &= O(1-\phi\lambda_1) \quad \text{as } \phi \to \lambda_1^{-1}. \end{aligned} \tag{19}$$

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

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

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 (18) it follows that $X' \Sigma_{\lambda_1^{-1}}^{-1} X$ 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 (19) 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 $X' \Sigma_{\lambda_1^{-1}}^{-1} X\mathbf{a} = X' \Sigma_{\lambda_1^{-1}}^{-1} \mathbf{y}$. 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

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

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. □

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

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

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 $\int_{\lambda_n^{-1}}^0 \pi^{J1}(\phi)\mathrm{d}\phi$. The result now follows since t^{-1} is not integrable around 0.
(ii) From Proposition 1 and Lemma 2, and noting that $\pi^{J1}(\phi|\mathbf{y}) \propto L^I(\phi; \mathbf{y})\pi^{J1}(\phi)$

$$\begin{aligned}\int_0^{\lambda_1^{-1}} \pi^{J1}(\phi \mid \mathbf{y})\mathrm{d}\phi &\propto \begin{cases} \int_0^{\lambda_1^{-1}} (1 - \phi\lambda_1)^{-1} h(\phi)\mathrm{d}\phi & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ \int_0^{\lambda_1^{-1}} (1 - \phi\lambda_1)^{-\frac{1}{2}} h(\phi)\mathrm{d}\phi & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases} \\ &= \begin{cases} \lambda_1^{-1} \int_0^1 t^{-1}\tilde{h}(t)\mathrm{d}t & \text{if } \mathbf{u}_1 \in \mathcal{C}(X) \\ \lambda_1^{-1} \int_0^1 t^{-\frac{1}{2}}\tilde{h}(t)\mathrm{d}t & \text{if } \mathbf{u}_1 \notin \mathcal{C}(X) \end{cases}, \end{aligned} \tag{20}$$

where $h(\phi)$ and $\tilde{h}(t)$ are as in (i); a similar identity holds for $\int_{\lambda_n^{-1}}^0 \pi^{J1}(\phi \mid \mathbf{y})\mathrm{d}\phi$ 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 $(1 - t)^k$

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

where $\tilde{h}(t)$ is as in (i); a similar identity holds for $\int_{\lambda_n^{-1}}^{0} \phi^k \pi^{J1}(\phi \mid \mathbf{y})\mathrm{d}\phi$ 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 $E\{(\sigma^2)^k|\mathbf{y}\}$ exists if $E\{(\sigma^2)^k|\phi, \mathbf{y}\}$ exists and is integrable with respect to $\pi^{J1}(\phi|\mathbf{y})$. A standard calculation shows that for any prior in (3) with $a = 1$, $\pi^{J1}(\sigma^2|\phi, \mathbf{y}) = \mathrm{IG}\left(\frac{n-p}{2}, \frac{1}{2}S_\phi^2\right)$, 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), (20) and the relation analogous to (20) for $\int_{\lambda_n^{-1}}^{0} \pi^{J1}(\phi \mid \mathbf{y})\mathrm{d}\phi$.

(v) The posterior moment $E\{\beta_j^k \mid \mathbf{y}\}$ exists if $E\{\beta_j^k \mid \phi, \mathbf{y}\}$ exists and is integrable with respect to $\pi(\phi \mid \mathbf{y})$. A standard calculation shows that for any prior in (3) with $a = 1$

$$(\boldsymbol{\beta}|\phi, \mathbf{y}) \sim t_p\left(\hat{\boldsymbol{\beta}}_\phi, \frac{S_\phi^2}{n-p}(X'\Sigma_\phi^{-1}X)^{-1}, n-p\right),$$

a p-variate t distribution with location vector $\hat{\boldsymbol{\beta}}_\phi$, scale matrix $\frac{S_\phi^2}{n-p}(X'\Sigma_\phi^{-1}X)^{-1}$ and $n-p$ degrees of freedom. So for any $j = 1, \ldots, p$ it holds that

$$\left(\beta_j - (\hat{\boldsymbol{\beta}}_\phi)_j|\phi, \mathbf{y}\right) \stackrel{d}{=} \left(\frac{S_\phi^2}{n-p}(X'\Sigma_\phi^{-1}X)^{-1}_{jj}\right)^{1/2} T,$$

where $(\hat{\boldsymbol{\beta}}_\phi)_j$ is the jth component of $\hat{\boldsymbol{\beta}}_\phi$, $(X'\Sigma_\phi^{-1}X)^{-1}_{jj}$ is the jth diagonal entry of $(X'\Sigma_\phi^{-1}X)^{-1}$ and $T \sim t_1(0, 1, n-p)$. (a standard t distribution with $n-p$ degrees of freedom.) Now $E\{\beta_j^k \mid \phi, \mathbf{y}\}$ exists if and only if $E\{(\beta_j - (\hat{\boldsymbol{\beta}}_\phi)_j)^k \mid \phi, \mathbf{y}\}$ exists, in which case

$$E\{(\beta_j - (\hat{\boldsymbol{\beta}}_\phi)_j)^k|\phi, \mathbf{y}\} = \left(\frac{S_\phi^2}{n-p}(X'\Sigma_\phi^{-1}X)^{-1}_{jj}\right)^{k/2} E\{T^k\}.$$

