

Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors

Matthew J. Keefe*, Marco A. R. Ferreira†, and Christopher T. Franck‡

Abstract. Bayesian hierarchical models are commonly used for modeling spatially correlated areal data. However, choosing appropriate prior distributions for the parameters in these models is necessary and sometimes challenging. In particular, an intrinsic conditional autoregressive (CAR) hierarchical component is often used to account for spatial association. Vague proper prior distributions have frequently been used for this type of model, but this requires the careful selection of suitable hyperparameters. In this paper, we derive several objective priors for the Gaussian hierarchical model with an intrinsic CAR component and discuss their properties. We show that the independence Jeffreys and Jeffreys-rule priors result in improper posterior distributions, while the reference prior results in a proper posterior distribution. We present results from a simulation study that compares frequentist properties of Bayesian procedures that use several competing priors, including the derived reference prior. We demonstrate that using the reference prior results in favorable coverage, interval length, and mean squared error. Finally, we illustrate our methodology with an application to 2012 housing foreclosure rates in the 88 counties of Ohio.

MSC 2010 subject classifications: Primary 62H11, 62F15; secondary 62M30.

Keywords: conditional autoregressive, hierarchical models, objective prior, reference prior, spatial statistics.

1 Introduction

Bayesian hierarchical models with intrinsic conditional autoregressive (CAR) priors are used for many statistical models for spatially dependent data in applications such as disease mapping (Clayton and Kaldor, 1987; Bell and Broemeling, 2000; Moraga and Lawson, 2012; Goicoa et al., 2016), image restoration (Besag et al., 1991), complex survey data (Mercer et al., 2015), and neuroimaging (Liu et al., 2016). The use of CAR specifications for modeling areal data was first introduced by Besag (1974), followed by the intrinsic CAR model as a prior for a latent spatial process proposed by Besag et al. (1991). Bayesian methods have been the dominating paradigm for spatial models including a CAR component (Sun et al., 1999; Hodges et al., 2003; Reich et al., 2006; Banerjee et al., 2014). Often in practice, it is difficult to subjectively choose informative priors for the parameters of the CAR component that are meaningful based on

*Department of Statistics, Virginia Tech, Blacksburg, VA 24061

†Department of Statistics, Virginia Tech, Blacksburg, VA 24061. Corresponding author: marf@vt.edu

‡Department of Statistics, Virginia Tech, Blacksburg, VA 24061

relevant domain knowledge. As a result, practitioners frequently use vague naïve priors (Bernardinelli et al., 1995; Best et al., 1999; Bell and Broemeling, 2000; Lee, 2013) that in the case of spatial models may lead to poor performance, such as slow Markov chain Monte Carlo (MCMC) convergence (Natarajan and McCulloch, 1998), unacceptably wide credible intervals, or larger mean squared error for estimation of parameters. To solve these problems, we introduce a novel objective prior that eliminates the need to subjectively choose priors for the parameters of the intrinsic CAR prior when no previous knowledge is available. Our objective prior can serve as an automatic prior for the popular Gaussian hierarchical model with an intrinsic CAR prior for the spatial random effects, thus enabling an automatic Bayesian analysis.

To address concerns with the use of subjective proper priors, research has been conducted to explore objective Bayesian analysis of various spatial models. In particular, Berger et al. (2001) have introduced objective priors for geostatistical models for spatially correlated data, which has been further developed by De Oliveira (2007) who has derived objective priors for a similar model that includes measurement error. The development of objective priors for CAR models for the observed data has been explored by Ferreira and De Oliveira (2007), De Oliveira (2012), and Ren and Sun (2013). While many use CAR models directly for observed data, sometimes it is preferred to use a CAR prior for spatial random effects to allow for a smoother spatial process. More recently, Ren and Sun (2014) have derived objective priors for autoregressive models incorporated into the model as a latent process. Ren and Sun (2014) have focused on the use of proper CAR priors for the latent process component. However, in practice spatial random effects are usually assigned intrinsic CAR priors (Best et al., 2005). In this work, we consider objective priors for hierarchical models that use intrinsic CAR priors for the spatial random effects.

In the spatial statistics literature, intrinsic autoregressions (or intrinsic CARs) carry the notion that they are improper “densities” that are made proper by imposing a constraint. The most frequently used constraint is a sum-zero constraint that ensures that the sum of the spatial random effects is equal to zero. For the spatial random effects, we formally define a sum-zero constrained intrinsic CAR prior which is actually a singular multivariate Gaussian distribution that is proper.

We provide further details and explanations in the following sections. In Section 2, we introduce notation and describe the Gaussian hierarchical model with unstructured random effects and intrinsic CAR spatial random effects. We also describe how to express the constrained intrinsic CAR prior as the limit of a proper CAR prior. We derive explicit expressions for the reference prior, the independence Jeffreys prior, and the Jeffreys-rule prior for this model and discuss posterior propriety in Section 3. We describe the details of the MCMC algorithm we use to simulate from the posterior distribution in Section 4. In Section 5, we present results of a simulation study to assess the frequentist properties of the Bayesian analyses using several competing priors. To illustrate and compare our proposed method to other common approaches, we conduct an analysis using the proposed reference prior and two frequently used subjective priors to model 2012 foreclosure rates in the 88 counties of Ohio in Section 6. Finally, we provide our conclusions and discussion of future work in Section 7. For clarity of

exposition, the proofs of the theoretical results are provided in the Appendix section of the Supplementary Material (Keefe et al., 2019).

2 Model Specification

2.1 Model

Consider a contiguous geographical region of interest that is partitioned into n disjoint subregions that collectively make up the entire region of interest. For example, a state could be divided into several counties. Additionally, suppose that a neighborhood structure is considered for the region of interest such that $\{N_j; j = 1, \dots, n\}$ denotes the set of subregions that are neighbors of subregion j . Typically, subregions that share a boundary are considered neighbors. Within this framework, consider the following model:

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\phi}, \quad (1)$$

where \mathbf{Y} is the $n \times 1$ vector containing the response variable, X is a $n \times p$ matrix of covariates, and $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed effect regression coefficients. Furthermore, we assume that $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_n)^T$ is a vector of unstructured random effects defined such that θ_i are independent and normally distributed with mean 0 and variance σ^2 for $i = 1, 2, \dots, n$. Finally, $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_n)^T$ is a vector of spatial random effects that is assigned an intrinsic CAR prior (Besag and Kooperberg, 1995), with the sum-zero constraint $\sum_{i=1}^n \phi_i = 0$. Additionally, we assume that $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are independent a priori.

In order to demonstrate how to obtain the sum-zero constrained intrinsic CAR prior, we first provide background information on how intrinsic CAR models are typically defined in Section 2.2. Then, we describe a formal procedure to obtain the sum-zero constrained intrinsic CAR prior in Section 2.3 that is used in our model formulation.

2.2 Background on Intrinsic CAR Models

To distinguish our formal specification of the intrinsic CAR spatial random effects from the traditional specification, we reserve $\boldsymbol{\phi}$ for our formal definition and use $\boldsymbol{\omega}$ to provide background information on intrinsic CAR models as they are frequently used in practice. Typically, intrinsic CAR models are defined by their conditional distributions, which are proper distributions. Consider the frequently used intrinsic CAR model for $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^T$ specified by its conditional distributions

$$p(\omega_i | \boldsymbol{\omega}_{\sim i}) \propto \exp \left\{ -\frac{\tau_\omega}{2} \left[\sum_{i=1}^n \omega_i^2 h_i - 2 \sum_{i < j} \omega_i \omega_j g_{ij} \right] \right\}, \quad (2)$$

where $\boldsymbol{\omega}_{\sim i}$ is the vector of the CAR elements for all subregions except subregion i and $\tau_\omega > 0$ is a precision parameter. In addition, $g_{ij} \geq 0$ is a measure of how similar subregions i and j are, $g_{ij} = g_{ji}$, and $h_i = \sum_{j=1}^n g_{ij}$. Alternatively, we may write the joint density for $\boldsymbol{\omega}$ as

$$p(\boldsymbol{\omega}) \propto \exp \left\{ -\frac{\tau_\omega}{2} \boldsymbol{\omega}^T H \boldsymbol{\omega} \right\}, \quad (3)$$

where H is a symmetric, positive semi-definite precision matrix defined as

$$(H)_{ij} = \begin{cases} h_i, & \text{if } i = j, \\ -g_{ij}, & \text{if } i \in N_j, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The matrix H is assumed to be fixed and known, as its structure is typically chosen as a modeling decision. One common choice for the similarity measure is $g_{ij} = 1$ if subregions i and j are neighbors, and $g_{ij} = 0$ if subregions i and j are not neighbors. This binary similarity measure implies that h_i is the number of neighbors of subregion i . Note that (3) is a multivariate normal kernel specified by its precision matrix $\tau_\omega H$. Furthermore, note that the matrix H is singular with one eigenvalue equal to zero and corresponding eigenvector $n^{-1/2}\mathbf{1}_n$, where $\mathbf{1}_n$ denotes a $n \times 1$ vector of ones. As a consequence, the density given in (3) does not change if we substitute $\boldsymbol{\omega}$ by $\boldsymbol{\omega} + a\mathbf{1}_n$, where a is any real constant. Because of that, (3) is an improper “density”. By improper, we mean that the integral of the “density” with respect to $\boldsymbol{\omega}$ diverges to ∞ . To make this density proper, practitioners usually assume the sum-zero constraint $\sum_{i=1}^n \omega_i = 0$. This constraint is frequently imposed in an informal manner with the vector of spatial random effects being recentered around its own mean at the end of each MCMC iteration. While ingenious, this mathematically informal way to impose the sum-zero constraint obscures the actual joint prior density of $\boldsymbol{\omega}$ under the constraint. Consequently, a hierarchical model with spatial random effects defined by (3) with the sum-zero constraint $\sum_{i=1}^n \omega_i = 0$ imposed within the MCMC algorithm does not yield itself to a formal objective Bayes analysis.

To enable a formal objective Bayes analysis, in Section 2.3 we consider an intrinsic CAR expressed as the limit of a proper CAR, rather than starting with an intrinsic CAR. We impose the sum-zero constraint before taking the limit. The resulting joint density for the spatial random effects after taking the limit is proper.

2.3 Intrinsic CAR as the Limit of a Proper CAR

In the following development, we consider the following notation to clearly demonstrate the transition between different CAR models. First, we consider a proper CAR for the random vector $\boldsymbol{\phi}^{**}$, then impose the sum-zero constraint to obtain the distribution of the random vector $\boldsymbol{\phi}^*$, and finally consider the limiting case that leads to the sum-zero constrained intrinsic CAR for the random vector $\boldsymbol{\phi}$. The intrinsic CAR prior is a limiting case of a proper CAR prior. Specifically, we use a signal-to-noise ratio parametrization to express the proper CAR prior as

$$\boldsymbol{\phi}^{**} \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c} \Sigma_\lambda\right), \quad (5)$$

where σ^2 and $\tau_c > 0$ are unknown parameters, $\lambda > 0$ is a propriety parameter, and $\Sigma_\lambda^{-1} = \lambda I_n + H$ with I_n being the $n \times n$ identity matrix and H is defined as in (4).

