Bayesian Analysis (2024) **19**, Number 4, pp. 985–1011

Objective Bayesian Model Selection for Spatial Hierarchical Models with Intrinsic Conditional Autoregressive Priors*

Erica M. Porter[†,¶], Christopher T. Franck[‡], and Marco A. R. Ferreira[§]

Abstract. We develop Bayesian model selection via fractional Bayes factors to simultaneously assess spatial dependence and select regressors in Gaussian hierarchical models with intrinsic conditional autoregressive (ICAR) spatial random effects. Selection of covariates and spatial model structure is difficult, as spatial confounding creates a tension between fixed and spatial random effects. Researchers have commonly performed selection separately for fixed and random effects in spatial hierarchical models. Simultaneous selection methods relieve the researcher from arbitrarily fixing one of these types of effects while selecting the other. Notably, Bayesian approaches to simultaneously select covariates and spatial model structure are limited. Our use of fractional Bayes factors allows for selection of fixed effects and spatial model structure under automatic reference priors for model parameters, which obviates the need to specify hyperparameters for priors. We also show the equivalence between two ICAR specifications and derive the minimal training size for the fractional Bayes factor applied to the ICAR model under the reference prior. We perform a simulation study to assess the performance of our approach and we compare results to the Deviance Information Criterion and Widely Applicable Information Criterion. We demonstrate that our fractional Bayes factor approach assigns low posterior model probability to spatial models when data is truly independent and reliably selects the correct covariate structure with highest probability within the model space. Finally, we demonstrate our Bayesian model selection approach with applications to county-level median household income in the contiguous United States and residential crime rates in the neighborhoods of Columbus, Ohio.

Keywords: Bayesian model selection, spatial statistics, areal data, ICAR random effects, fractional Bayes factor.

1 Introduction

Bayesian hierarchical models are often used to model spatial data because they are flexible enough to accommodate both fixed regression effects and spatial random effects. In particular, hierarchical models with conditional autoregressive (CAR) structure (Besag, 1974) and intrinsic conditional autoregressive (ICAR) structure (Besag et al., 1991)

*The work of Ferreira was supported in part by National Science Foundation Awards DMS 1853549 and DMS 2054173.

†Department of Statistics, Virginia Tech, Blacksburg, VA 24061, ericamp@vt.edu
‡Department of Statistics, Virginia Tech, Blacksburg, VA 24061
§Department of Statistics, Virginia Tech, Blacksburg, VA 24061
¶Corresponding author.

 https://doi.org/10.1214/23-BA1375

for spatial random effects have been used for inference and prediction in fields such as ecology (Ver Hoef et al., 2018), neuroscience (Liu et al., 2016), disease mapping (Reich et al., 2006; Lee, 2011; Jin et al., 2005; White et al., 2017), and public policy (Logan et al., 2020). While modeling and estimation methods for models with CAR and ICAR structures have seen a variety of methodological developments and applications, simultaneous selection of spatial model structure and covariates in hierarchical models with ICAR spatial random effects has seen limited development. Thus, when choosing which specific covariates to include and whether spatial dependence persists in the presence of these covariates, researchers often resort to two-stage procedures. We focus our attention on Bayesian selection methods in this work. For example, researchers have adapted selection techniques to compare various proposed CAR models (Lee, 2011; Song and De Oliveira, 2012; Best et al., 2005) and to determine the impact of covariates when a spatial correlation structure is assumed (Best et al., 1999). Such approaches require researchers to either fix the spatial model structure and select covariates or fix the covariates and assess the need for spatial model structure. The order of selection is arbitrary and has been seen in case studies to potentially provide conflicting results. For example, Lee and Mitchell (2013) studied two covariates in a disease mapping application. They first fit a Bayesian model with both covariates and no spatial random effects. In the non-spatial model, 95% credible intervals revealed both covariates were non-null, however a Moran's I test indicated spatial correlation in the residuals. Upon finding spatial correlation in the residuals, the authors fit spatial models including the Besag-York-Mollié (BYM, Besag et al., 1991), Leroux (Leroux et al., 1999), and locally adaptive spatial models. In some of these spatial models, one of the covariates was then found to be plausibly null. This sort of recursive approach can lead to uncertainty about covariates in the model since it does not directly assess spatial effects and covariates simultaneously. While Lee and Mitchell (2013) successfully developed novel methods for accommodating localized spatial dependence, this multi–stage approach that selects the mean and covariance structures in separate stages indicates a need for more cohesive and simultaneous selection methods for areal data. Bayesian methods for simultaneous selection of spatial model structure and covariates have seen less development. Since simultaneous selection of spatial model structure and regressors would provide a framework for researchers to make concurrent probabilistic decisions in spatial contexts, we propose a Bayesian approach for simultaneous selection of fixed effects and spatial model structure in Gaussian hierarchical models with ICAR priors.

Current literature on model selection for hierarchical models with ICAR priors has suffered from the crucial limitation that, until recently, these models did not have a fully specified expression for the likelihood function with integrated out random effects. Without such an expression, the development of formal Bayesian model selection was not possible. Fortunately, Keefe et al. (2018) recently proposed a formal specification of sum-zero constrained ICAR models that fully specifies the constant of proportionality in these models. We explore these recent results to develop formal Bayesian model selection for hierarchical models with ICAR random effects.

To the best of our knowledge, the closest published work related to our proposed method is by Song and De Oliveira (2012). Specifically, Song and De Oliveira (2012) proposed the use of posterior model probabilities to choose between classes of Gaussian CAR models and simultaneous autoregressive (SAR) models with default priors for model parameters. However, their proposed methodology differs from ours in two important ways. First, Song and De Oliveira (2012) considered CAR and SAR models as direct models for the observations, whereas we consider the more usual framework in the statistical literature of using CAR priors for spatial random effects in a hierarchical model. Second, their approach assumes that all competing models have the same mean structure, and it is restricted to model selection of covariance structure. Hence, their approach cannot perform covariate selection. In contrast, we provide an approach to perform joint selection of covariates (fixed effects) and spatial model structure.

There are several published criteria other than posterior probabilities for model selection in hierarchical models with CAR and ICAR spatial random effects. Currently, the most frequently used model selection criterion for such hierarchical models is the Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)). For example, the DIC has been used for model selection in the contexts of generalized multivariate CAR models (Jin et al., 2005), co-regionalized models for areal data (Jin et al., 2007), locally adaptive spatial CAR models (Lee and Mitchell, 2013), and disease mapping (Martinez-Beneito et al., 2017). Another model comparison tool, proposed by White et al. (2017) for hierarchical models with CAR random effects, is cross-validation, where a part of the sample is used for model fitting and the other observations are held out for model evaluation. In such cross-validation setting, White et al. (2017) proposed as model comparison criteria predictive interval coverage, predictive mean square error, and predictive mean absolute error. Although ingenious and practically useful, these published model selection criteria do not provide the natural and straightforward quantification of model uncertainty provided by posterior model probabilities.

To enable simultaneous selection of fixed effects and spatial model structure, we develop a Bayesian model selection method for hierarchical models with an ICAR component. In particular, we examine a sum-zero constrained ICAR prior for spatial random effects in a Bayesian hierarchical model (Keefe et al., 2018). We devise a fractional Bayes factor (O'Hagan, 1995) approach for model selection via posterior model probabilities. Fractional Bayes factors use a portion of the likelihood to update priors on parameters, which enables our automatic Bayesian model selection with an improper reference prior on model parameters. Model selection consistency, which refers to the method's ability to select the true model as sample size increases if the true model is in the candidate set, is a well-known result when using fractional Bayes factors (O'Hagan, 1995). Thus our approach provides consistent, simultaneous selection of fixed effects and spatial model structure in Bayesian hierarchical models and allows for direct probabilistic statements about inclusion of covariates and spatial model structure.

We describe our formal Bayesian model selection approach in the following sections. In Section 2 we introduce the hierarchical model with a sum-zero constrained ICAR prior, prove an equivalence result for two ICAR specifications, and provide the reference prior we consider for the parameters of the ICAR component. In Section 3 we present the motivation and implementation of our proposed method which uses fractional Bayes factors for simultaneous selection of fixed effects and spatial model structure in Gaussian hierarchical models with ICAR priors. In Section 4 we study the performance of our proposed method with a simulation study that includes varying levels of spatial dependence. Section 5 demonstrates two applications of our method, including median income and socioeconomic data at the county-level for the contiguous United States (US) in 2017 and residential crime rates in the neighborhoods of Columbus, Ohio (OH) in 1980. Finally, in Section 6 we discuss the practical impact of our fractional Bayes factor approach and future avenues for research. Proofs of theoretical results and additional model selection background and simulation results are provided in the Supplementary Material (Porter et al., 2023).

2 Hierarchical Model Specification

We consider a hierarchical model for areal data measured over a contiguous region that is partitioned into n disjoint subregions indexed by $1, \ldots, n$. Consider the following hierarchical model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\phi}, \tag{1}$$

where $\mathbf{Y}$ is a $n \times 1$ response vector, X is a $n \times p$ matrix of covariates, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients corresponding to fixed effects. Additionally, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T$ is a $n \times 1$ vector of independent unstructured random effects with distribution $N(\mathbf{0}, \sigma^2 I_n)$ and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)^T$ is a $n \times 1$ vector of spatial random effects. The first column of the matrix X is assumed to be a vector of ones and, thus, the first element of $\boldsymbol{\beta}$ is an intercept. The vectors of random effects, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, are assumed to be independent *a priori*.