Also $E\{T^k\}$ exists if and only if $n - p \geq k + 1$, in which case

$$E\{T^k\} = \begin{cases} 0 & \text{if } k \text{ is odd} \\ c_k & \text{if } k \text{ is even} \end{cases}, \quad \text{with} \quad c_k = \frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-p-k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-p}{2}\right)}(n-p)^{k/2}.$$

Hence for k odd it is clear that $E\{(\beta_j - (\hat{\boldsymbol{\beta}}_\phi)_j)^k|\mathbf{y}\}$ exists and equals zero. For k even only the case when $\mathbf{u}_1 \notin \mathcal{C}(X)$ needs to be considered since it is in this case that the posterior of the model parameters is proper. It was shown in the proof of Lemma 1 that in this case $X_o'\Sigma_{\lambda_1}^{-1}X_o$ is nonsingular, and so is $X'\Sigma_{\lambda_1}^{-1}X$ [see (18)]. Then $(X'\Sigma_\phi^{-1}X)^{-1} = O(1)$ as $\phi \to \lambda_1^{-1}$. From this, (7), Proposition 1 and Lemma 2 follows that $E\{(\beta_j - (\hat{\boldsymbol{\beta}}_\phi)_j)^k \mid \phi, \mathbf{y}\}\pi(\phi \mid \mathbf{y})$ is continuous on $(0, \lambda_1^{-1})$ and $O\left((1-\phi\lambda_1)^{-1/2}\right)$ as $\phi \to \lambda_1^{-1}$, and so integrable on $(0, \lambda_1^{-1})$. The integrability on $(\lambda_n^{-1}, 0)$ follows along the same lines. □

Acknowledgments I thank two anonymous referees for their helpful comments and suggestions that lead to an improved article. I also thank Mark Arnold for assistance in proving Lemma 1 and Hiroshi Kurata for comments on an earlier version of the manuscript. This work was partially supported by the US National Science Foundation Grant DMS-0719508.

References

Banerjee, S., Carlin, B., Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: Chapman & Hall.

Bapat, R. B., Raghavan, T. E. S. (1997). *Nonnegative matrices and applications*. Cambridge: Cambridge University Press.

Bell, B. S., Broemeling, L. D. (2000). A Bayesian analysis of spatial processes with application to disease mapping. *Statistics in Medicine, 19*, 957–974.

Berger, J. O., Bernardo, J. M. (1992). On the development of the reference prior method. In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian statistics* (Vol. 4, pp. 35–60). New York: Oxford University Press.

Berger, J. O., De Oliveira, V., Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association, 96*, 1361–1374.

Bernardo, J. M. (1979). Reference posterior distributions for Bayes inference (with discussion). *Journal of the Royal Statistical Society Series B, 41*, 113–147.

Besag, J., Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika, 82*, 733–746.

Buck, C. E., Cavanagh, W. G., Litton, C. D. (1988). The spatial analysis of site phosphate data. In S. P. Q. Rahtz (Ed.), *Computer applications and quantitative methods in archeology* (Vol. 446, pp. 151–160). Oxford: British Archaeological Reports, International Series.

Case, A. C., Rosen, H. S., Hines, J. R. (1993). Budget spillovers and fiscal policy interdependence. *Journal of Public Economics, 52*, 285–307.

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

Cressie, N., Kapat, P. (2008). Some diagnostics for Markov random fields. *Journal of computational and graphical statistics, 17*, 726–749.

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

Cressie, N. A. C. (1993). *Statistics for spatial data* (rev. ed.). New York: Wiley.

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

Gilks, W. R., Best, N. G., Tan, K. K. C. (1995). Adaptive rejection Metropolis sampling within Gibbs sampling. *Applied Statistics, 44*, 455–472.

Mardia, K. V., Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika, 71*, 135–146.

Militino, A. F., Ugarte, M. D., 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., 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., Lasserre, V. (1992). Spatial linear models with autocorrelated error structure. *The Statistician, 41*, 539–557.

Rue, H., Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. Boca Raton: Chapman & Hall.

Schott, J. R. (2005). *Matrix analysis for statistics* (2nd ed.). Hoboken: Wiley.

Sun, D., Tsutakawa, R. K., Speckman, P. L. (1999). Posterior distribution of hierarchical models using CAR(1) distributions. *Biometrika, 86*, 341–350.