Note that rather than using a CAR model directly for the data, we consider a hierarchical model with both conditional autoregressive spatial random effects and unstructured random effects. Because of this, we use a signal-to-noise ratio parametrization of the CAR prior where the variance of the spatial random effects is expressed as a

function of the variance of the unstructured random effects, as seen in expression (5). When $\lambda \rightarrow 0$, ϕ^{**} in (5) approaches an intrinsic CAR (Besag et al., 1991; Besag and Kooperberg, 1995). In practice, the intrinsic CAR spatial random effects are constrained to sum to zero to ensure propriety of the prior. This is often performed within the posterior sampling algorithm by centering the sampled random effects after each iteration. Rather than taking this approach, we first consider the proper CAR in (5), then we impose the sum-zero constraint, and finally we take the limit as $\lambda \rightarrow 0$ to obtain a constrained intrinsic CAR.

To impose the sum-zero constraint for the CAR prior in (5), we project ϕ^{**} onto the subspace of \mathbb{R}^n that is orthogonal to the subspace spanned by the vector $n^{-1/2}\mathbf{1}_n$ to obtain ϕ^* . By construction, $\sum_{i=1}^n \phi_i^* = 0$. Specifically, we define $\phi^* = P\phi^{**}$, where $P = (I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^T)$ is a centering (projection) matrix. After imposing the constraint, the distribution of ϕ^* is given by

$$\phi^* \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c} \Sigma_{\phi\lambda}\right), \tag{6}$$

where $\Sigma_{\phi\lambda} = P\Sigma_\lambda P^T$.

Now, suppose that the spectral decomposition of H is given by $H = QDQ^T$ where $Q = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n)$ is a $n \times n$ matrix comprised of columns which are the normalized eigenvectors of H and $D = \text{diag}(d_1, d_2, \dots, d_n)$ where $d_1 \geq d_2 \geq \dots \geq d_n$ are the ordered eigenvalues of H . Additionally, if all of the subregions that completely partition the region of interest are connected (i.e. there is a path connecting any two subregions), then H has rank $n - 1$. Consequently, 0 is an eigenvalue of H with multiplicity 1 (e.g. see De Oliveira and Ferreira (2011)). As a result, $d_{n-1} > d_n = 0$ and d_n has a corresponding eigenvector $\mathbf{q}_n = n^{-1/2}\mathbf{1}_n$.

Lemma 1. *If all of the subregions that completely partition the region of interest are connected, then $\Sigma_{\phi\lambda}$ has rank $n - 1$ and the spectral decomposition of $\Sigma_{\phi\lambda}$ is $\Sigma_{\phi\lambda} = QM_\lambda Q^T$, where $M_\lambda = \text{diag}((\lambda + d_1)^{-1}, \dots, (\lambda + d_{n-1})^{-1}, 0)$. Finally, $\Sigma_{\phi\lambda}$ can be written as*

$$\Sigma_{\phi\lambda} = \sum_{i=1}^{n-1} \frac{1}{\lambda + d_i} \mathbf{q}_i \mathbf{q}_i^T. \tag{7}$$

Proof. See Appendix in Supplementary Material □

Now we take the limit as $\lambda \rightarrow 0$ to obtain the final density for ϕ which appears in our model from (1) given by

$$\phi \sim N\left(\mathbf{0}, \frac{\sigma^2}{\tau_c} \Sigma_\phi\right), \tag{8}$$

where $\Sigma_\phi = \lim_{\lambda \rightarrow 0} \Sigma_{\phi\lambda} = QMQ^T$ and $M = \text{diag}(d_1^{-1}, \dots, d_{n-1}^{-1}, 0)$. Hence, $\Sigma_\phi = H^+$ is the Moore–Penrose pseudoinverse of H (Penrose, 1955). Thus, the vector of spatial random effects ϕ has a singular Gaussian distribution (Muirhead, 2009; Ferreira et al., 2011). Further, we note that the integral of the corresponding density with respect to

ϕ is equal to one. Hence, this singular Gaussian distribution is proper. In addition, Ferreira (2018) shows that the singular Gaussian distribution in (2.8) is the limiting distribution of a one-at-a-time Gibbs sampler applied to the intrinsic CAR prior in (2.2) with centering on the fly. Therefore, our reference prior is directly applicable to the Gaussian hierarchical models with intrinsic CAR priors widely used in practice.

A possible concern raised by Lavine and Hodges (2012) is that there are several ways of approaching an improper intrinsic CAR as the limit of a proper CAR. Without imposing the sum-zero constraint, Lavine and Hodges (2012) showed that different ways to approach the limit may lead to different functions of τ_ω appearing in the constant of proportionality in (2.3). As a consequence, the likelihood function for τ_ω would not be well-defined and formal Bayesian inference would not be possible. To address this issue, Keefe et al. (2018) have shown that different ways of approaching the limit are equivalent to using different initial proper CAR models. In particular, Keefe et al. (2018) consider proper CAR models given by (2.5) but substituting the identity matrix I_n by a matrix K to obtain the precision matrix $\Sigma_\lambda^{-1} = \lambda K + H$. Keefe et al. (2018) consider for K the class of symmetric positive semidefinite matrices such that the sum of its elements is positive. This construction can represent two of the ways of approaching the limit considered by Lavine and Hodges (2012). Keefe et al. (2018) have shown that for any matrix K in this class, if the sum-zero constraint is formally imposed before taking the limit, then sum-zero constrained intrinsic CAR model does not depend on K and is, therefore, unique. As a consequence, for any such matrix K , the sum-zero constrained intrinsic CAR density has a well-defined constant of proportionality that is a unique function of τ_c . Therefore, for the hierarchical model with the sum-zero constrained intrinsic CAR random effects we consider, the likelihood function is well-defined. The resulting sum-zero constrained intrinsic CAR prior for ϕ in (2.8) is a singular Gaussian distribution, which is amenable for an objective Bayes analysis.

Thus, it follows that the model for the response \mathbf{Y} with the spatial random effects ϕ integrated out is

$$(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \tau_c) \sim N(X\boldsymbol{\beta}, \sigma^2(I_n + \tau_c^{-1}\Sigma_\phi)). \quad (9)$$

2.4 Remarks

Note that we assume a signal-to-noise ratio parametrization for the variance components of the random components of the model, as is done in De Oliveira (2007). One of the benefits of this parametrization is that it leads to simpler expressions for the objective priors. The unknown model parameters for this hierarchical spatial model for areal data are $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma^2, \tau_c) \in \mathbb{R}^p \times (0, \infty) \times (0, \infty)$. Of particular interest is the parameter τ_c , which controls the strength of spatial dependence. Conditional on the neighborhood structure chosen by the user, Σ_ϕ is fixed and the correlation between the response for two neighboring subregions Y_i and Y_j is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\tau_c^2 + \tau_c\sigma_{ii} + \tau_c\sigma_{jj} + \sigma_{ii}\sigma_{jj}}}, \quad (10)$$

where σ_{ij} denotes the element of Σ_ϕ located in the i^{th} row and j^{th} column and σ_{ii} and σ_{jj} denote the i^{th} and j^{th} diagonal elements of Σ_ϕ , respectively. Since ρ_{ij} is a decreasing function of τ_c , as $\tau_c \rightarrow 0$, ρ_{ij} increases implying that the spatial dependence also increases. As is commonly the case in stochastic processes, an increase in correlation is often accompanied by an increase in marginal variance. Take for example the first-order autoregressive model (AR(1)) with error variance δ^2 and autoregressive coefficient $0 < \rho < 1$. The correlation function for lag h is $\text{Corr}(x_t, x_{t-h}) = \rho^h$ and the marginal variance is $\delta^2/(1-\rho^2)$. Therefore, for the widely used AR(1) time series model, increase in ρ corresponds to increase in both correlation and marginal variance. Note that in our model, $\text{Var}(y_i) = \sigma^2(1 + \tau_c^{-1}\sigma_{ii})$. So, it is also the case for our model that an increase in correlation is accompanied by an increase in marginal variance.

Specifying the model using a sum-zero constrained intrinsic CAR simplifies application of the results of De Oliveira (2007) to derive several objective priors. Although these priors take on similar forms as those presented by De Oliveira (2007), the properties of the derived priors differ in our case. The reason for the difference is that for the geostatistical models considered by De Oliveira (2007), the matrix Σ_ϕ is full rank, whereas we consider a hierarchical model for areal data where Σ_ϕ is singular.

3 Objective Priors

3.1 Likelihood Functions

In order to obtain objective priors, such as the reference, independence Jeffreys, and Jeffreys-rule priors for this model, we consider the likelihood and integrated likelihood functions. The likelihood of $\boldsymbol{\eta}$ given the observed response \mathbf{y} and the matrix X is given by

$$L(\boldsymbol{\eta}; \mathbf{y}, X) \propto (\sigma^2)^{-n/2} |\Omega|^{-1/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - X\boldsymbol{\beta})^T \Omega^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \right], \quad (11)$$

where $\Omega = I_n + \frac{1}{\tau_c} \Sigma_\phi$ and $|A|$ denotes the determinant of the matrix A. All of the objective priors we derive with respect to the integrated likelihoods in the following sections conveniently fall into a class of priors of the form

$$\pi(\boldsymbol{\eta}) \propto \frac{\pi(\tau_c)}{(\sigma^2)^a}, \quad (12)$$

where $a \in \mathbb{R}$ is a hyperparameter and $\pi(\tau_c)$ is referred to as the marginal prior for τ_c .

In order to obtain the objective priors, it is first necessary to obtain the integrated likelihoods. If we first integrate $L(\boldsymbol{\eta}; \mathbf{y}, X)$ with respect to $\boldsymbol{\beta}$, we obtain the integrated likelihood of (σ^2, τ_c) . This integration leads to the log-integrated likelihood given by

$$\log L^I(\sigma^2, \tau_c; \mathbf{y}, X) = -\frac{1}{2} \left[(n-p) \log(\sigma^2) + \log(|\Omega|) + \log(|X^T \Omega^{-1} X|) + \frac{S^2}{\sigma^2} \right], \quad (13)$$

where $S^2 = \mathbf{y}^T [\Omega^{-1} - \Omega^{-1} X (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1}] \mathbf{y}$. Furthermore, obtaining the integral of $L^I(\sigma^2, \tau_c; \mathbf{y}, X) (\sigma^2)^{-a}$ with respect to σ^2 yields the integrated likelihood of τ_c , which is given by

$$L^I(\tau_c; \mathbf{y}, X) \propto (|\Omega| |X^T \Omega^{-1} X|)^{-1/2} (S^2)^{-\left(\frac{n-p}{2} + a - 1\right)}. \quad (14)$$

Unfortunately, it is difficult to calculate the reference, independence Jeffreys, and Jeffreys-rule priors directly using expressions (13) and (14). To overcome this challenge, it is particularly useful to express (13) in a simplified form involving the eigenvalue decompositions of matrices that are functions of the matrices X and Σ_ϕ . Consider the following lemma (Verbyla, 1990; Dietrich, 1991; Kuo, 1999; De Oliveira, 2007):

Lemma 2. *Suppose $X_{n \times p}$ is of full rank, with $n > p$, and Σ is an $n \times n$ symmetric positive definite matrix. Then there exists a full rank $n \times (n - p)$ matrix L with the following properties:*

- (i) $L^T X = 0$
- (ii) $L^T L = I_{n-p}$
- (iii) $\Sigma^{-1} - \Sigma^{-1} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} = L (L^T \Sigma L)^{-1} L^T$
- (iv) $\log(|\Sigma|) + \log(|X^T \Sigma^{-1} X|) = \log(|L^T \Sigma L|) + c$, where c depends on X but not on Σ
- (v) $L^T \Sigma L$ is an $(n - p) \times (n - p)$ diagonal matrix with positive diagonal elements.