The spatial random effects $\boldsymbol{\phi}$ are assigned the formal sum-zero constrained intrinsic conditional autoregressive (ICAR) prior (Keefe et al., 2018, 2019). We consider a signal-to-noise ratio parameterization where τ/σ^2 represents the precision for the vector of spatial random effects, where the parameter $\tau \in (0, \infty)$ controls the strength of spatial dependence and σ^2 denotes the variance of the unstructured random effects. Small values of τ indicate strong spatial dependence while values tending towards infinity correspond to independent data. The sum-zero constraint $\sum_{i=1}^n \phi_i = 0$ appears explicitly in the density for $\boldsymbol{\phi}$ as follows.

$$p(\boldsymbol{\phi}|\sigma^2, \tau) = (2\pi)^{-(n-1)/2} \Big(\frac{\tau}{\sigma^2}\Big)^{(n-1)/2} \left(\prod_{i=1}^{n-1} d_i\right)^{1/2} \exp\Big\{ -\frac{\tau}{2\sigma^2} \boldsymbol{\phi}^T H \boldsymbol{\phi}\Big\} \mathbb{1}(\mathbf{1}_n^T \boldsymbol{\phi} = 0), \tag{2}$$

where $\mathbf{1}_n$ is a vector of ones and H is a positive semi-definite precision matrix defined as

$$(H)_{ij} = \begin{cases} h_i, & \text{if } i = j, \\ -g_{ij}, & \text{if } i \in N_j, \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

and $d_1 \geq d_2 \geq \ldots \geq d_{n-1} > d_n = 0$ are the ordered eigenvalues of H. The matrix H is fixed and is chosen by the researcher to specify the neighborhood structure of the study region. For example, a common choice for H classifies two subregions as neighbors if they share a border. In that case, $\{N_j; j = 1 \ldots n\}$ denotes the set of regions that are neighbors to region j, h_i indicates the total number of neighbors of region i, and $g_{ij} = 1$ if regions i and j are neighbors and $g_{ij} = 0$ if regions i and j are not neighbors. We assume that there are no islands in the region of interest, that is, all of the subregions are connected. As a consequence, H has rank $n-1$ and one null eigenvalue (e.g., see Ferreira and De Oliveira, 2007; De Oliveira and Ferreira, 2011). Let the spectral decomposition of H be $H = QDQ^T$, where $Q = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n)$ is a $n \times n$ matrix whose columns are the normalized eigenvectors of H and $D = \text{diag}(d_1, d_2, \ldots, d_n)$ is a diagonal matrix with the ordered eigenvalues of H along the diagonal. Let $\tilde{Q} = (\mathbf{q}_1, \ldots, \mathbf{q}_{n-1})$. Then the distribution of the spatial random effects $\boldsymbol{\phi}$ can be written as the following singular Gaussian distribution (Keefe et al., 2018, 2019):

$$\boldsymbol{\phi} \sim N\Big(\mathbf{0}, \frac{\sigma^2}{\tau}\Sigma_\phi\Big), \tag{4}$$

where $\Sigma_\phi = \tilde{Q}\text{diag}(d_1^{-1}, \ldots, d_{n-1}^{-1})\tilde{Q}^T$ is the Moore-Penrose pseudoinverse (Penrose, 1955) of the precision matrix H. See Auxiliary Fact A3 in the Supplementary Material for details concerning how the sum-zero constraint is represented in (2) and (4).

2.1 Equivalence Between ICAR Specifications

There are three main ways to impose the sum-zero constraint on the ICAR prior. The first way is through centering on the fly, where the spatial random effects are centered to sum to zero at each iteration of the Markov Chain Monte Carlo (MCMC) algorithm. The second way is through obtaining the full conditional distribution of the spatial random effects, which as we explain below is a proper multivariate Gaussian distribution, and then use standard multivariate Gaussian results to obtain the full conditional distribution of the spatial random effects conditional on their sum being equal to zero. Finally, the third way is to use the sum-zero constrained ICAR model proposed by Keefe et al. (2018, 2019). Ferreira (2019) and Ferreira et al. (2021) have shown that the first and the third ways are equivalent for Gaussian hierarchical models. In this section, we show that the second and third ways are equivalent.

The following propositions and theorem present results about the distribution of spatial random effects $\boldsymbol{\phi}$ and show the equivalence between sampling from the improper ICAR prior conditional on a sum-zero constraint and sampling from the formal sum-zero constrained prior.

Proposition 2.1 (Ferreira et al. (2021))**.** *Assume the hierarchical model given by (1) and (2). Partition the design matrix as $X = [\mathbf{1}_n, F]$ and, similarly, partition the vector of regression coefficients as $\boldsymbol{\beta} = (\alpha, \boldsymbol{\nu}^T)^T$ where α is an intercept. Then, the full conditional distribution of $\boldsymbol{\phi}$ is*

$$\boldsymbol{\phi}|\tau, \sigma^2, \mathbf{Y}, \boldsymbol{\beta} \sim N(\tilde{Q}\mathbf{s}, \sigma^2\tilde{Q}D^*\tilde{Q}^T), \tag{5}$$

where $D^* = diag((1+\tau d_1)^{-1}, \ldots, (1+\tau d_{n-1})^{-1})$, and $\mathbf{s} = D^*\tilde{Q}^T(\mathbf{Y} - F\boldsymbol{\nu})$.

Next, let $\boldsymbol{\omega}$ be a vector of spatial random effects that *a priori* follows the improper ICAR prior (Besag et al., 1991). Then, the prior density for $\boldsymbol{\omega}$ is defined up to a constant of proportionality and, with the signal-to-noise ratio parameterization, is given by

$$p(\boldsymbol{\omega}|\tau,\sigma^2) \propto \exp\Big\{-\frac{\tau}{2\sigma^2}\boldsymbol{\omega}^T H\boldsymbol{\omega}\Big\}. \tag{6}$$

If we substitute $\boldsymbol{\phi}$ by $\boldsymbol{\omega}$ in (1) and use the prior for $\boldsymbol{\omega}$ given in (6), straightforward application of Bayes' Theorem yields the full conditional distribution (e.g., see Ferreira et al., 2021)

$$\boldsymbol{\omega}|\tau,\sigma^2,\mathbf{Y},\boldsymbol{\beta} \sim N(g, V), \tag{7}$$

where $V = \sigma^2(I_n + \tau H)^{-1} = \sigma^2 Q(I_n + \tau D)^{-1}Q^T$ and $g = (I_n + \tau H)^{-1}(\mathbf{Y} - X\boldsymbol{\beta}) = Q(I_n + \tau D)^{-1}Q^T(\mathbf{Y} - X\boldsymbol{\beta})$. Note that $(I_n + \tau H)$ is diagonally dominant and, therefore, non-singular. Hence, the matrix V is well defined. However, assigning a flat prior for the intercept in the model and the improper CAR prior given in (6) leads to an improper posterior distribution for ω.

Alternatively to sampling the sum-zero-constrained spatial random effects vector $\boldsymbol{\phi}$ from its full conditional distribution (5), we may consider using the Besag spatial random effects vector $\boldsymbol{\omega}$ sampled from its full conditional distribution (7) conditional on the constraint $\mathbf{1}^T\boldsymbol{\omega} = 0$. The details of this distribution are given in Proposition 2.2.

Proposition 2.2. *Assume the model given by (1) but substituting $\boldsymbol{\phi}$ by $\boldsymbol{\omega}$. In addition, assume for $\boldsymbol{\omega}$ the prior given by (6). Then, the full conditional distribution of $\boldsymbol{\omega}$ conditional on the constraint $\mathbf{1}_n^T\boldsymbol{\omega} = 0$ is*

$$\boldsymbol{\omega}|\mathbf{1}_n^T\boldsymbol{\omega} = 0, \tau, \sigma^2, \mathbf{Y}, \boldsymbol{\beta} \sim N(\mu^*, \Sigma^*), \tag{8}$$

where $\mu^* = g - V\mathbf{1}_n(\mathbf{1}_n^T V\mathbf{1}_n)^{-1}(\mathbf{1}_n^T g - 0)$ and $\Sigma^* = V - V\mathbf{1}_n(\mathbf{1}_n^T V\mathbf{1}_n)^{-1}\mathbf{1}_n^T V$.

Proof. See Proofs of Main Results in the Supplementary Material. □

The following theorem shows the equivalence between sampling from the full conditional distribution of $\boldsymbol{\phi}$ implied by the sum-zero constrained ICAR prior of Keefe et al. (2018, 2019) and sampling from the full conditional distribution of the spatial random effects $\boldsymbol{\omega}$ implied by the improper ICAR prior of Besag et al. (1991) with respect to the sum-zero constraint, that is, sampling $\boldsymbol{\omega}$ conditional on the spatial random effects summing to zero.

Theorem 2.1. *Assume that all subregions are connected. Then, sampling from the full conditional distribution given in (8) of the spatial random effects $\boldsymbol{\omega}$ implied by the improper ICAR prior (6) conditional on the sum-zero constraint $\mathbf{1}^T\boldsymbol{\omega} = 0$ is equivalent to sampling from the full conditional distribution given in (5) implied by the sum-zero constrained ICAR prior given in (2).*

Proof. See Proofs of Main Results in the Supplementary Material. □

The equivalence result given in Theorem 2.1 is of fundamental importance because it implies that the model selection approach we propose can be applied to Gaussian hierarchical models with Besag ICAR spatial random effects such as implemented in the widely used R package R-INLA, which implements the Integrated Nested Laplace Approximation (INLA) method by Rue et al. (2009).