Here we propose one way to obtain a matrix L with the above properties by the following steps:

1. Obtain the orthogonal complement of the matrix $G = X(X^T X)^{-1} X^T$ given by $G^* = I_n - X(X^T X)^{-1} X^T$.
2. Calculate the matrix Q^* such that the columns of Q^* are the normalized eigenvectors corresponding to the non-zero eigenvalues of G^* .
3. Denote the spectral decomposition of $Q^{*T} \Sigma_\phi Q^*$ as $U \Psi U^T$.
4. Compute $L = Q^* U$.

The resulting L matrix has properties that allow for convenient simplifications in the integrated likelihoods. Let $\xi_1 \geq \xi_2 \geq \dots \geq \xi_{n-p} > 0$ be the ordered eigenvalues of $Q^{*T} \Sigma_\phi Q^*$. Then, by the way we obtain the matrix L , we have that $L^T \Sigma_\phi L = \Psi = \text{diag}(\xi_1, \xi_2, \dots, \xi_{n-p})$. Then the following results hold (De Oliveira, 2007)

$$\log(|\Omega|) + \log(|X^T \Omega^{-1} X|) = \sum_{j=1}^{n-p} \log\left(1 + \frac{1}{\tau_c} \xi_j\right) + c, \quad (15)$$

$$S^2 \stackrel{d}{=} \sigma_*^2 \sum_{j=1}^{n-p} \left(\frac{1 + \tau_{c*}^{-1} \xi_j}{1 + \tau_c^{-1} \xi_j} \right) Z_j^2, \quad (16)$$

where σ_*^2 and τ_{c*} are the true values of σ^2 and τ_c , respectively, and $\{Z_j^2\} \stackrel{iid}{\sim} \chi_1^2$. Using these results, the log-integrated likelihood in (13) satisfies

$$\begin{aligned} & \log L^I(\sigma^2, \tau_c; \mathbf{y}, X) \\ & \stackrel{d}{=} -\frac{1}{2} \left[(n-p) \log(\sigma^2) + \sum_{j=1}^{n-p} \left\{ \log\left(1 + \frac{1}{\tau_c} \xi_j\right) + \frac{\sigma_*^2}{\sigma^2} \left(\frac{1 + \tau_{c*}^{-1} \xi_j}{1 + \tau_c^{-1} \xi_j} \right) Z_j^2 \right\} \right]. \end{aligned} \quad (17)$$

The posterior distribution of $\boldsymbol{\eta}$ using a prior of the form given in (12) is proper if and only if

$$0 < \int_0^\infty L^I(\tau_c; \mathbf{y}, X) \pi(\tau_c) d\tau_c < \infty. \quad (18)$$

Proposition 1. Consider the model given in (9) along with the prior given in (12). Then, $L^I(\tau_c; \mathbf{y}, X)$ is a continuous function on $(0, \infty)$ and $L^I(\tau_c; \mathbf{y}, X) = O(\tau_c^{1-a})$ as $\tau_c \rightarrow 0$ and $L^I(\tau_c; \mathbf{y}, X) = O(1)$ as $\tau_c \rightarrow \infty$.

Proof. See Appendix in Supplementary Material □

The condition in (18) and the tail behavior of the integrated likelihood in Proposition 1 provide justification for determining whether or not the proposed priors lead to proper posterior distributions. For example, the tail behavior of the posterior distribution as $\tau_c \rightarrow 0$ must be $O(\tau_c^m)$ where $m > -1$ in order to guarantee posterior propriety. Likewise, the tail behavior of the posterior distribution as $\tau_c \rightarrow \infty$ must be $O(\tau_c^m)$ where $m < -1$ in order to guarantee posterior propriety.

3.2 Reference Prior

In order to obtain a reference prior for the model parameters $\boldsymbol{\eta}$, it is necessary to identify an order of importance for the parameters. Here, we consider $\boldsymbol{\beta}$ to be a nuisance parameter, while (σ^2, τ_c) is the parameter vector of interest. Then, we use exact marginalization to find the reference prior (Berger et al., 2001). First, the joint prior distribution of $\boldsymbol{\eta}$ is factored as $\pi^R(\boldsymbol{\eta}) = \pi^R(\boldsymbol{\beta} | \sigma^2, \tau_c) \pi^R(\sigma^2, \tau_c)$. Then given (σ^2, τ_c) , the conditional reference prior for $\boldsymbol{\beta}$ is $\pi^R(\boldsymbol{\beta} | \sigma^2, \tau_c) \propto 1$ (Bernardo and Smith, 1994). Finally, we use the Jeffreys-rule algorithm on the integrated likelihood $L^I(\sigma^2, \tau_c; \mathbf{y}, X)$ to obtain $\pi^R(\sigma^2, \tau_c)$.

Theorem 1. Consider the model given in (9). Then, the reference prior of $\boldsymbol{\eta}$ is of the form (12) with

$$a = 1 \text{ and } \pi^R(\tau_c) \propto \frac{1}{\tau_c} \left[\sum_{j=1}^{n-p} \left(\frac{\xi_j}{\tau_c + \xi_j} \right)^2 - \frac{1}{n-p} \left\{ \sum_{j=1}^{n-p} \left(\frac{\xi_j}{\tau_c + \xi_j} \right) \right\}^2 \right]^{1/2}. \quad (19)$$

Proof. See Appendix in Supplementary Material □

The following proposition and corollary describe the properties of the proposed reference prior and its resulting posterior distribution.

Proposition 2. *Suppose that $\xi_1, \xi_2, \dots, \xi_{n-p}$ are not all equal. Then, the marginal reference prior of τ_c in (19) is a continuous function on $(0, \infty)$ where:*

- (i) $\pi^R(\tau_c) = O(1)$ as $\tau_c \rightarrow 0$
- (ii) $\pi^R(\tau_c) = O(\tau_c^{-2})$ as $\tau_c \rightarrow \infty$.

Proof. See Appendix in Supplementary Material □

Corollary 1. *Using the model given by (9), we have*

- (i) *The reference prior $\pi^R(\tau_c)$ given in (19) and its resulting posterior $\pi^R(\boldsymbol{\eta}|\mathbf{y}, X)$ are both proper.*
- (ii) *The k^{th} moment of the marginal reference posterior $\pi^R(\tau_c|\mathbf{y}, X)$ does not exist for $k \geq 1$.*

Proof. See Appendix in Supplementary Material □

Notice that the reference prior depends on the covariates X through the eigenvalues $\xi_1, \xi_2, \dots, \xi_{n-p}$ of $L^T \Sigma_\phi L$. We also note that since our proposed reference prior has the same mathematical expression as that of De Oliveira (2007), the results regarding the marginal reference prior, the joint posterior distribution, and the marginal posterior distribution are all the same. However, the same is not true for the independence Jeffreys and Jeffreys-rule priors that follow. This is due to the fact that, in our case, $L^T \Sigma_\phi L$ is full rank, while Σ_ϕ is not full rank.

3.3 Independence Jeffreys Prior

The independence Jeffreys prior is obtained by assuming that $\boldsymbol{\beta}$ and (σ^2, τ_c) are independent a priori. Then, the Jeffreys rule is used to find the marginal prior of each parameter, assuming the other parameters are known. We denote the ordered eigenvalues of Σ_ϕ as $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_{n-1} > \gamma_n = 0$. It is important to note that in our case, we are dealing with areal data, and thus $\gamma_n = 0$ since Σ_ϕ is of rank $n - 1$. This is a key distinction between our work and that of De Oliveira (2007), which leads to different results regarding posterior propriety.

Theorem 2. *Consider the model given in (9). Then, the independence Jeffreys prior of $\boldsymbol{\eta}$ is of the form (12) with*

$$a = 1 \text{ and } \pi^{J1}(\tau_c) \propto \frac{1}{\tau_c} \left[\sum_{i=1}^n \left(\frac{\gamma_i}{\tau_c + \gamma_i} \right)^2 - \frac{1}{n} \left\{ \sum_{i=1}^n \left(\frac{\gamma_i}{\tau_c + \gamma_i} \right) \right\}^2 \right]^{1/2}. \quad (20)$$

Proof. See Appendix in Supplementary Material □

Proposition 3. *The marginal independence Jeffreys prior of τ_c in (20) is a continuous function on $(0, \infty)$ where $\pi^{J1}(\tau_c) = O(\tau_c^{-1})$ as $\tau_c \rightarrow 0$.*

Proof. See Appendix in Supplementary Material □

Corollary 2. *Using the model given by (9), the independence Jeffreys prior $\pi^{J1}(\tau_c)$ given in (20) and its resulting posterior $\pi^{J1}(\boldsymbol{\eta}|\mathbf{y}, X)$ are both improper.*

Proof. See Appendix in Supplementary Material □

3.4 Jeffreys-Rule Prior

The Jeffreys-rule prior is obtained by $\pi^{J2}(\boldsymbol{\eta}) \propto |I(\boldsymbol{\eta})|^{1/2}$, where $I(\boldsymbol{\eta})$ is the Fisher information matrix of $\boldsymbol{\eta}$.

Theorem 3. *Consider the model given in (9). Then, the Jeffreys-rule prior of $\boldsymbol{\eta}$, is of the form (12) with*

$$a = 1 + \frac{p}{2} \text{ and } \pi^{J2}(\tau_c) \propto \left(\frac{\prod_{j=1}^{n-p} (1 + \tau_c^{-1} \xi_j)}{\prod_{i=1}^n (1 + \tau_c^{-1} \gamma_i)} \right)^{1/2} \cdot \pi^{J1}(\tau_c). \quad (21)$$

Proof. See Appendix in Supplementary Material □

Proposition 4. *The marginal Jeffreys-rule prior of τ_c in (21) is a continuous function on $(0, \infty)$ where $\pi^{J2}(\tau_c) = O(\tau_c^{\frac{p-3}{2}})$ as $\tau_c \rightarrow 0$.*

Proof. See Appendix in Supplementary Material □

Corollary 3. *Using the model given by (9), the posterior distribution obtained by using the Jeffreys-rule prior $\pi^{J2}(\tau_c)$ given in (21) is improper.*

Proof. See Appendix in Supplementary Material □

Similar to the reference prior, notice that the Jeffreys-rule prior also depends on the covariates X through the eigenvalues $\xi_1, \xi_2, \dots, \xi_{n-p}$ of $L^T \Sigma_\phi L$.

3.5 Remarks

There are a few key distinctions between our results and those of other researchers. Specifically, our reference prior leads to a proper posterior distribution, while the independence Jeffreys and Jeffreys-rule priors lead to improper posterior distributions, unlike De Oliveira (2007). Although we consider areal data while De Oliveira (2007) has considered geostatistical data, the reference prior in (19) has the same mathematical expression as that proposed by De Oliveira (2007) because we choose to formulate our model with the intrinsic CAR component as the limit of a proper CAR. However, the matrix Σ_ϕ is not full rank in the areal data setting. Thus, the smallest eigenvalue of Σ_ϕ is $\gamma_n = 0$, which leads to improper posterior distributions when the independence Jeffreys and Jeffreys-rule priors are used.

Ren and Sun (2014) have defined their proper CAR prior using *conditional distributions* with a constant variance that does not depend on the number of neighbors a subregion has, which limits utility for many applications. In contrast, the conditional specification of the CAR prior we use has a variance that depends on the number of neighboring subregions. This allows the conditional distribution of ϕ_i given $\phi_{\sim i}$ for subregions with more neighbors to have a smaller variance. Furthermore, the objective priors that have been derived by Ren and Sun (2014) do not have explicit mathematical forms. Specifically, the priors of Ren and Sun (2014) are written as functions of determinants of Fisher information matrices that do not resemble our reference prior. In contrast, the expression of our reference prior can be easily implemented with its explicit mathematical form.

4 Sampling from the Posterior Distribution

In order to perform posterior inference about the model parameters, we propose a MCMC algorithm. Specifically, we have implemented a MCMC sampler (Gelfand and Smith, 1990) that considers the integrated likelihood given in (9) obtained by integrating out the spatial random effects ϕ . Our MCMC sampler is a Metropolis-within-Gibbs algorithm that includes a Gibbs step for β and a joint Metropolis-Hastings step for τ_c and σ^2 . Additionally, we simulate ϕ using composite sampling.

The use of the signal-to-noise ratio parametrization results in a Fisher information matrix of $\log(\sigma^2)$ and $\log(\tau_c)$ which does not depend on the value of σ^2 . This implies a uniform prior for $\log(\sigma^2)$. Thus, in sampling from the posterior distribution, we jointly propose values for $\log(\sigma^2)$ and $\log(\tau_c)$ with a Metropolis-Hastings step using Normal proposal distributions, and accept proposed values based on the appropriate Metropolis-Hastings acceptance probability (Gamerman and Lopes, 2006; Robert and Casella, 2004). Posterior inference can then be conducted directly from the MCMC samples after discarding an appropriate number of iterations as burn-in. While we choose to use a MCMC sampler, there are numerous alternative methods that could be used to obtain samples from the posterior distribution. For example, De Oliveira (2007) has implemented adaptive rejection Metropolis sampling (Gilks et al., 1995). The steps of our MCMC algorithm are provided in Algorithm 1.

Since the matrix Σ_ϕ can be conveniently expressed as a function of eigenvectors and eigenvalues of H , as shown in (7), several computational equivalences can be used to speed up the MCMC algorithm. Specifically, execution of the MCMC algorithm relies on computation of the determinant and inverse of the matrix $(I + \tau_c^{-1}\Sigma_\phi)$. In this work, these quantities can be expressed by the following equivalences

$$I + \tau_c^{-1}\Sigma_\phi = Q \text{diag} \left(1 + \frac{\tau_c^{-1}}{\lambda + d_1}, \dots, 1 + \frac{\tau_c^{-1}}{\lambda + d_{n-1}}, 1 \right) Q^T, \quad (22)$$

$$(I + \tau_c^{-1}\Sigma_\phi)^{-1} = Q \text{diag} \left(\left[1 + \frac{\tau_c^{-1}}{\lambda + d_1} \right]^{-1}, \dots, \left[1 + \frac{\tau_c^{-1}}{\lambda + d_{n-1}} \right]^{-1}, 1 \right) Q^T, \quad (23)$$

$$|\sigma^2(I + \tau_c^{-1}\Sigma_\phi)| = (\sigma^2)^n \prod_{i=1}^{n-1} \left(1 + \frac{\tau_c^{-1}}{\lambda + d_i} \right). \quad (24)$$

Algorithm 1 MCMC Algorithm

1. Initialize $\eta^{(0)} = (\beta^{(0)}, \sigma^{2(0)}, \tau_c^{(0)})$

For i in 1 to K

{

2. Generate $\log(\sigma^{2*}) \sim N(\sigma^{2(i-1)}, \delta_s)$ and $\log(\tau_c^*) \sim N(\tau_c^{(i-1)}, \delta_t)$
3. Compute joint acceptance probability for σ^{2*} and τ_c^* :

$$\alpha = \min \left[1, \frac{P(\boldsymbol{\eta}^* | \text{Data}) q(\boldsymbol{\eta} | \boldsymbol{\eta}^*)}{P(\boldsymbol{\eta}^{(i)} | \text{Data}) q(\boldsymbol{\eta}^* | \boldsymbol{\eta})} \right]$$

4. Generate $(\beta^* | \mathbf{y}, \sigma^2, \tau_c, X) \sim N_p(\mu^*, \Sigma^*)$, where

$$\begin{aligned} \mu^* &= \left(X^T (I_n + \tau_c^{-1(i-1)} \Sigma_\phi)^{-1} X \right)^{-1} X^T (I_n + \tau_c^{-1(i-1)} \Sigma_\phi)^{-1} \mathbf{y}, \\ \Sigma^* &= \sigma^{2(i-1)} \left(X^T (I_n + \tau_c^{-1(i-1)} \Sigma_\phi)^{-1} X \right)^{-1}. \end{aligned}$$

}

5. Use composite sampling to generate $\phi^{(i)}$ ($i = 1, \dots, K$) from its full conditional distribution which is provided in the Appendix in the Supplementary Material.
-

5 Simulation Study

5.1 Comparison of Priors

We use Monte Carlo simulation to compare the performance of the proposed reference prior with two commonly used vague naïve prior distributions from the literature. Because the independence Jeffreys and Jeffreys-rule priors lead to improper posterior distributions, we have not included them in this simulation study. Performance is as-

sessed using frequentist properties of Bayesian procedures, including interval coverage rate, average interval length (IL), and mean squared error (MSE).

The vague naïve prior distributions we consider come from the `CARBayes` R package (Lee, 2013) and Best et al. (1999). Both of these approaches assume gamma(shape, rate) prior distributions for the precisions of both unstructured and spatial random effects. Specifically, the `CARBayes` package (version 4.0) has implemented gamma(0.001, 0.001) prior distributions as the default for both precision parameters. Best et al. (1999) has used gamma(0.001, 0.001) and gamma(0.1, 0.1) prior distributions for the precisions of the unstructured and spatial random effects, respectively. Previously, the `CARBayes` package used uniform prior distributions which are no longer available. We adopt the prior distributions implied by the `CARBayes` package and Best et al. (1999) under our signal-to-noise ratio parametrization. For the purpose of reporting, we refer to these methods as *CARBayes* and *NB* (lead author initials), respectively.

Since we have adopted a signal-to-noise ratio parametrization in this work, the gamma(α_1, β_1) and gamma(α_2, β_2) prior distributions for the precisions of the unstructured and spatial random effects, respectively, are re-parametrized. The implied marginal prior for τ_c under the signal-to-noise ratio parametrization is given by

$$\pi^V(\tau_c) = \frac{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} (\beta_1 + \beta_2 \tau_c)^{-(\alpha_1 + \alpha_2)} \tau_c^{\alpha_2 - 1}, \quad (25)$$

where $\Gamma(\cdot)$ is the Gamma function.

When comparing priors, it is important to consider not only the regions in the parameter space that receive the most mass, but also the rate of decay of the tail of each prior. It is clear from (25) that with regard to tail behavior, as $\tau_c \rightarrow \infty$, $\pi^V(\tau_c) = O(\tau_c^{-\alpha_1 - 1})$ and as $\tau_c \rightarrow 0$, $\pi^V(\tau_c) = O(\tau_c^{\alpha_2 - 1})$. Thus, for values of α_1 and α_2 close to 0, the vague naïve prior is close to impropriety. From Proposition 2, we can see that the tail behavior of the reference prior is considerably different than that of the `CARBayes` and `NB` priors. Consequently, the slower decay rate as $\tau_c \rightarrow \infty$ implies that the `CARBayes` and `NB` priors place significantly more mass on large values of τ_c . Figure 1(a) provides a plot of each of the prior distributions for τ_c that we consider. The plot of $\log \pi(\tau_c)$ versus $\log_{10} \tau_c$ provided in Figure 1(b) shows that the reference prior places more mass on values of τ_c between 0.01 and 56 (`CARBayes`) and 0.02 and 35 (`NB`), while the `CARBayes` and `NB` priors place significantly more mass on large values of τ_c .

In addition, the `CARBayes` and `NB` priors put more mass on very small values of τ_c . Our practical experience indicates that real spatial datasets will usually have estimated values of τ_c between 0.1 and 10. For example, in the application considered in Section 6 the estimated values of τ_c for the three considered priors range from about 0.15 to about 0.25. Further, because we use a signal-to-noise ratio parametrization, values of τ_c much less than 0.01 are not of practical use. Note that $\tau_c < 0.01$ implies that the conditional variance of the structured random effects is greater than 100 times the variance of the unstructured random effects. In that case, the analyst should consider a model that does not include unstructured random effects. Hence, in the rare case when for a given dataset the estimate of τ_c is less than 0.01, then instead of the hierarchical model given