Next, we denote the vector of unknown model parameters for the hierarchical spatial model by $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma^2, \tau)$. Following the approach of Keefe et al. (2018), we impose the formal sum-zero constraint that implies as prior for the spatial random effects $\boldsymbol{\phi}$ the singular Gaussian distribution given in (4). After that, we integrate out the vector of spatial random effects $\boldsymbol{\phi}$ to obtain for $\mathbf{Y}|\boldsymbol{\eta}$ the Gaussian distribution

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

We are interested in selecting which covariates to include in the model by choosing among competing $X\boldsymbol{\beta}$, and whether to include spatial dependence. We use the ordinary linear model (OLM) as the independent data model, which does not accommodate spatial correlation. In the case of the OLM, the unknown parameters are $\boldsymbol{\beta}$ and σ^2, and the response $\mathbf{Y}$ follows the Gaussian distribution

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma^2 \sim N(X\boldsymbol{\beta}, \sigma^2 I_n). \tag{10}$$

2.2 Priors on Model Parameters

We adopt a Bayesian approach and specify priors for $\boldsymbol{\eta}$. We consider the recently proposed reference prior for the parameters $\boldsymbol{\beta}$, σ^2, and τ of the hierarchical model given in (1) and (4) (Keefe et al., 2019), which serves as an automatic prior with favorable properties for inference in Gaussian hierarchical models with ICAR priors. The joint reference prior for $\boldsymbol{\eta}$ in the hierarchical spatial model is given by

$$\pi(\boldsymbol{\beta}, \sigma^2, \tau) \propto \frac{1}{\sigma^2}\frac{1}{\tau}\left[\sum_{j=1}^{n-p}\left(\frac{\xi_j}{\tau+\xi_j}\right)^2 - \frac{1}{n-p}\left\{\sum_{j=1}^{n-p}\left(\frac{\xi_j}{\tau+\xi_j}\right)\right\}^2\right]^{\frac{1}{2}}, \tag{11}$$

where $\xi_1, \ldots, \xi_{n-p}$ are the ordered eigenvalues of $Q^{*T}\Sigma_\phi Q^*$ such that the columns of Q^* are normalized eigenvectors corresponding to the non-zero eigenvalues of the projection matrix $G = I_n - X(X^TX)^{-1}X^T$. The prior on σ^2 is $\pi(\sigma^2) \propto 1/\sigma^2$, and the conditional reference prior on the vector of regression coefficients is $\pi(\boldsymbol{\beta}|\sigma^2, \tau) \propto 1$. Thus $\pi(\tau)$ takes the form of (11) excluding $1/\sigma^2$. Note that improper priors must be treated carefully when used for model selection, which we address further in Section 3.

Subjective information for setting hyperparameters is not always available and expert elicitation is challenging, as evidenced by the absence of such approaches in the spatial statistics literature. The reference prior obviates the need to choose hyperparameters for priors in hierarchical models for areal data and has been shown to perform well for estimation in spatial ICAR models. Keefe et al. (2019, Section 5 and supplementary material) show that inference procedures based on the reference prior have favorable performance in terms of frequentist coverage rate, average interval length, and mean squared error (MSE) for $\boldsymbol{\beta}$, σ^2, and τ. Thus, the reference prior in (11) can be used to reliably estimate all model parameters in $\boldsymbol{\eta}$ and to identify appropriate subsets of covariates. Finally, for the OLM, the joint reference prior for $\boldsymbol{\eta}$ is $\pi(\boldsymbol{\eta}) \propto 1/\sigma^2$. In Section 3 we describe how the reference prior in (11) can be used with fractional Bayes factors to simultaneously select spatial model structure and covariates.

3 Bayesian Model Selection via Fractional Bayes Factors

We next describe simultaneous Bayesian model selection for spatial dependence and covariates based on models (9) and (10). Bayesian model selection relies on integrated likelihoods of the form

$$p(\mathbf{Y}|M_c) = \int p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c|M_c)d\boldsymbol{\eta}_c, \tag{12}$$

where model M_c has corresponding parameter vector $\boldsymbol{\eta}_c$ and $c = 1, \dots, C$.

To compare two models M_1 and M_2, we may use the Bayes factor BF_{12} that is defined as a ratio of the two models' integrated likelihoods

$$BF_{12} = \frac{p(\mathbf{Y}|M_1)}{p(\mathbf{Y}|M_2)}. \tag{13}$$

To alleviate numerical underflow when forming all posterior model probabilities, we form all Bayes factors with respect to a single baseline model, M_l, which has the largest integrated likelihood of all models in the model space $\mathcal{M} = \{M_c,\ c = 1, \dots, C\}$, where C is the total number of candidate models. From the set of Bayes factors $\{BF_{1l}, \dots, BF_{Cl}\}$ formed with respect to baseline M_l, the posterior probability of a single model M_r in the model space can then be found using Bayes' Rule:

$$P(M_r|\mathbf{Y}) = \frac{p(\mathbf{Y}|M_r)P(M_r)}{\sum_{c=1}^{C} p(\mathbf{Y}|M_c)P(M_c)} = \left(\sum_{c=1}^{C} BF_{cl}P(M_c)\right)^{-1} \times BF_{rl} \times P(M_r). \tag{14}$$

Formulation of posterior model probabilities in (14) requires prior probabilities to be assigned to the competing models. For a moment, consider covariate subset selection for K total candidate predictors in the class of OLMs, where $K = p - 1$. We follow the recommendations of Scott and Berger (2010) by first assigning a uniform prior on all groups of models with a fixed number of covariates k, then evenly splitting the share of probability among models in that set. For example, if $K = 2$, then the candidate models all include either zero, one, or two covariates. The model with zero covariates receives $1/3$ of the prior probability, as does the model with two covariates. Each model with a single covariate receives $1/6$ of the prior model probability. This approach imparts Bayesian multiplicity correction to the selection procedure. We modify the subset selection approach slightly to accommodate both OLMs and spatial models. In this work, we set prior probability for independence at $1/2$, and also give $1/2$ prior probability to spatial dependence. Thus, we further divide prior probabilities suggested by Scott and Berger (2010) for the independence models in half, and incorporate all possible subset models with the inclusion of spatial dependence, so that the full candidate model set contains all possible combinations of candidate predictors in both spatial dependence and independence settings. Then the prior probability for a model M_c with k_c covariates is

$$P(M_c) = \frac{1}{2(K+1)}\binom{K}{k_c}^{-1}. \tag{15}$$

We develop Bayesian model selection via fractional Bayes factors with the following motivation in mind. If improper priors are assigned to parameters in competing models, the full Bayes factor is defined only up to an undefined constant and thus cannot be used for valid model comparison. In particular, the conditional reference prior on $\boldsymbol{\beta}$ is improper and cannot be used in the full Bayes factor when we consider selection of covariates. For example, consider a comparison between two candidate models, M_1 and M_2, where M_1 represents an intercept-only spatial model and M_2 represents a spatial model with $k \geq 1$ covariates. Then $\pi(\boldsymbol{\eta}) = a_1 \cdot \frac{1}{\sigma^2}\pi(\tau)$ for M_1 and $\pi(\boldsymbol{\eta}) = a_2 \cdot \frac{1}{\sigma^2}\pi(\tau)$ for M_2, where $a_1 \neq a_2$ due to differing sizes for $\boldsymbol{\beta}$. Then the Bayes factor, BF_{12}, becomes

$$BF_{12} = \frac{a_1 \cdot \int p(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \tau, M_1)\pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau)d\boldsymbol{\beta}d\sigma^2 d\tau}{a_2 \cdot \int p(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \tau, M_2)\pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau)d\boldsymbol{\beta}d\sigma^2 d\tau}, \tag{16}$$

where a_1/a_2 is an undefined constant and thus BF_{12} is not well-defined. In addition, while σ^2 appears in every model in $\mathcal{M}$, use of the full Bayes factor with $\pi(\sigma^2) \propto 1/\sigma^2$ tacitly assumes the normalizing constant for $\pi(\sigma^2)$ is the same across all models, where σ^2 may include variation from important regressors that are missing in some models. Thus, the same improper prior on σ^2 for models with different specifications of $\boldsymbol{\beta}$ may not be reasonable when performing model selection via full Bayes factors. Assigning proper priors to all model parameters elicits a proper Bayes factor as defined in (13). However, specification of sensible priors for spatial dependence models is difficult. The sensitivity of Bayes factors and the resulting model selection to hyperparameter specification is well established (Kass and Raftery, 1995; Berger and Pericchi, 1996; Chipman et al., 2001; Franck and Gramacy, 2020).

Rather than approaching Bayesian model selection with proper priors, approaches that use training samples to calibrate reference improper priors, including partial Bayes factors and fractional Bayes factors (FBF), have been proposed (O'Hagan, 1995; Berger and Pericchi, 1996). The partial Bayes factor separates out a subset of the data as a training sample, which is then used to update the priors on parameters. The partial Bayes factor uses for training the joint distribution of the specific observations selected for training. Selecting training observations for correlated data is difficult, as a subset of randomly selected points may not contain much information about the dependence structure and the τ parameter. The underlying Markovian structure may be lost by splitting the likelihood based on spatially correlated observations between training and selection, and training based on observations that do not properly reflect the overall dependence structure may result in poor model selection. The intrinsic Bayes factor averages over partial Bayes factors obtained from some or all possible training samples (Berger and Pericchi, 1996). However, this process is computationally expensive and it is not clear if all possible minimal training sets would have the same size necessary to capture the dependence structure. To overcome this difficulty, we develop a FBF approach, which uses a fraction of the likelihood rather than reserving specific observations for training. We thus use FBF methodology to approximate partial Bayes factors. The FBF updates the prior on model parameters $\pi(\boldsymbol{\eta}_c)$ using a fraction $b = m/p$ of the likelihood, obtaining the updated prior

$$\pi^*(\boldsymbol{\eta}_c) = \frac{\pi(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^b}{\int \pi(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^b d\boldsymbol{\eta}_c}. \tag{17}$$

Then the fractional integrated likelihood, $q_c(b, \mathbf{Y})$, for a single model M_c using the updated prior is

$$q_c(b, \mathbf{Y}) = \int \pi^*(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^{1-b} d\boldsymbol{\eta}_c, \tag{18}$$

where $p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^{1-b}$ is the likelihood remaining to calculate the fractional integrated likelihood. Finally, note that we can rewrite the fractional integrated likelihood as

$$\begin{aligned} q_c(b, \mathbf{Y}) &= \int \frac{\pi(\boldsymbol{\eta}_c)p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\{p^{1-b}(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}d\boldsymbol{\eta}_c}{\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c} \\ &= \frac{\int p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c}{\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c}. \end{aligned} \tag{19}$$

Since the fractional integrated likelihood uses a ratio of two integrals each containing the same $\pi(\boldsymbol{\eta}_c)$, all undefined normalizing constants discussed in (16) cancel out in the computation of the fractional integrated likelihood. The FBF, which we denote by BF_{12}^b, for two models M_1 and M_2 is defined as the ratio of two fractional integrated likelihoods. That is,

$$BF_{12}^b = \frac{q_1(b, \mathbf{Y})}{q_2(b, \mathbf{Y})}. \tag{20}$$

We use this FBF approach to form posterior model probabilities that are model selection consistent (O'Hagan, 1995), which we demonstrate with a simulation study in Section 4.