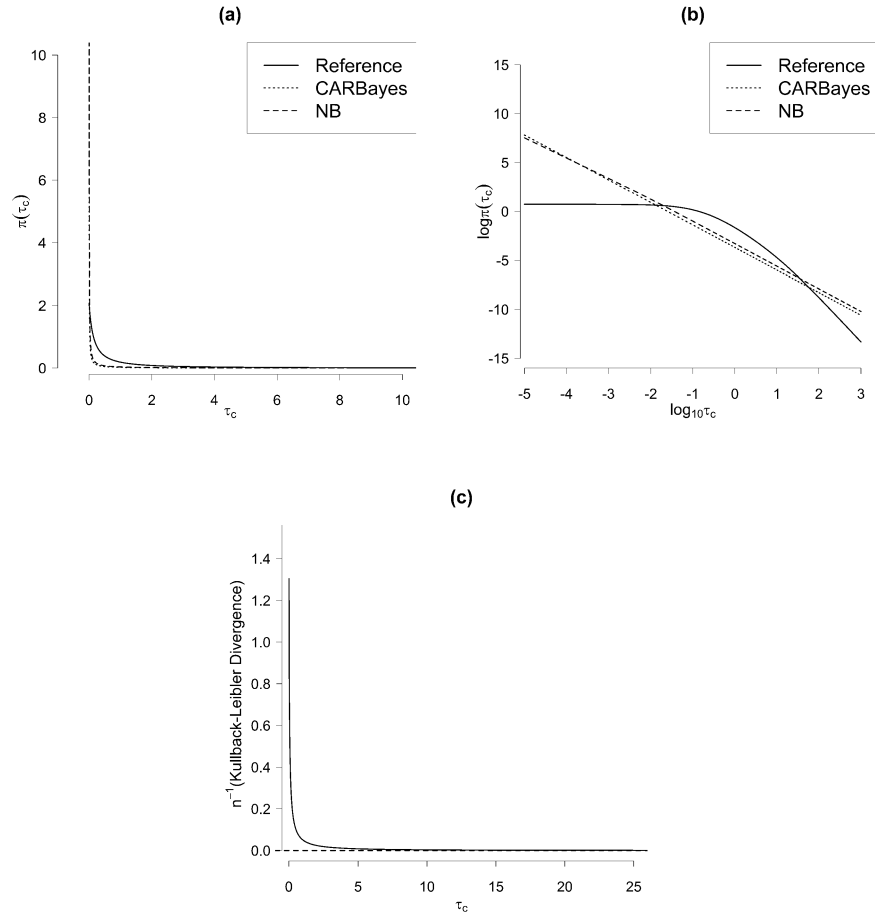


Figure 1: (a) Raw plot of $\pi(\tau_c)$ and (b) plot of $\log \pi(\tau_c)$ vs. $\log_{10} \tau_c$ for the reference, CARBayes, and NB priors. The CARBayes and NB priors both approach ∞ as $\tau_c \rightarrow 0$, while the reference prior does not. Panel (c) shows the Kullback–Leibler divergence per observation between the spatial model and the independent data model across values of τ_c , indicating that large values of τ_c correspond to a model for nearly independent data.

in (1), the analyst should consider a proper CAR model directly for the observations (e.g., see Ferreira and De Oliveira, 2007).

In order to better understand the importance of the parameter τ_c as it relates to the strength of spatial dependence, we consider the Kullback–Leibler divergence between the spatial model and the independent data model as a function of τ_c . Specifically, if p_i and p_s correspond to the independent and spatial models, respectively, we consider the Kullback–Leibler divergence given by

$$\begin{aligned}
KL(p_i \parallel p_s) &= \int p_i(x) \log \left[\frac{p_i(x)}{p_s(x)} \right] dx \\
&= -\frac{n}{2} + \frac{1}{2} \log |\Omega| + \frac{1}{2} \sum_{i=1}^n (1 + \tau_c^{-1} d_i^{-1})^{-1}.
\end{aligned} \tag{26}$$

Note that as $\tau_c \rightarrow \infty$, $KL(p_i \parallel p_s) \rightarrow 0$ indicating that the spatial and independent data models are nearly identical for large values of τ_c . Specifically, Figure 1(c) shows that for values of $\tau_c > 10$, the Kullback–Leibler divergence per observation is nearly zero, indicating that the spatial model we consider is only relevant for smaller values of τ_c . The CARBayes and NB priors unnecessarily place significantly more mass on large values of τ_c . In contrast, the reference prior intuitively places mass on values of τ_c that are typical of spatially dependent data.

5.2 Simulation Design

For the design of our simulations, we consider square regions with three different sample sizes of $n = 5^2, 7^2, 10^2$. We fix $\sigma^2 = 2$, while considering values of $\tau_c = 0.01, 0.032, 0.1, 0.32, 1, 3.2, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$. To explore how the number of neighbors which exert spatial dependence on a subregion affects inferential performance, we investigate both first- and second-order neighborhood structure (i.e., diagonal subregions are not/are considered neighbors, respectively). We also consider $p = 1$ (intercept only) with $\beta = 1$ and $p = 6$ with $\beta = (-3, -2, -1, 1, 2, 3)^T$. All covariates are generated from a normal distribution with mean 0 and variance 1. We generate results based on 1,000 simulated data sets for each combination of these levels of n , τ_c , neighborhood structure, and p . For each data set, 15,000 MCMC iterations are obtained with the first 5,000 iterations discarded as burn-in. The Gelman–Rubin convergence diagnostic (Gelman and Rubin, 1992) for τ_c and σ^2 for three MCMC chains using one setting of parameter values was used to determine the number of MCMC iterations needed to obtain convergence. Gelman–Rubin convergence diagnostics for τ_c and σ^2 were calculated to be less than 1.01, indicating that 15,000 MCMC iterations is sufficient for our simulation study. The variances of the proposal distributions described in Section 4 were chosen to be $\delta_s = \delta_t = 0.5$ yielding an acceptance rate close to 40%. Due to the heavily right-skewed posterior density for τ_c , we consider 95% highest posterior density (HPD) regions for τ_c , while 95% equal-tailed credible intervals are considered for σ^2 and β .

Performance results pertaining to τ_c less than or equal to 10, first-order neighborhood structure, and $p = 6$ are displayed graphically and discussed in Section 5.3. For the sake of brevity, performance results pertaining to β , σ^2 , second-order neighborhood structure, or $p = 1$ are described in Section 5.3 and tabulated in the Supplementary Material. To investigate the case of a larger number of regressors, we have also performed a simulation study with $p = 20$ when $n = 100$; because the results are qualitatively similar to the results discussed in Section 5.3, we do not include them here. Simulation results for values of $\tau_c > 10$ indicate that the performance of all considered priors deteriorates as τ_c increases (see Simulation Results in Supplementary Material; To make the information easier to digest, in addition to the values of τ_c less or equal to 10, the tables provided in the Supplementary Material include only three representative values

of τ_c greater than 10: 40, 70, and 100.). In particular, the frequentist coverages of the reference and NB credible intervals for τ_c decrease substantially as τ_c increases much beyond 10. The frequentist coverage of the CARBayes credible interval remains above nominal, but the CARBayes credible interval is so wide that it becomes of no practical use. The deterioration of performance of the reference credible interval is not a problem, because as illustrated by the Kullback–Leibler divergence shown in Figure 1(c), larger values of τ_c indicate that the observations are nearly independent, which may imply that a model without spatial dependence would be preferred. We focus on values of $\tau_c < 10$, since our reference prior is intended for a model that accounts for spatial dependence.

5.3 Simulation Results

Figure 2 displays the frequentist interval coverage rate and average IL (on the \log_{10} scale) for τ_c resulting from the reference, CARBayes, and NB priors across a range of values for τ_c (on the \log_{10} scale) for various sample sizes, assuming $p = 6$ and first-order neighborhood structure. Figure 2 demonstrates that the reference prior achieves nominal interval coverage and low average interval length as n increases across a range of values for τ_c . Further, the CARBayes and NB priors exhibit a dip in interval coverage for values of τ_c corresponding to strong spatial dependence. The CARBayes prior yields particularly wide average interval lengths, potentially detracting from the value of its close-to-nominal coverage. These results demonstrate that, when compared to the two gamma priors, the reference prior leads to favorable coverage and interval length.

Mean squared error of the posterior median of τ_c is shown in Figure 3, which demonstrates that use of the reference prior leads to advantageous estimation properties across all sample sizes, especially smaller sample sizes. Use of the CARBayes prior leads to considerably higher MSE than that of the other priors.

All credible intervals for each of the regression coefficients β achieve the nominal level for frequentist coverage with comparable interval lengths across all three priors. Note that for the tabulated results in the Supplementary Material for the case of $p = 6$, the frequentist coverage and average interval length are averaged across all six of the regression coefficients since results were comparable for each of the regression coefficients. Additionally, all three priors show reasonable performance in terms of frequentist coverage and average interval length for σ^2 . Results for second-order neighborhood structure are qualitatively similar to those presented for first-order neighborhood structure. Likewise, results for $p = 1$ are qualitatively similar to those for $p = 6$.

An intriguing result is that, even though Figure 1(b) seems to show that the marginal CARBayes and NB priors for τ_c are fairly similar, the frequentist properties of the CARBayes and NB posteriors can be very different. The issue here is that in the CARBayes and NB specifications, τ_c and σ^2 are not independent a priori. Specifically, it is straightforward to show that in both CARBayes and NB specifications, the conditional prior for σ^2 given τ_c is $IG(\alpha_1 + \alpha_2, \beta_1 + \beta_2\tau_c)$, where $IG(\cdot, \cdot)$ corresponds to an inverse gamma distribution defined by its shape and scale parameters, respectively. Hence, the joint posteriors of σ^2 and τ_c (and the marginal posteriors of τ_c) under the CARBayes and

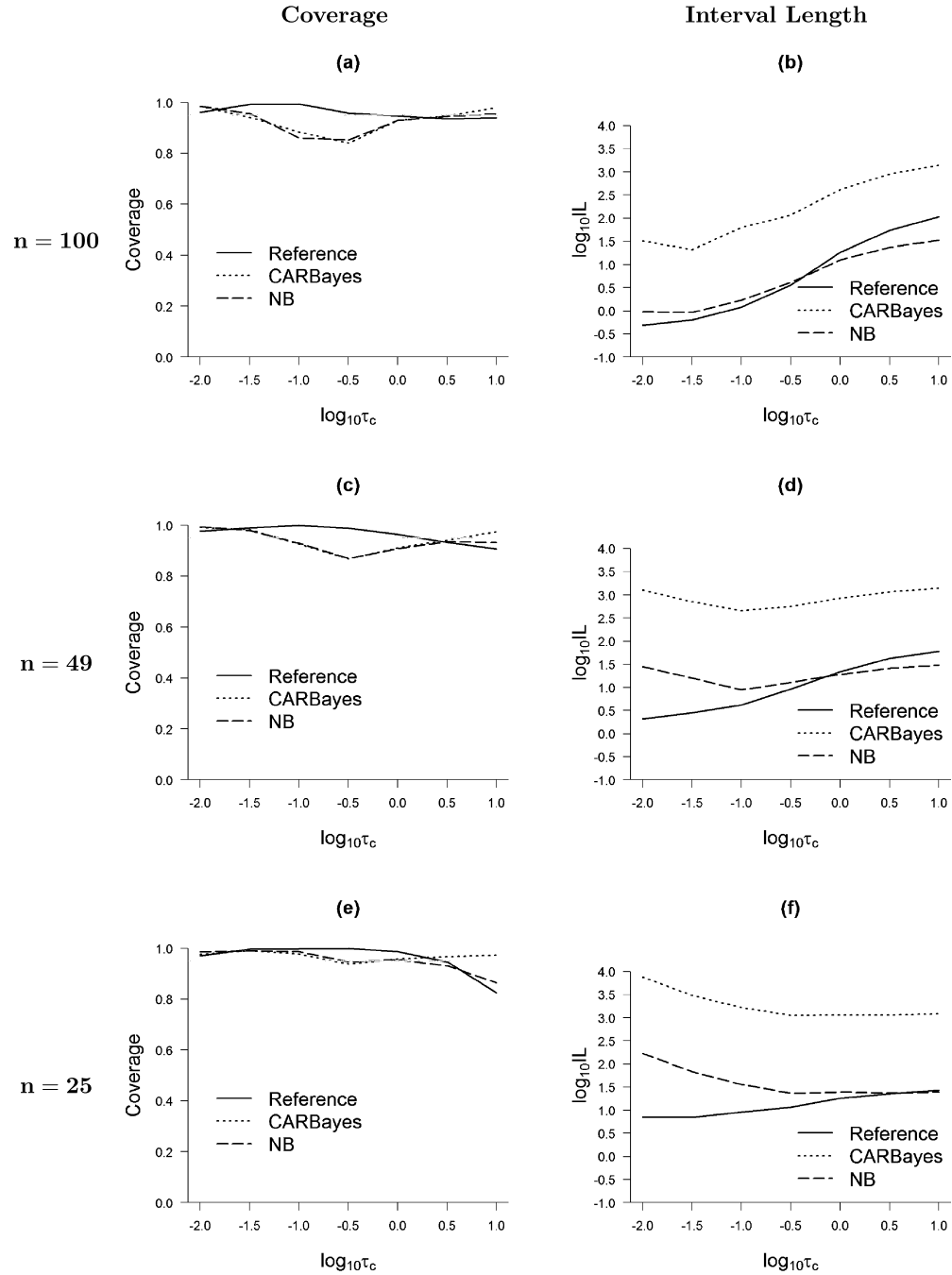


Figure 2: Frequentist coverage and \log_{10} average interval length (IL) for τ_c for $n = 100$ (top row), $n = 49$ (middle row), and $n = 25$ (bottom row). Reference prior shows favorable performance in terms of frequentist coverage and average interval length.

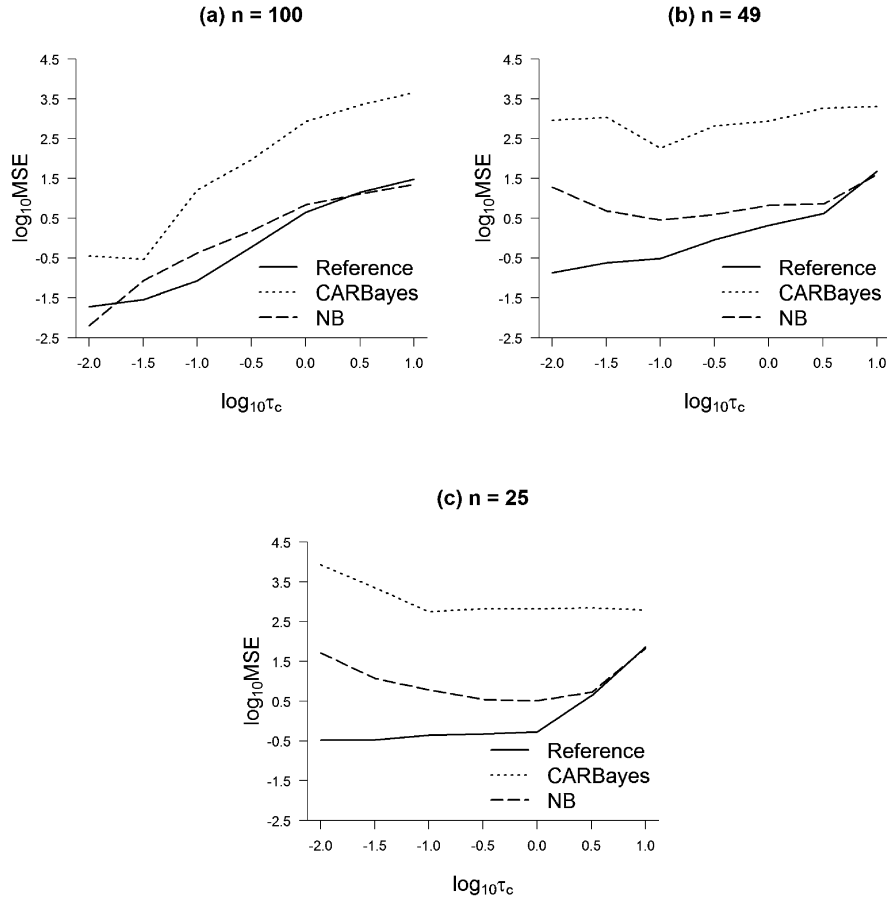


Figure 3: $\log_{10}\text{MSE}$ for the posterior median of τ_c for (a) $n = 100$, (b) $n = 49$, and (c) $n = 25$. Reference prior leads to favorable performance in terms of estimation of τ_c . Note that since the y-axis is on the \log_{10} scale, the difference in MSE of the posterior median of τ_c between the reference prior and the CARBayes prior is considerable.

NB specifications may differ substantially. Therefore, even though the marginal priors for τ_c under the CARBayes and NB specifications behave similarly, the frequentist properties of CARBayes and NB procedures for τ_c may be very different.

This is illustrated in Figure S1 (Supplementary Material) that shows the impact of the joint prior of (σ^2, τ_c) on the posterior inference for these parameters for a difficult dataset. Specifically, Figure S1 shows contour plots of (a) integrated likelihood of (σ^2, τ_c) , and joint posterior density of (σ^2, τ_c) based on (b) reference prior, (c) NB prior, and (d) CARBayes prior. The dataset was simulated with a first-order neighborhood structure, square regular grid, $n = 49$, and parameters $\beta = 1, \sigma^2 = 2$, and $\tau_c = 0.1$. Contours correspond to HPD regions with credible levels equal to 10%, 20%, ..., 90%.

The integrated likelihood behaves badly for this dataset and at first glance seems not to provide much information about σ^2 and τ_c . We have found empirically that the probability of such a bad integrated likelihood behavior increases as the Moran-I statistic decreases. For the particular dataset used to produce Figure S1, the Moran-I statistic is equal to 0.203 with a corresponding p-value of 0.0181 (null hypothesis is of no spatial dependence). This is not an extreme dataset; under the conditions used to simulate this dataset, the probability of the Moran-I statistic being less than 0.203 is about 0.12. For this dataset, the CARBayes analysis provides unacceptably large HPD regions. The NB analysis provides somewhat smaller HPD regions than the CARBayes analysis. Finally, in this case the reference analysis provides much smaller HPD regions. True parameter values are indicated with a red dot, and the posterior median under each prior is indicated with a red star. The posterior medians of τ_c are equal to 1.0, 2.2, and 3.2 for the reference, NB, and CARBayes analyses, respectively. Hence, analyses based on the reference, NB, and CARBayes priors may differ substantially. Finally, for this dataset the reference analysis provides much tighter HPD regions and a much better estimate of τ_c .

Finally, Figure S2 in the Supplementary Material shows the behavior of the integrated likelihood and the posterior densities for a dataset with Moran-I statistic of moderate size. Specifically, for this dataset the Moran-I statistic is equal to 0.363 with a corresponding p-value of 0.00017 (null hypothesis is of no spatial dependence). This Moran-I statistic is close to the median of the sampling distribution of the Moran-I statistics. The simulation setting was the same as in the previous paragraph: first-order neighborhood structure, square regular grid, $n = 49$, and parameters $\beta = 1$, $\sigma^2 = 2$, and $\tau_c = 0.1$. For this dataset, the integrated likelihood behaves much better than the one discussed in the previous paragraph. However, the integrated likelihood tends to put much of its mass at values of σ^2 and τ_c larger than the true values. The reference prior corrects the integrated likelihood behavior and leads to a reference posterior density that has much of its mass close to the true values of the parameters. For this dataset, the NB and CARBayes priors tend to shrink the posterior density too much towards zero. Specifically, the posterior medians of τ_c are equal to 0.095 (reference), 0.017 (NB), and 0.015 (CARBayes), while the posterior medians of σ^2 are 2.8 (reference), 0.6 (NB), and 0.6 (CARBayes). Thus, for this dataset the NB and CARBayes analyses are actually fairly similar and severely underestimate both σ^2 and τ_c . Meanwhile, the reference analysis provides a reasonable correction to the integrated likelihood and yields estimates for σ^2 and τ_c much closer to the true values of these parameters. Such beneficial correction of badly behaved likelihood functions by reference (and other non-informative) priors has been previously reported in other statistical inference problems such as for example in the analysis of generalized linear models (Firth, 1993), linear regression with Student-t errors (Fonseca et al., 2008) and exponential power errors (Salazar et al., 2012; Ferreira and Salazar, 2014), and analysis of elapsed times in continuous time Markov chains (Ferreira and Suchard, 2008) used in phylogenetic tree reconstruction (Bouckaert et al., 2014).

In summary, the proposed reference prior exhibits a favorable combination of high interval coverage, short average interval length, and low MSE relative to the CARBayes and NB priors, in addition to philosophical appeal and practical convenience relative to vague naïve priors previously used in the literature.

6 Case Study

6.1 Data Description

To illustrate an objective Bayesian analysis using the model in (9) along with the reference prior, we consider a data set containing foreclosure rates as a proportion of all housing transactions for each of the 88 counties in the state of Ohio for the year 2012. This data set is a subset of a larger database of approximately 54 million records from municipalities across the United States between 2005 and 2014 that we obtained from our partnerships with the Virginia Center for Housing Research and Metrostudy, a Hanley Wood Company. The data include very fine spatial and temporal information, including latitude, longitude, and sale date of each closing record. We choose to aggregate the data at the county level to illustrate this analysis. For more details describing how records were classified as foreclosures, see Keefe et al. (2017). In addition to foreclosure rates, county unemployment rates are also available as a potential covariate to be used in modeling (Bureau of Labor Statistics, 2012). For each county, the total number of housing transactions and the observed number of foreclosures are available. Analogously to what is typically done in disease mapping, we consider the standardized morbidity ratio (SMR) for each county, defined by $SMR_i = O_i/E_i$, where O_i is the observed number of foreclosures in county i and E_i is the expected number of foreclosures in county i . The expected counts are calculated by $E_i = n_i(\sum_j O_j/\sum_j n_j)$, where n_i is the total number of housing transactions in county i . A map of the SMRs for foreclosure rates in Ohio counties for 2012 is shown in Figure 4(a). Figure 4(b) shows a map of the unemployment rates in 2012 for each county.

6.2 Modeling

Using the 2012 foreclosure rate data set, we fit the model given in (9) using the reference, CARBayes, and NB priors in order to compare the results using the proposed reference prior to the results using the priors used frequently in practice. We consider $y_i = \log(SMR_i)$ as the response variable and unemployment rate as a covariate. The posterior distribution was sampled using the MCMC algorithm described in Section 4 with 25,000 iterations where the first 5,000 were discarded as burn-in. Gelman–Rubin convergence diagnostics were calculated using four MCMC chains with the reference prior implemented with different starting values for $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, τ_c and σ^2 . The 25,000 MCMC iterations with 5,000 burn-in iterations yielded Gelman–Rubin diagnostics that were all below 1.005, indicating acceptable posterior convergence. Posterior summaries, including posterior medians and credible intervals for the parameters are provided in Table 1. Note that 95% HPD credible intervals are used for τ_c , while 95% equal-tailed credible intervals are used for $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and σ^2 .