3.1 Parameter Estimation and Model Selection Under the FBF

An advantage of a fractional Bayes framework is that the trained prior, denoted $\pi^*(\boldsymbol{\eta}_c)$ for parameter vector $\boldsymbol{\eta}_c$, multiplied by the likelihood after training, is proportional to the same posterior distribution as if $\pi(\boldsymbol{\eta}_c)$ is used with the full likelihood. That is,

$$\begin{aligned} \pi^*(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^{1-b} &= \frac{\pi(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^b}{\int \pi(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^b d\boldsymbol{\eta}_c} \times \{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^{1-b} \\ &\propto \pi(\boldsymbol{\eta}_c)\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^b\{p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\}^{1-b} \\ &\propto \pi(\boldsymbol{\eta}_c)p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c) \\ &\propto p(\boldsymbol{\eta}_c|\mathbf{Y}, M_c). \end{aligned} \tag{21}$$

Thus, point and interval estimation for parameters $\boldsymbol{\eta}_c$ is unchanged when using a FBF approach for model selection, and the trained prior that results from using the FBF resolves a tension for use of priors for either estimation or model selection. Therefore, researchers may select prior $\pi(\boldsymbol{\eta}_c)$ for the purpose of parameter estimation and benefit from a FBF approach with $\pi^*(\boldsymbol{\eta}_c)$ for model selection without conflict.

3.2 Integrated Likelihood Methods

To form the fractional integrated likelihood $q_c(b, \mathbf{Y})$ of a single model M_c using a FBF approach, we need both $\int p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$ and $\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$. Under the Gaussian hierarchical model with reference prior presented in Section 2.2, parameters $\boldsymbol{\beta}$ and σ^2 can be analytically integrated out of the integrated likelihood, but τ must be integrated out via approximation methods. Therefore, the integrated likelihoods for the independent model have tractable expressions. Note that, since we use a prior for τ that depends on a projection of the matrix of covariates for a given model, the fractional integrated likelihood for model M_c depends on the model-specific matrix of covariates. That is, $\boldsymbol{\beta}_c$ is the vector of regression coefficients for covariates contained in M_c, X_c is the matrix of covariates corresponding to $\boldsymbol{\beta}_c$, p_c is the length of $\boldsymbol{\beta}_c$, and $\xi_{c1}, \ldots, \xi_{c,n-p}$ are the ordered eigenvalues of $Q_c^{*T}\Sigma_\phi Q_c^*$ such that the columns of Q_c^* are normalized eigenvectors corresponding to the non-zero eigenvalues of the projection matrix $I_n - X_c(X_c^TX_c)^{-1}X_c^T$. Then the denominator of the fractional integrated likelihood in (19) for a spatial model under the reference prior reduces to the one-dimensional integral

$$\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c \propto \int_0^\infty |\Omega|^{-\frac{b}{2}}|X_c^T\Omega^{-1}X_c|^{-\frac{1}{2}}(\tau)^{-1}\left[\frac{b}{2}S_c^2\right]^{\frac{p_c-nb}{2}} \times \qquad (22)$$
$$\left[\sum_{j=1}^{n-p_c}\left(\frac{\xi_{cj}}{\tau+\xi_{cj}}\right)^2 - \frac{1}{n-p_c}\left\{\sum_{j=1}^{n-p_c}\left(\frac{\xi_{cj}}{\tau+\xi_{cj}}\right)\right\}^2\right]^{\frac{1}{2}} d\tau,$$

where $\Omega = I_n + \tau^{-1}\Sigma_\phi$ and $S_c^2 = \mathbf{Y}^T(\Omega^{-1} - \Omega^{-1}X_c(X_c^T\Omega^{-1}X_c)^{-1}X_c^T\Omega^{-1})\mathbf{Y}$.

We use an adaptive quadrature approach to approximate integrals $\int p(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$ and $\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$ over τ for the FBF. See the Supplementary Material for further details and complete integrated likelihood expressions for the OLM and spatial ICAR model. Next, Section 3.3 discusses the choice of training fraction, specifically the minimal training fraction that will make the integrals in the FBF finite so that adaptive quadrature can be applied.

3.3 FBF Training Fraction

The training fraction for the FBF should be chosen to be small while still ensuring propriety of the fractional integrated likelihood. Consider a training fraction of the form $b = m/n$, where m is the corresponding training size. The minimal training size, which we use in this work, is the smallest integer value for m such that $\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$ is finite for all models considered in $\mathcal{M} = \{M_c,\ c = 1, \ldots, C\}$. In particular, if m is chosen to be too small, the integral $\int p^b(\mathbf{Y}|\boldsymbol{\eta}_c, M_c)\pi(\boldsymbol{\eta}_c)d\boldsymbol{\eta}_c$ will diverge for one or more models in $\mathcal{M}$ and the corresponding FBF cannot be formed for all models. Additionally, as m increases beyond the minimal training size, the posterior model probabilities more closely resemble the prior model probabilities. Over–training the prior at the cost of the likelihood forfeits statistical power and reduces the ability of the FBF to detect signal. Thus, the minimal training size should be used when known for a class of models. To understand the behavior of the denominator of the fractional integrated likelihood for the ICAR model, we first consider expressions leading to the integral over τ as in (22) by using the eigenvalue decomposition of functions of X and Σ_ϕ.

The following propositions outline the fractional integrated likelihood results which lead to the minimal training size for the FBF with the reference prior for spatial ICAR models. The above text has demonstrated that it is essential for the FBF-trained prior to be proper. The purpose of the upcoming Propositions 3.1 and 3.2 is to find the smallest value of the training size m that will yield a proper FBF-trained prior, overcome the arbitrary constant issue in (16), and lead to valid Bayesian model selection. Proposition 3.1 addresses propriety with respect to σ^2 in the denominator of the fractional integrated likelihood and in the updated prior (see (19) and (17)). For notational convenience, let $p^{(b)}(\mathbf{Y}|\sigma^2,\tau,M) = \int p^b(\mathbf{Y}|\boldsymbol{\beta},\sigma^2,\tau,M)\pi(\beta)d\boldsymbol{\beta}$ and $p^{(b)}(\mathbf{Y}|\tau,M) = \int\int p^b(\mathbf{Y}|\boldsymbol{\beta},\sigma^2,\tau,M)\pi(\boldsymbol{\beta})\pi(\sigma^2)d\boldsymbol{\beta}d\sigma^2$.

Proposition 3.1. *To ensure $\int p^b(\mathbf{Y}|\boldsymbol{\eta},M)\pi(\boldsymbol{\eta})d\boldsymbol{\eta} < \infty$, consider the tail behavior of $p^{(b)}(\mathbf{Y}|\sigma^2,\tau,M)\pi(\sigma^2)$ over σ^2. For a given value of τ,*

(i) $p^{(b)}(\mathbf{Y}|\sigma^2,\tau,M)\pi(\sigma^2) = O\big((\sigma^2)^{\frac{p-nb}{2}-1}\big)$ *as* $\sigma^2 \to \infty$.

(ii) $p^{(b)}(\mathbf{Y}|\sigma^2,\tau,M)\pi(\sigma^2) = O\big(\exp\{\frac{-b}{2\sigma^2}S^2\}\big)$ *as* $\sigma^2 \to 0$.

Proof. See Proofs of Main Results in Supplementary Material. □

Next, Proposition 3.2 addresses propriety of the trained prior and $p^{(b)}(\mathbf{Y}|\tau,M)$ with respect to τ after both $\boldsymbol{\beta}$ and σ^2 have been integrated out analytically.

Proposition 3.2. *Assume $\frac{nb-p}{2} > 0$. To ensure $\int p^b(\mathbf{Y}|\boldsymbol{\eta},M)\pi(\boldsymbol{\eta})d\boldsymbol{\eta} < \infty$, consider the behavior of $p^{(b)}(\mathbf{Y}|\tau,M)$ over τ. $p^{(b)}(\mathbf{Y}|\tau,M)$ is a continuous function on $\tau \in (0,\infty)$ and*

(i) $p^{(b)}(\mathbf{Y}|\tau,M) = O(\tau^{\frac{1-b}{2}+1})$ *as* $\tau \to 0$.

(ii) $p^{(b)}(\mathbf{Y}|\tau,M) = O(1)$ *as* $\tau \to \infty$.

Proof. See Proofs of Main Results in Supplementary Material. □

Proposition 3.3 demonstrates that the reference prior for τ is proper.

Proposition 3.3 (Keefe et al. (2019))**.** *The marginal reference prior for τ is a continuous function on $(0,\infty)$ where*

(i) $\pi(\tau) = O(1)$ *as* $\tau \to 0$.

(ii) $\pi(\tau) = O(\tau^{-2})$ *as* $\tau \to \infty$.

Theorem 3.1 provides the minimum training size for the application of our FBF approach.

Theorem 3.1. *Consider model (9) and the reference prior in (11). The minimal training size for the FBF is $m = p + 1$.*

Proof. See Proofs of Main Results in Supplementary Material. □

Finally, it is straightforward to show that results similar to that of Theorem 3.1 also hold for the OLM with a reference prior as well as for the SAR model for spatial areal data with the independence Jeffreys prior developed by De Oliveira and Song (2008). For both of these models, the minimal training size for the FBF is also $m = p + 1$.

4 Simulation Study

To investigate the utility of our FBF approach to simultaneously select covariates and presence of spatial dependence, we perform Monte Carlo simulations for 100 square grid regions of size $n = 10^2, 20^2, 30^2$. To study selection performance for a variety of spatial settings, we examine varying levels of the signal-to-noise ratio. In particular, we fix $\sigma^2 = 1$ for simulations and consider $\tau = 0.01, 0.1, 1, 10, \infty$ for the response, where small values of τ correspond to strong spatial dependence, and infinite τ corresponds to independence. Further, we consider $k = 5$ covariates with $\boldsymbol{\beta} = (5, 2, 1, 0, 0, 0)^T$. Since many spatial applications also contain spatially correlated covariates, covariates are generated from a model of the form (9) with mean 0, $\sigma^2 = 1$, and strong spatial correlation where $\tau = 0.1$ for each covariate. We obtain model selection results based on 100 simulated data sets for each combination of these levels of sample size and spatial dependence.

For each simulated data set our method computes posterior model probabilities via FBFs for each of the $2^6 = 64$ models, including 32 OLMs and 32 spatial models with all possible combinations of the $k = 5$ covariates. Additionally, since we use Bayesian model selection, posterior inclusion probabilities for individual fixed effects are easy to calculate. Figure 1 displays boxplots of the posterior inclusion probabilities for each of the 5 covariates alongside the probability of selecting a spatial model at each sample size. Red diamonds correspond to the mean probability across the 100 data sets represented in each boxplot. Overall our method correctly assigns high posterior model probability to the correct model, both in terms of spatial structure and identification of null and non-null covariates. For small τ the true model contains spatial dependence, and as sample size increases the probability of selecting a spatial model quickly approaches 1 for $\tau \in \{0.01, 0.1, 1\}$ and moves toward 0 for τ at ∞, which corresponds to the OLM. Note that only τ at ∞ corresponds to true independence, reducing the model in (9) to (10). However, practically, most real spatial data sets have small τ less than 10. Thus, for larger finite values of τ, e.g. $10 \leq \tau < \infty$, the true dependence structure is spatial, but there is very little spatial signal to detect. As demonstrated in Figures 1b, 1d, and 1f, the decision to select a spatial model or OLM at $\tau = 10$ is much more equivocal than for small values of τ, due to the diminishing strength of spatial correlation between the observations. Figure 1f indicates that for $n = 900$, our method has correctly identified all spatial data sets with small $\tau \in \{0.01, 0.1, 1\}$ as requiring spatial random effects. The most difficult case of non-null covariate selection occurs in Figure 1a under $n = 100$ and the strongest level of spatial dependence with $\tau = 0.01$, where we assigned x_2 a low signal-to-noise ratio. Results for all three sample sizes indicate accurate identification of covariates, as the posterior inclusion probabilities for the three null covariates x_3, x_4, and x_5 move towards 0 from the prior, and posterior inclusion probabilities for the non-null covariates x_1 and x_2 quickly approach 1. For any single covariate x, if you sum over the model priors corresponding to each model that contains x, that probability is 1/2. Additionally, recall from Section 3 that the prior probability of spatial dependence is 1/2. The horizontal bar in Figure 1 corresponds to this prior probability at 1/2, where all posterior probabilities can be seen moving off of this prior.

These results indicate that our method tends to select the correct model in terms of both spatial and fixed effects as n increases. When data is truly independent corresponding to $\tau = \infty$, our method correctly selects the simpler OLM without spatial random effects the majority of the time, it does not attribute high probability to null covariates, and it assigns high probability to non-null covariates with relatively small signal-to-noise ratios. Performance quickly improves as the sample size increases, indicating that our FBF approach provides consistent model selection for fixed and spatial effects simultaneously.

In addition to our FBF model selection approach, we also considered the Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)) and the Widely Applicable Information Criterion (WAIC, Watanabe (2010)). The DIC and WAIC are most often computed using the likelihood that includes $\boldsymbol{\phi}$ with estimated values for the spatial random effects plugged into the criteria calculations. The likelihood for $\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \sigma^2 \sim N(X\boldsymbol{\beta} + \boldsymbol{\phi}, \sigma^2 I_n)$ follows as:

$$p(\mathbf{Y}|\boldsymbol{\phi}, \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - X\boldsymbol{\beta} - \boldsymbol{\phi})^T(\mathbf{Y} - X\boldsymbol{\beta} - \boldsymbol{\phi})\Big\}. \quad (23)$$

We also consider a version of the DIC that uses the likelihood with the spatial random effects integrated out:

$$p(\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \tau) = (2\pi)^{-n/2}(\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\Big\{ -\frac{1}{2\sigma^2}(\mathbf{Y} - X\boldsymbol{\beta})^T\Omega^{-1}(\mathbf{Y} - X\boldsymbol{\beta})\Big\}, \quad (24)$$

where $\Omega = I_n + \tau^{-1}\Sigma_\phi$. The likelihood in (24) corresponds to the distribution in (9) used in our FBF approach. This version of DIC, called type 2 DIC by Celeux et al. (2006) in the context of missing data models, was studied in Ferreira et al. (2021) for the spatial hierarchical models considered here.

We use an MCMC algorithm to compute the DIC, WAIC, and type 2 DIC for all 64 models for each data set in the simulation study as described above (Gelman et al., 2014). We use the same reference priors when computing DIC, WAIC, and type 2 DIC as for the FBF, so parameters in the ICAR models are assigned the prior in (11) and parameters in the OLMs are sampled with prior $\pi(\boldsymbol{\beta}, \sigma^2) \propto 1/\sigma^2$. For each ICAR model, we sample parameters in $\boldsymbol{\eta}$ using a Metropolis-within-Gibbs algorithm with a Gibbs step for $\boldsymbol{\beta}$ and a joint Metropolis-Hastings step for τ and σ^2 (Keefe et al., 2019). To obtain samples from likelihood (23), we simulate the spatial random effects $\boldsymbol{\phi}$ using composite sampling. The full conditional distribution for $\boldsymbol{\phi}$ and the complete algorithm for sampling from the parameters of the ICAR model are listed in the Supplementary