From Table 1, we see that posterior estimates for the fixed effect regression coefficients are all similar. The estimates for τ_c and σ^2 vary quite a bit for the different priors. In particular, the analysis using the NB prior implied much stronger spatial dependence than the reference prior with an estimate of τ_c closer to 0. The estimated values of $\log_{10} \tau_c$ for all three analyses are between -0.8 and -0.55 , which according to our simulation study are values of τ_c for which the proposed reference prior provides

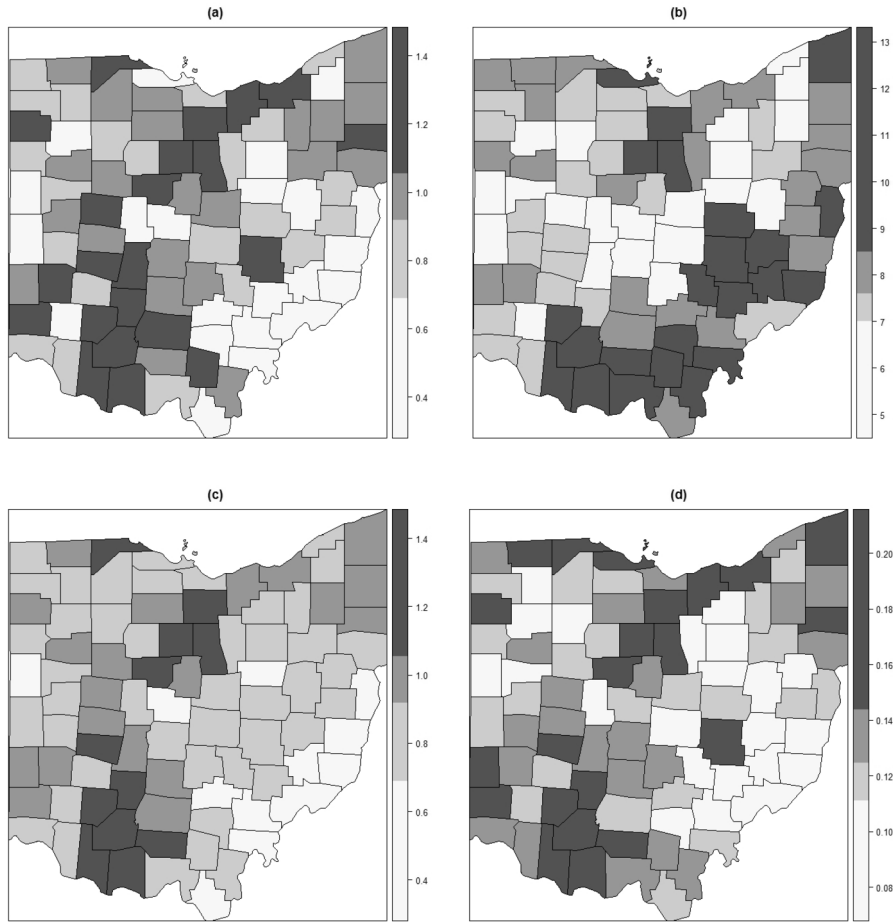


Figure 4: Map of (a) 2012 SMRs for foreclosure rates, (b) 2012 unemployment rates (%), (c) posterior median $E[SMR]$, and (d) posterior standard deviation of $E[SMR]$ in Ohio counties.

superior results in terms of interval coverage, average interval length, and MSE for τ_c . Thus, we should trust the resulting inference based on the reference prior more than the other priors. Furthermore, note that the credible interval for τ_c is wider when the CARBayes prior is used, while use of the reference and NB priors results in more narrow credible intervals. This supports the findings of our simulation study regarding the average interval length for τ_c . Although the credible interval for τ_c using the CARBayes prior is only slightly wider in the case study, the simulation study results illustrate the potential danger of using this particular vague prior. While the analysis of foreclosure rates does not result in undesirably wide credible intervals for τ_c when the CARBayes prior is used, our simulation study shows that other data sets may lead to credible intervals for τ_c that are orders of magnitude wider than those obtained when the reference

Prior	Parameter	Estimate	95% Credible Interval
Reference	β_0 (intercept)	-0.5561	(-0.9388, -0.1710)
	β_1 (unemployment rate)	0.0488	(0.0008, 0.0967)
	τ_c	0.2519	(0.0017, 0.9781)
	σ^2	0.0432	(0.0046, 0.0854)
CARBayes	β_0 (intercept)	-0.5612	(-0.9482, -0.1708)
	β_1 (unemployment rate)	0.0496	(0.0009, 0.0977)
	τ_c	0.1959	(0.0009, 1.0231)
	σ^2	0.0385	(0.0013, 0.0819)
NB	β_0 (intercept)	-0.5687	(-0.9516, -0.1802)
	β_1 (unemployment rate)	0.0504	(0.0020, 0.0985)
	τ_c	0.1589	(0.0010, 0.6290)
	σ^2	0.0345	(0.0019, 0.0755)

Table 1: Posterior Summaries for Foreclosure Rate Case Study for Bayesian Analyses Using Reference, CARBayes, and NB Priors.

prior is used. Unreasonably wide credible intervals for τ_c could lead practitioners to decide that a spatial model is not necessary for the data, resulting in the incorrect use of a non-spatial model. We thus recommend use of the reference prior.

The estimate of the regression coefficient for unemployment rate is positive and its credible interval does not contain zero in all three cases. This implies that, on average, as unemployment rate increases, foreclosure rates also increase. Consider the posterior distribution of the expected SMR in county i given by $E[\text{SMR}_i] = \exp\{\beta_0 + \beta_1 x_i + \phi_i\}$, where x_i is the unemployment rate of county i for $i = 1, 2, \dots, n$. Figures 4(c) and 4(d) show maps of the posterior median and posterior standard deviation, respectively, of the expected SMR values computed directly from the results of the MCMC algorithm using the reference prior. These maps show that counties with higher unemployment rates tend to have higher risk of foreclosure than counties with lower unemployment rates.

As shown in the simulation study, the main selling point of the reference prior is that, in the absence of prior information, it allows a more realistic assessment of the uncertainty for the several parameters. This actually makes difference in our case study. For example, we may be interested in identifying clusters of counties with risk higher than predicted by the regressors. If that is the case, we would analyze the spatial random effects. A common decision rule would be to say that counties for which $P(\phi_i > 0 | \mathbf{Y}) > 0.95$ would have risk higher than predicted by the regressors. With respect to counties with risk higher than predicted by the regressor, the reference, CARBayes, and NB analyses identify 3, 4, and 9 counties respectively. Similarly to the results presented in Table 1, for this particular dataset the CARBayes analysis provides results that are in between those of the reference and the NB analyses. Further, the NB analysis identifies many more counties than the reference analysis. Thus, the analyses based on the different priors lead to very different conclusions in terms of what counties have risk of foreclosure higher than predicted by the regressor. That is because such analyses are based on tail probabilities that depend crucially on an appropriate assess-

ment of uncertainty through the variance parameters σ^2 and τ_c . As our simulation study indicates, amongst the analyses considered in this study, the reference prior provides the best uncertainty quantification for these parameters and is, therefore, preferable.

It is often difficult to choose sensible hyperparameters for the commonly used priors for this type of spatial model. For instance, it might be difficult for a housing market expert to articulate their understanding of the spatial dependence among counties' foreclosure rates in such a way as to inform prior specification for a Bayesian hierarchical model. The proposed reference prior is an appealing alternative because it is automatic. It does not require the choice of hyperparameters and has favorable performance. This reference prior is useful in situations where researchers are unsure of how to subjectively choose priors for areal data with a nugget effect.

7 Discussion

This paper presents a fully Bayesian analysis for a commonly used Gaussian hierarchical model for spatial data. We have derived explicit expressions for several objective priors, including the reference, independence Jeffreys, and Jeffreys-rule priors for the Gaussian hierarchical model with an intrinsic CAR prior for the spatial random effects. We have shown that the reference prior results in a proper posterior distribution, while the independence Jeffreys and Jeffreys-rule priors lead to improper posterior distributions. Furthermore, we have studied frequentist properties of Bayesian procedures using the proposed reference prior and two commonly used priors for this model.

We have determined that the reference prior leads to a combination of favorable frequentist coverage, average interval length, and mean squared error relative to two commonly used priors. We have demonstrated using the Kullback–Leibler divergence that focus should be placed on small values of τ_c , which correspond to strong spatial dependence. So, while inferential performance deteriorates across all considered priors for larger values of τ_c , this is not a concern since perhaps a model without spatial dependence would be preferred in this situation. More importantly, the reference prior performs better than the competing priors for smaller values of τ_c in terms of frequentist coverage, average interval length, and mean squared error.

Additionally, the reference prior approach obviates the need to subjectively specify hyperparameters, allowing for an automatic Bayesian analysis. This philosophy works well for hierarchical modeling when interpretation of hyperparameters and elicitation of meaningful priors is difficult. Often times, practitioners choose to use a vague naïve prior in the absence of prior information or understanding of the problem. However, in situations such as the one considered here, where an improper prior for τ_c would lead to an improper posterior, the use of a vague naïve prior only masks the impropriety issue, rather than solving it (Berger, 2006). By contrast, an objective prior like the reference prior proposed here, will lead to a proper posterior distribution and will let the data speak for themselves.

Although much research has been done on objective Bayesian analysis for spatial models, such as geostatistical models (Berger et al., 2001; De Oliveira, 2007), proper CAR models for the observed data (Ferreira and De Oliveira, 2007; De Oliveira, 2012;

Ren and Sun, 2013), and proper CAR models for latent process models for areal data (Ren and Sun, 2014), the hierarchical model with an intrinsic CAR prior that we consider here is one of the most popularly used models in applications of disease mapping and other areas of research. All intrinsic CAR models require subjective decisions regarding rigorous specification and additional constraints to ensure propriety (Lavine and Hodges, 2012). Keefe et al. (2018) have shown that if we first impose the sum-zero constraint and then consider the limit to obtain the sum-zero constrained intrinsic CAR, there is a broad class of proper CAR priors that all result in the unique intrinsic CAR prior given in Equation (8). Our development of an appropriate objective Bayesian analysis for this model will hopefully help researchers analyze areal data.

There are many possible avenues for future research. For example, this work could be extended by developing objective priors for areal models for non-Gaussian responses, such as counts and rates that can be more accessible to disease mapping applications such as the one described in Section 6. Other future work may also include the study and development of objective priors for spatio-temporal models for areal data, as well as the effect of the choice of priors for these types of models.

Supplementary Material

Supplementary Material of Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors (DOI: [10.1214/18-BA1107SUPP](https://doi.org/10.1214/18-BA1107SUPP); .pdf).