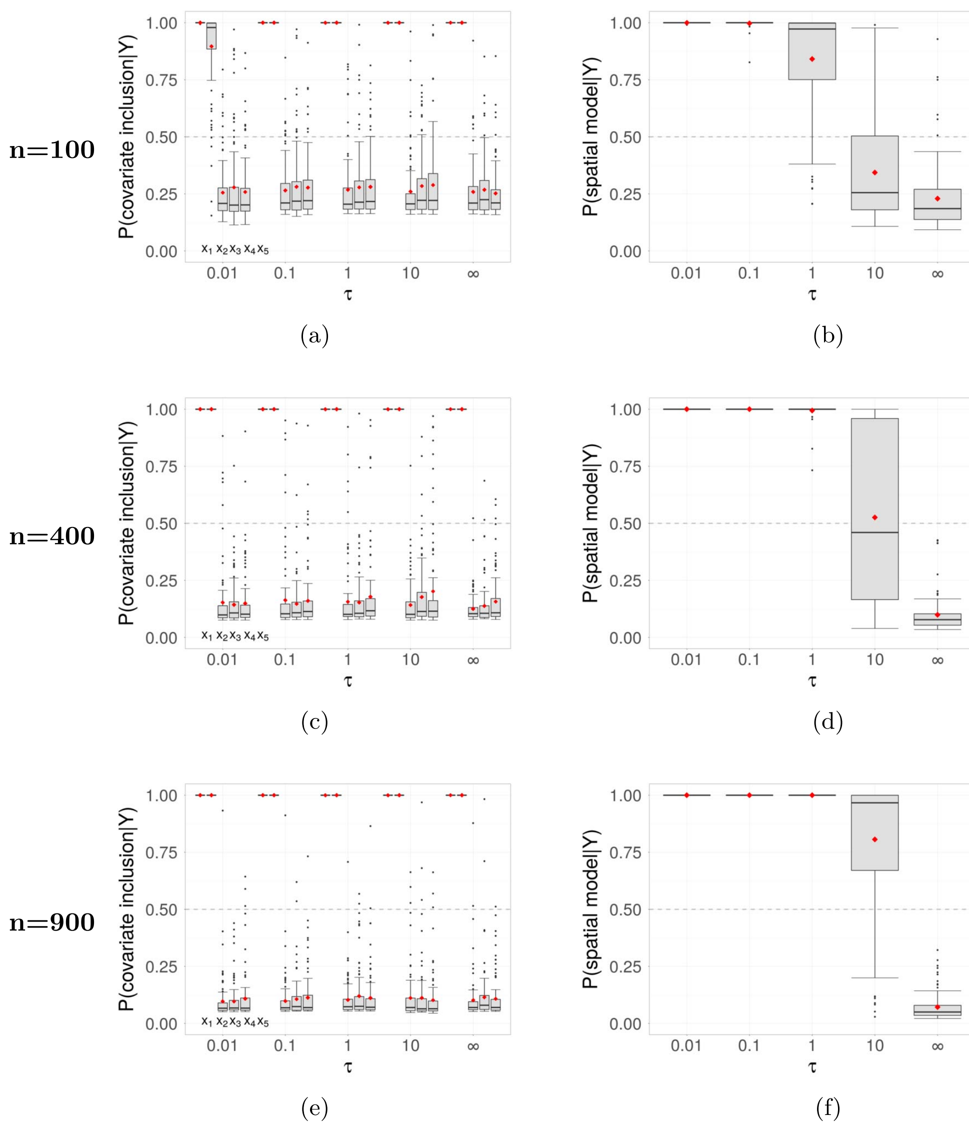


Figure 1: Covariate posterior inclusion probabilities (left) and probability of selecting a spatial model (right) for $\tau \in \{0.01, 0.1, 1, 10, \infty\}$ for n = 100 (top row), n = 400 (middle row), and n = 900 (bottom row). Each boxplot represents probabilities from 100 simulated data sets. The reference FBF selection method correctly assigns high probability to non-null covariates, and to spatial models for small τ. The posterior inclusion probabilities for non-null covariates x_1 and x_2 are exactly 1 for all simulated data sets where $n = 400$ and $n = 900$, and for $n = 100$ with $\tau > 0.01$. Thus, the boxplots for these covariates appear as lines at 1.

Material. For each OLM, we sample σ^2 from its marginal posterior $p(\sigma^2|\mathbf{Y})$ and use a Gibbs sampler to sample $\boldsymbol{\beta}$ from its conditional posterior $p(\boldsymbol{\beta}|\sigma^2, \mathbf{Y})$. For each model, 30,000 MCMC iterations are obtained with the first 10,000 iterations discarded as burn-in.

The computations for all results reported here were performed using a $2 \times \mathrm{E5} - 2683\mathrm{v}42.1\mathrm{GHz}$ (Broadwell) CPU supercomputer from Advanced Research Computing at Virginia Tech. Our FBF selection approach takes 9.28, 65.89, and 365.73 seconds for a single data set with sample size equal to 100, 400, and 900, respectively. The DIC selection approach takes 953.39; 2,491.87; and 6,315.48 seconds for all models for a single data set with sample size equal to 100, 400, and 900, respectively. The WAIC takes 1,041.75; 2,254.34; and 6,156.46 seconds for a single data set with sample size equal to 100, 400, and 900. Finally, the type 2 DIC selection approach takes 1,059.91; 2,350.69; and 6,835.21 seconds for a single data set with sample size equal to 100, 400, and 900, respectively.

For each simulated data set, we calculated the DIC, WAIC, and type 2 DIC for all 64 candidate models using MCMC and identified the model with lowest DIC, WAIC and type 2 DIC values, and the model with highest posterior model probability according to our FBF approach. Figure 2 plots the proportion of data sets for which DIC, WAIC, type 2 DIC (abbreviated as DIC-2), and our FBF approach correctly identify the correct covariate structure containing only x_1 and x_2 and the correct spatial model structure, where the true model for τ at ∞ is the OLM with no spatial random effects. As discussed above and seen in Figure 1, the true dependence structure when $\tau = 10$ is spatial, but spatial correlation is weak in this setting, making the need for spatial random effects in the model ambiguous. Therefore we compare selection results only at values of $\tau \in \{0.01, 0.1, 1, \infty\}$ in Figure 2.

Each panel of Figure 2 demonstrates that the FBF approach performs better than DIC, WAIC, and type 2 DIC for selection in all data settings considered here. Our FBF approach also successfully identifies the correct model with respect to spatial random effects more than 80% of the time for $n = 100$ at all levels of spatial dependence, and correctly identifies the spatial model structure for every data set for $n = 400$ and $n = 900$. Additional simulation results generated with covariates with no spatial dependence for the sample sizes and coefficient vector described above appear in the Supplementary Material. These results exhibit similar patterns to those in Figure 2, as the performance of the FBF is superior to that of the DIC, WAIC, and type 2 DIC for all data settings. Thus, including results provided in the Supplementary Material, for all n and τ considered here, the FBF performs better in each setting in this simulation study. This simulation study demonstrates the reliability of our fully automatic FBF approach to accurately and simultaneously select both spatial model structure and covariate structure.

5 Case Studies

To illustrate the practical application of our FBF approach to simultaneously select covariates and spatial random effects in spatial areal datasets of varying sizes, we per-

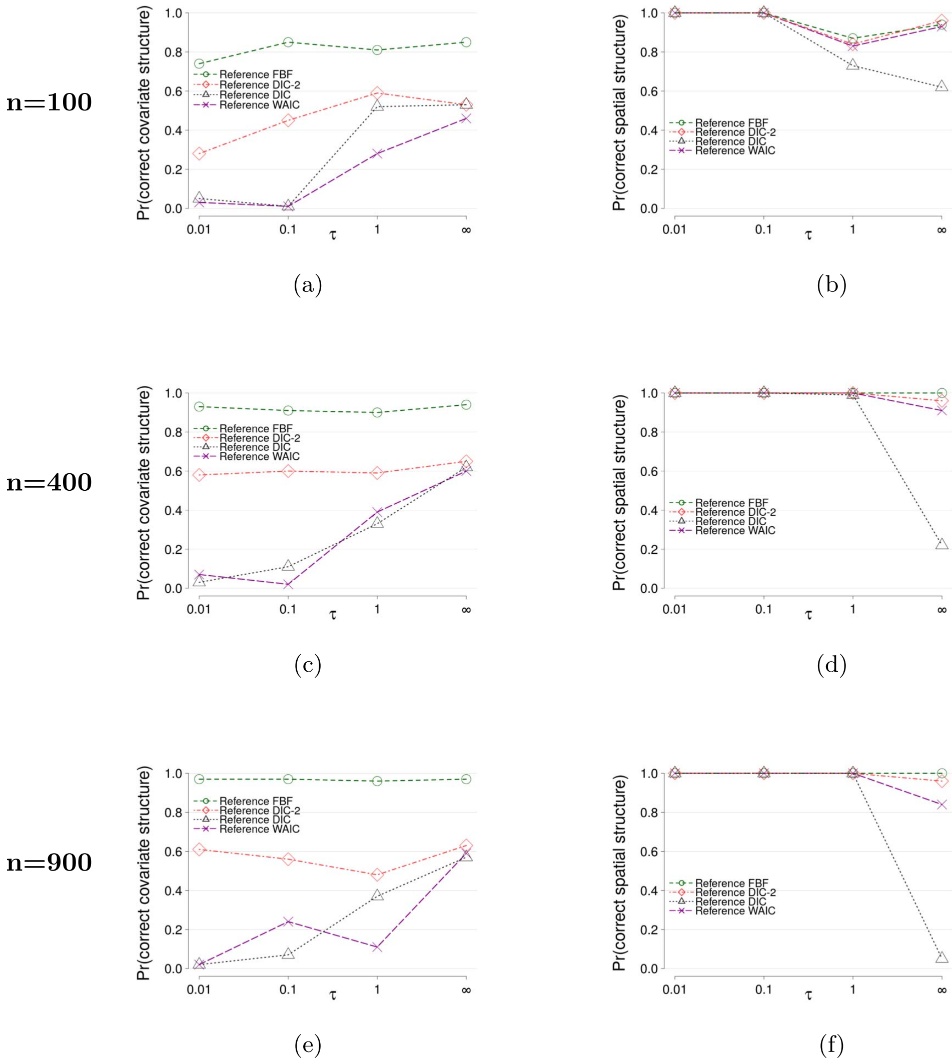


Figure 2: Proportion of times out of 100 simulated data sets that the reference FBF, reference DIC-2, reference DIC, and reference WAIC methods select the correct covariate and spatial dependence structure for $\tau \in \{0.01, 0.1, 1, \infty\}$ for n = 100 (top row), n = 400 (middle row), and n = 900 (bottom row). The reference FBF selection method reliably selects covariates and spatial dependence for all values of τ and performs better than DIC-2, DIC, and WAIC for selection in all data settings.

form selection for two existing datasets, whose responses include county-level median household income in the contiguous United States and residential crime rates in the neighborhoods of Columbus, Ohio.

To demonstrate the breadth of our method, we show that our FBF approach can also be used to select between different types of spatial random effects. In particular, for the two case studies that follow, we also considered selection with the class of simultaneous autoregressive (SAR) models in the model space $\mathcal{M}$. We adopt the model form and independence Jeffreys prior for the SAR model from De Oliveira and Song (2008). The SAR model for response $\mathbf{Y}$ is given by the following autoregression.

$$\mathbf{Y} = X\boldsymbol{\beta} + (I_n - B)^{-1}\boldsymbol{\epsilon}, \tag{25}$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$ and $B = \gamma W$, with unknown spatial parameter γ and $W = (w_{ij})_{n\times n}$ is a known, symmetric weight matrix with all $w_{ij} \geq 0$ and $w_{ij} > 0$ if $i \in N_j$. As with the ICAR model, we treat adjacent subregions as neighbors with the SAR model. The spatial parameter $\gamma \in (\lambda_n^{-1}, \lambda_1^{-1})$, where $\lambda_1 \geq \lambda_2 \ldots, \geq \lambda_n$ are the ordered eigenvalues of W, and $\gamma = 0$ corresponds to the OLM with distribution $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$.

Then SAR response $\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \gamma$ has the following Gaussian distribution.

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma^2, \gamma \sim N(X\boldsymbol{\beta}, (I_n - B)^{-1} M (I_n - B^T)^{-1}), \tag{26}$$

where $M = \sigma^2 I_n$. We consider the independence Jeffreys prior $\pi^J(\boldsymbol{\beta}, \sigma^2, \gamma)$ for SAR model parameters $\boldsymbol{\beta}$, σ^2, and γ (De Oliveira and Song, 2008).

$$\pi^J(\boldsymbol{\beta}, \sigma^2, \gamma) \propto \frac{1}{\sigma^2}\left\{ \sum_{i=1}^{n} \left(\frac{\lambda_i}{1-\gamma\lambda_i}\right)^2 - \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\lambda_i}{1-\gamma\lambda_i}\right]^2 \right\}^{\frac{1}{2}}. \tag{27}$$

We use the same training size, $m = p + 1$, for the SAR model with independence Jeffreys prior, as the prior induces similar behavior in the integrated likelihood to that produced by the reference prior for the ICAR model. Derivation of the fractional integrated likelihood $q_c(b, \mathbf{Y})$ for the SAR model with independence Jeffreys prior is detailed in the Fractional Integrated Likelihood Calculations section of the Supplementary Material.

Upon including SAR models in the candidate set, we adjust the model priors from Section 3. This results in a 50/25/25 split between prior probability for OLMs, ICAR models, and SAR models, with the remaining probability within each class attributed by model size, as described in Section 3. Thus, the prior probability for an OLM M_c with k_c covariates is

$$P(M_c) = \frac{1}{2(K+1)}\binom{K}{k_c}^{-1}, \tag{28}$$

and the prior probability for an ICAR or SAR model M_c with k_c covariates is

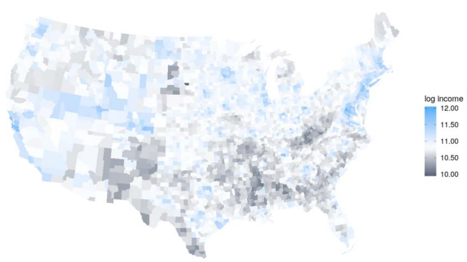
$$P(M_c) = \frac{1}{4(K+1)}\binom{K}{k_c}^{-1}. \tag{29}$$

The following case studies perform selection using the FBF with minimal training size $m = p + 1$ where OLM, ICAR, and SAR models are included in the model space.

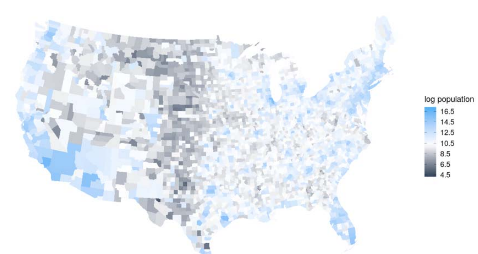
5.1 Case Study: US Socioeconomic Application

To demonstrate our formal Bayesian model selection approach for areal data, we first consider an application to median household income by county in the contiguous United States in 2017. We consider the logarithm of median household income as the response variable and we select among five candidate predictors: logarithm of the county population in 2017; logarithm of the unemployment rate in 2017; and three indicator variables for whether the county belongs to a large metropolitan area, a medium metropolitan area, or a small metropolitan area. The baseline covariate level corresponds to a non-metropolitan county. Figure 3 plots the response variable and all the candidate covariates over a map of the contiguous US counties.

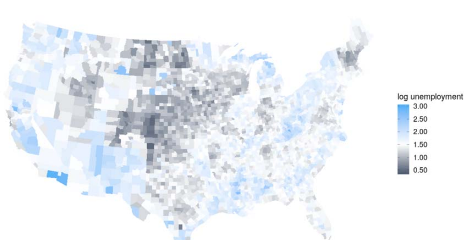
Following the approach to simultaneous selection of spatial model structure and fixed effects presented in Section 3 and adapted as described above to include SAR models in the candidate set, we form posterior model probabilities for all 96 models that include either an ICAR, SAR, or independent model structure as in (9), (26), and (10) and every combination of the five covariates. We use a training size of $m = 7$, according to the minimal training fraction found in Section 3.3. Model selection using the FBF approach



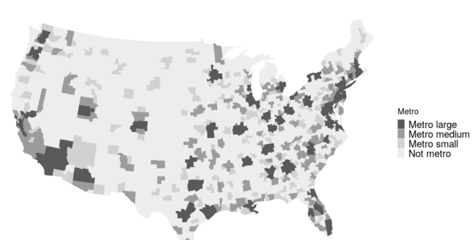
(a) Log median income by county



(b) Log population by county



(c) Log unemployment rate by county



(d) Metro area classification by county

Figure 3: Map of United States socioeconomic variables by county in 2017: (a) logarithm of median household income; (b) logarithm of population; (c) logarithm of unemployment rate; (d) metro area classification.

selects with probability 1 the ICAR model of form (9) with all five candidate predictors. To assess the impact of priors on the model space, we also performed selection among the 96 models using uniform priors on every model. Note that between the ICAR and SAR sets this causes 2/3 prior probability of selecting a spatial model.

The prior probability of the ICAR model with all covariates is equal to 0.0417 under (15) and decreases to 0.0104 under the use of equal prior probability for every model. But even with the latter prior specification, the posterior probability of the ICAR model with all five covariates is also equal to 1. Therefore, the model selection result in this case study does not appear to be highly sensitive to reasonable specifications of model prior probabilities.

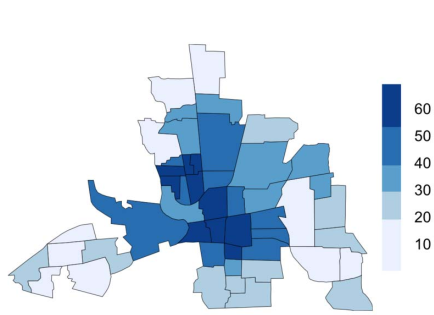
In keeping with the results seen in Figure 2, the type 2 DIC also selects the ICAR model with all covariates, with corresponding value equal to −3926.387. In contrast, both the DIC and WAIC select the OLM with all five candidate predictors, with criteria values equal to −2,306.965 and −2,306.213, respectively. The estimated τ value for this data set is 0.1575, which indicates strong spatial dependence. Finally, the results from the simulation study presented in Section 4 indicate that the reference FBF selection method is more reliable. These results demonstrate performance of our reference FBF selection method when applied to large spatial data sets.

5.2 Case Study: Columbus, OH Crime Rates

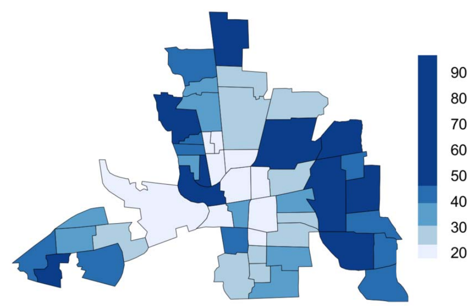
Next we consider a data set containing crime rates in the 49 neighborhoods of Columbus, OH in 1980. This data set has been previously analyzed by Anselin (1988) and Banerjee et al. (2015) and can be obtained from the `spData` package in R (Bivand et al., 2019). The response variable is residential burglaries and vehicle thefts per thousand households in each of the $n = 49$ neighborhoods of Columbus, OH. We consider five available candidate predictors: housing value, household income, open space in the neighborhood, percentage of housing units without plumbing, and distance to the Columbus business district. Using the minimal training size $m = 7$ to select between the 96 candidate models, our FBF approach selects with probability 0.1422 the OLM with three covariates: housing value, household income, and distance to the Columbus business district. Table 1 lists the candidate predictors and their corresponding posterior inclusion probabilities; the selected model contains the three covariates with the largest posterior inclusion probabilities. Figure 4 plots the response variable and the three selected covariates over a map of the 49 neighborhoods in Columbus, OH. In contrast to the previous case study, $n = 49$ is a small sample size and thus posterior probabilities do not move as far off the model priors. In particular, the prior probability for an OLM with three covariates was 0.0083. The total posterior model probability of selecting an OLM was 0.6770, indicating that the decision about spatial structure for this application has moved only slightly off the 1/2 prior probability of selecting an OLM. Table 2 lists the covariate structure, dependence structure, posterior probability, and DIC-2, DIC, and WAIC values for the top models indicated by the FBF approach. The top five models are OLMs and the model with sixth highest probability is the ICAR model containing the same covariates as the selected model. Covariates household income and distance to the Columbus business district have the two largest

Covariate description	**Posterior inclusion probability**
housing value	0.733
household income	0.931
open space in the neighborhood	0.302
percentage of housing units without plumbing	0.432
distance to Columbus business district	0.918

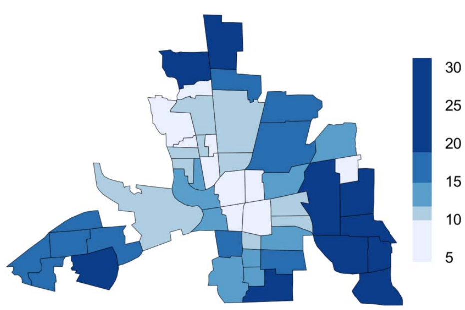
Table 1: Description and posterior inclusion probability for each of the 5 candidate covariates available for modeling theft and burglary rates in the neighborhoods of Columbus, OH.