References

- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press. MR3362184. 181
- Bell, B. S. and Broemeling, L. D. (2000). “A Bayesian analysis for spatial processes with application to disease mapping.” *Statistics in Medicine*, 19(7): 957–974. 181, 182
- Berger, J. (2006). “The case for objective Bayesian analysis.” *Bayesian Analysis*, 1(3): 385–402. MR2221271. doi: <https://doi.org/10.1214/06-BA115>. 204
- Berger, J. O., De Oliveira, V., and Sansó, B. (2001). “Objective Bayesian analysis of spatially correlated data.” *Journal of the American Statistical Association*, 96(456): 1361–1374. MR1946582. doi: <https://doi.org/10.1198/016214501753382282>. 182, 189, 204
- Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). “Bayesian estimates of disease maps: how important are priors?” *Statistics in Medicine*, 14(21–22): 2411–2431. 182
- Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. New York: Wiley. MR1274699. doi: <https://doi.org/10.1002/9780470316870>. 189
- Besag, J. (1974). “Spatial interaction and the statistical analysis of lattice systems.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 192–236. MR0373208. 181

- Besag, J. and Kooperberg, C. (1995). “On conditional and intrinsic autoregressions.” *Biometrika*, 82(4): 733–746. MR1380811. 183, 185
- Besag, J., York, J., and Mollié, A. (1991). “Bayesian image restoration, with two applications in spatial statistics.” *Annals of the Institute of Statistical Mathematics*, 43(1): 1–20. MR1105822. doi: <https://doi.org/10.1007/BF00116466>. 181, 185
- Best, N., Richardson, S., and Thomson, A. (2005). “A comparison of Bayesian spatial models for disease mapping.” *Statistical Methods in Medical Research*, 14(1): 35–59. MR2135924. doi: <https://doi.org/10.1191/0962280205sm388oa>. 182
- Best, N., Waller, L., Thomas, A., Conlon, E., and Arnold, R. (1999). “Bayesian models for spatially correlated disease and exposure data.” In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, volume 6, 131. Oxford University Press. MR1723496. 182, 194
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M., Rambaut, A., and Drummond, A. J. (2014). “BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.” *PLoS Computational Biology*, 10(4): e1003537. 200
- Bureau of Labor Statistics (2012). “Local Area Unemployment Statistics.” <http://www.bls.gov/lau/>. Accessed: 2014-07-14. 201
- Clayton, D. and Kaldor, J. (1987). “Empirical Bayes estimates of age-standardized relative risks for use in disease mapping.” *Biometrics*, 43(3): 671–681. 181
- De Oliveira, V. (2007). “Objective Bayesian analysis of spatial data with measurement error.” *Canadian Journal of Statistics*, 35(2): 283–301. MR2393610. doi: <https://doi.org/10.1002/cjs.5550350206>. 182, 186, 187, 188, 190, 192, 204
- De Oliveira, V. (2012). “Bayesian analysis of conditional autoregressive models.” *Annals of the Institute of Statistical Mathematics*, 64(1): 107–133. MR2878920. doi: <https://doi.org/10.1007/s10463-010-0298-1>. 182, 204
- De Oliveira, V. and Ferreira, M. A. R. (2011). “Maximum likelihood and restricted maximum likelihood estimation for a class of Gaussian Markov random fields.” *Metrika*, 74(2): 167–183. MR2822154. doi: <https://doi.org/10.1007/s00184-009-0295-7>. 185
- Dietrich, C. (1991). “Modality of the restricted likelihood for spatial Gaussian random fields.” *Biometrika*, 78(4): 833–839. MR1147020. doi: <https://doi.org/10.1093/biomet/78.4.833>. 188
- Ferreira, M. A. R. (2018). “The Limiting Distribution of the Gibbs Sampler for the Intrinsic Conditional Autoregressive Model.” Technical report, Department of Statistics, Virginia Tech. 186
- Ferreira, M. A. R. and De Oliveira, V. (2007). “Bayesian reference analysis for Gaussian Markov random fields.” *Journal of Multivariate Analysis*, 98(4): 789–812. MR2322129. doi: <https://doi.org/10.1016/j.jmva.2006.07.005>. 182, 195, 204

- Ferreira, M. A. R., Holan, S. H., and Bertolde, A. I. (2011). “Dynamic multiscale spatiotemporal models for Gaussian areal data.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5): 663–688. MR2867453. doi: <https://doi.org/10.1111/j.1467-9868.2011.00774.x>. 185
- Ferreira, M. A. R. and Salazar, E. (2014). “Bayesian reference analysis for exponential power regression models.” *Journal of Statistical Distributions and Applications*, 1(1): 1–12. 200
- Ferreira, M. A. R. and Suchard, M. A. (2008). “Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains.” *Canadian Journal of Statistics*, 36: 355–368. MR2456010. doi: <https://doi.org/10.1002/cjs.5550360302>. 200
- Firth, D. (1993). “Bias reduction of maximum likelihood estimates.” *Biometrika*, 80: 27–38. MR1225212. doi: <https://doi.org/10.1093/biomet/80.1.27>. 200
- Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. (2008). “Objective Bayesian analysis for the Student- t regression model.” *Biometrika*, 95(2): 325–333. MR2521587. doi: <https://doi.org/10.1093/biomet/asn001>. 200
- Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. CRC Press. MR2260716. 192
- Gelfand, A. E. and Smith, A. F. (1990). “Sampling-based approaches to calculating marginal densities.” *Journal of the American Statistical Association*, 85(410): 398–409. MR1141740. 192
- Gelman, A. and Rubin, D. B. (1992). “Inference from iterative simulation using multiple sequences.” *Statistical Science*, 7(4): 457–472. 196
- Gilks, W. R., Best, N., and Tan, K. (1995). “Adaptive rejection Metropolis sampling within Gibbs sampling.” *Applied Statistics*, 44(4): 455–472. 192
- Goicoa, T., Ugarte, M., Etxeberria, J., and Militino, A. (2016). “Age–space–time CAR models in Bayesian disease mapping.” *Statistics in Medicine*, 35(14): 2391–2405. MR3513520. doi: <https://doi.org/10.1002/sim.6873>. 181
- Hodges, J. S., Carlin, B. P., and Fan, Q. (2003). “On the precision of the conditionally autoregressive prior in spatial models.” *Biometrics*, 59(2): 317–322. MR1987398. doi: <https://doi.org/10.1111/1541-0420.00038>. 181
- Keefe, M. J., Ferreira, M. A. R., and Franck, C. T. (2018). “On the formal specification of sum-zero constrained intrinsic conditional autoregressive models.” *Spatial Statistics*, 24: 54–65. doi: <https://doi.org/10.1016/j.spasta.2018.03.007>. 186, 205
- Keefe, M. J., Ferreira, M. A. R., and Franck, C. T. (2019). “Supplementary Material of Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/18-BA1107SUPP>. 183
- Keefe, M. J., Franck, C. T., and Woodall, W. H. (2017). “Monitoring foreclosure rates with a spatially risk-adjusted Bernoulli CUSUM chart for concurrent observations.”

- Journal of Applied Statistics*, 44(2): 325–341. MR3607739. doi: <https://doi.org/10.1080/02664763.2016.1169257>. 201
- Kuo, B.-S. (1999). “Asymptotics of ML estimator for regression models with a stochastic trend component.” *Econometric Theory*, 15(01): 24–49. MR1680058. doi: <https://doi.org/10.1017/S0266466699151028>. 188
- Lavine, M. L. and Hodges, J. S. (2012). “On rigorous specification of ICAR models.” *The American Statistician*, 66(1): 42–49. MR2934740. doi: <https://doi.org/10.1080/00031305.2012.654746>. 186, 205
- Lee, D. (2013). “CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors.” *Journal of Statistical Software*, 55(13): 1–24. 182, 194
- Liu, Z., Berrocal, V. J., Bartsch, A. J., and Johnson, T. D. (2016). “Pre-surgical fMRI Data Analysis Using a Spatially Adaptive Conditionally Autoregressive Model.” *Bayesian Analysis*, 11: 599–625. MR3472004. doi: <https://doi.org/10.1214/15-BA972>. 181
- Mercer, L. D., Wakefield, J., Pantazis, A., Lutambi, A. M., Masanja, H., and Clark, S. (2015). “Space–time smoothing of complex survey data: Small area estimation for child mortality.” *The Annals of Applied Statistics*, 9(4): 1889–1905. MR3456357. doi: <https://doi.org/10.1214/15-A0AS872>. 181
- Moraga, P. and Lawson, A. B. (2012). “Gaussian component mixtures and CAR models in Bayesian disease mapping.” *Computational Statistics & Data Analysis*, 56(6): 1417–1433. MR2892351. doi: <https://doi.org/10.1016/j.csda.2011.11.011>. 181
- Muirhead, R. J. (2009). *Aspects of Multivariate Statistical Theory*, volume 197. John Wiley & Sons. MR0652932. 185
- Natarajan, R. and McCulloch, C. E. (1998). “Gibbs sampling with diffuse proper priors: A valid approach to data-driven inference?” *Journal of Computational and Graphical Statistics*, 7(3): 267–277. 182
- Penrose, R. (1955). “A generalized inverse for matrices.” In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 51, 406–413. Cambridge Univ Press. MR0069793. 185
- Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). “Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models.” *Biometrics*, 62(4): 1197–1206. MR2307445. doi: <https://doi.org/10.1111/j.1541-0420.2006.00617.x>. 181
- Ren, C. and Sun, D. (2013). “Objective Bayesian analysis for CAR models.” *Annals of the Institute of Statistical Mathematics*, 65(3): 457–472. MR3067433. doi: <https://doi.org/10.1007/s10463-012-0377-6>. 182, 204
- Ren, C. and Sun, D. (2014). “Objective Bayesian analysis for autoregressive models with nugget effects.” *Journal of Multivariate Analysis*, 124: 260–280. MR3147325. doi: <https://doi.org/10.1016/j.jmva.2013.11.003>. 182, 192, 205

- Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2nd edition. MR2080278. doi: <https://doi.org/10.1007/978-1-4757-4145-2>. 192
- Salazar, E., Ferreira, M. A. R., and Migon, H. S. (2012). “Objective Bayesian analysis for exponential power regression models.” *Sankhya – Series B*, 74: 107–125. MR3021041. doi: <https://doi.org/10.1007/s13571-012-0045-0>. 200
- Sun, D., Tsutakawa, R. K., and Speckman, P. L. (1999). “Posterior distribution of hierarchical models using CAR (1) distributions.” *Biometrika*, 86(2): 341–350. MR1705418. doi: <https://doi.org/10.1093/biomet/86.2.341>. 181
- Verbyla, A. P. (1990). “A conditional derivation of residual maximum likelihood.” *Australian Journal of Statistics*, 32(2): 227–230. 188

Acknowledgments

The authors would like to thank Dr. Andrew McCoy, Director of the Virginia Center for Housing Research, and Metrostudy, a HanleyWood Company for providing the data used in the foreclosure case study. The authors also appreciate the helpful feedback from Dr. Victor De Oliveira of the University of Texas at San Antonio Department of Management Science and Statistics. The authors also thank the insightful comments and suggestions of two anonymous reviewers, the Associate Editor, and the Editor.