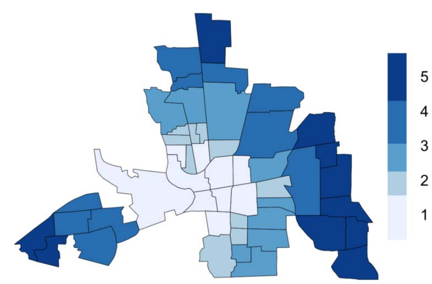
(a) Crimes per 1,000



(b) Housing value in $1,000



(c) Household income in $1,000



(d) Distance to business district

Figure 4: Map of Columbus, OH variables by neighborhood in 1980: (a) crimes per 1,000; (b) housing value; (c) household income; (d) distance to Columbus business district.

posterior inclusion probabilities and are included in all of the top 6 models. We also performed selection for this data set assigning uniform priors on the model space. This setup selected with posterior probability 0.1458 the OLM with the covariates housing value, household income, and distance to the Columbus business district. The prior

Covariate					**Model type**	**Post. prob.**	**DIC-2**	**DIC**	**WAIC**
value	income	open	plumb	dist					
✓	✓			✓	OLM	0.142	369.7	369.8	370.7
✓	✓		✓	✓	OLM	0.126	369.9	369.8	370.7
	✓			✓	OLM	0.123	371.3	371.3	372.6
✓	✓	✓	✓	✓	OLM	0.081	372	372	372.2
✓	✓	✓		✓	OLM	0.061	371.7	371.6	372.1
✓	✓			✓	ICAR	0.060	370.7	330.2	343.8

Table 2: Top 6 models for the Columbus, OH crime data according to the reference FBF approach. The first set of 5 columns indicates which of the covariates are in the model with the following abbreviations: value (housing value), income (household income), open (open space in the neighborhood), plumb (percentage of housing units without plumbing), and dist (distance to Columbus business district). Columns 6-10 provide the corresponding model type, posterior model probability, DIC-2, DIC, and WAIC values. The top 5 models are OLMs and the model with 6_{th} highest posterior model probability is the ICAR model with the same covariate structure as the model selected by the FBF approach.

probability for an OLM with three covariates increased to 0.0104, so the model selected by our initial FBF setup received even more prior mass from uniform priors. The total posterior model probability of selecting an OLM was 0.5089, which is an increase from the prior probability of 0.3333 of selecting an OLM. The posterior inclusion probabilities for the candidate predictors are 0.6827, 0.9033, 0.1816, 0.3002, and 0.8830 when performing selection with uniform model priors. Among OLM and ICAR models, the DIC selects the ICAR model with the covariate housing value and WAIC selects the ICAR model with covariates housing value and open space in the neighborhood. The DIC-2 selects the same OLM as the FBF approach. The DIC-2, DIC and WAIC values for their chosen models are 369.736, 265.814 and 288.481, respectively. This coincides with the simulation study in Section 4, which indicates that the DIC and WAIC criteria in particular tend to select spatial models over OLMs more often than the FBF approach does. The estimated value of τ for this data set is 1.9794, which does not indicate strong spatial dependence among the observations. Despite the low sample size, this application highlights the ability of our FBF method to select both independent and spatial data models in real spatial applications.

6 Discussion

We have presented a FBF approach that enables automatic, objective Bayesian model selection for hierarchical models with ICAR spatial random effects. We have derived integrated likelihood expressions and the resulting FBFs under the reference prior for areal data, which acts as an automatic prior. We found the minimal training size for the

FBF for the hierarchical model with an ICAR prior when the reference prior is assigned to all model parameters, and showed through simulation that our approach provides consistent simultaneous selection of fixed effects and spatial model structure. Notably, our FBF approach provides superior results, in terms of both detection of covariates and spatial dependence, to the widely used model selection criteria DIC and WAIC. When compared to the type 2 DIC, which is calculated using a likelihood with the spatial random effects integrated out, the performance from our FBF approach is superior and more reliable in simulations. However, the type 2 DIC selects the same model as the FBF approach in each of the two case studies presented in Section 5. We have demonstrated in Section 4 that the FBF approach implemented with the reference prior performs well for selection in spatial ICAR models, and Keefe et al. (2019) established that the reference prior to have favorable properties for estimation. Thus, the FBF approach provides the ability to use a single prior for both estimation and model selection for spatially dependent areal data. Finally, we showed that our FBF selection approach can be applied to spatial areal data sets of many sizes, and can be generalized to select between different types of spatial random effects (e.g. ICAR versus SAR). As is the case for other variable subset selection approaches, the model space grows exponentially with the number of candidate predictors and, as is well known, exhaustive search becomes computationally burdensome for large p. In this work we examine problems where the entire model space can be enumerated and assigned posterior model probabilities. When the model space is too large for exhaustive search, our FBF-based model selection approach can still be used in conjunction with a stochastic search algorithm to explore the model space such as the genetic algorithm used by Wu et al. (2020).

There are many possible avenues for future research. First, we note that our reference FBF approach provides posterior model probabilities, which could be used in future research to provide Bayesian model averaging for prediction. Other future work may include developing model selection for data in the exponential family. In particular, ongoing work addresses ICAR effects for Poisson and Binomial response data, which commonly occurs in disease-mapping and health data. In principle, one could extend the proposed methodology to the Poisson case by developing automatic and/or objective priors for ICAR random effects in the Poisson context, deriving the minimal training size, and applying FBF methodology to produce Bayesian model selection for count data models.

Supplementary Material

Supplementary Material for "Objective Bayesian Model Selection for Spatial Hierarchical Models with Intrinsic Conditional Autoregressive Priors" (DOI: 10.1214/23-BA1375SUPP; .pdf). Supplementary material for "Objective Bayesian Model Selection for Spatial Hierarchical Models with Intrinsic Conditional Autoregressive Priors" provides proofs for theoretical results and additional simulation results.

References

Anselin, L. (1988). *Spatial econometrics: methods and models*. Springer Science & Business Media. 1004

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, F: CRC Press, 2nd edition. MR3362184. 1004

Berger, J. O. and Pericchi, L. R. (1996). "The Intrinsic Bayes Factor for Model Selection and Prediction." *Journal of the American Statistical Association*, 91(433): 109–122. MR1394065. doi: https://doi.org/10.2307/2291387. 993

Besag, J. (1974). "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2): 192–225. MR0373208. doi: https://doi.org/10.1111/j.2517-6161.1974.tb00999.x. 985

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. 985, 986, 989, 990

Best, N., Arnold, R., Thomas, A., Waller, L., and M. Conlon, E. (1999). "Bayesian Models for Spatially Correlated Disease and Exposure Data." *Bayesian Statistics*, 6: 65–82. MR1723496. 986

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. PMID: 15690999. MR2135924. doi: https://doi.org/10.1191/0962280205sm388oa. 986

Bivand, R., Nowosad, J., and Lovelace, R. (2019). *spData: Datasets for Spatial Analysis*. R package version 0.3.2. URL https://CRAN.R-project.org/package=spData. 1004

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). "Deviance information criteria for missing data models." *Bayesian Analysis*, 1(4): 651–673. MR2282197. doi: https://doi.org/10.1214/06-BA122. 998

Chipman, H., George, E. I., and McCulloch, R. E. (2001). *The Practical Implementation of Bayesian Model Selection*, 67–116. Model Selection. Beachwood, OH: Institute of Mathematical Statistics. MR2000752. doi: https://doi.org/10.1214/lnms/1215540964. 993

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: 167–183. MR2822154. doi: https://doi.org/10.1007/s00184-009-0295-7. 989

De Oliveira, V. and Song, J. J. (2008). "Bayesian Analysis of Simultaneous Autoregressive Models." *Sankhya: The Indian Journal of Statistics, Series B*, 70(2): 323–350. URL http://www.jstor.org/stable/41234438. MR2563993. 997, 1002

Ferreira, M. A. R. (2019). "The limiting distribution of the Gibbs sampler for the intrinsic conditional autoregressive model." *Brazilian Journal of Probability and Statistics*, 33(4): 734–744. MR3996314. doi: https://doi.org/10.1214/19-bjps435. 989

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. 989

Ferreira, M. A. R., Porter, E. M., and Franck, C. T. (2021). "Fast and scalable computations for Gaussian hierarchical models with intrinsic conditional autoregressive spatial random effects." *Computational Statistics and Data Analysis*, 162: 107264. MR4256170. doi: https://doi.org/10.1016/j.csda.2021.107264. 989, 990, 998

Franck, C. T. and Gramacy, R. B. (2020). "Assessing Bayes Factor Surfaces Using Interactive Visualization and Computer Surrogate Modeling." *The American Statistician*, 74(4): 359–369. MR4168256. doi: https://doi.org/10.1080/00031305.2019.1671219. 993

Gelman, A., Hwang, J., and Vehtari, A. (2014). "Understanding predictive information criteria for Bayesian models." *Statistics and computing*, 24(6): 997–1016. MR3253850. doi: https://doi.org/10.1007/s11222-013-9416-2. 998

Jin, X., Banerjee, S., and Carlin, B. P. (2007). "Order-free co-regionalized areal data models with application to multiple-disease mapping." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 817–838. MR2368572. doi: https://doi.org/10.1111/j.1467-9868.2007.00612.x. 987

Jin, X., Carlin, B. P., and Banerjee, S. (2005). "Generalized hierarchical multivariate CAR models for areal data." *Biometrics*, 61(4): 950–61. MR2216188. doi: https://doi.org/10.1111/j.1541-0420.2005.00359.x. 986, 987

Kass, R. E. and Raftery, A. E. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90(430): 773–795. MR3363402. doi: https://doi.org/10.1080/01621459.1995.10476572. 993

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. MR3795058. doi: https://doi.org/10.1016/j.spasta.2018.03.007. 986, 987, 988, 989, 990, 991

Keefe, M. J., Ferreira, M. A. R., and Franck, C. T. (2019). "Objective Bayesian Analysis for Gaussian Hierarchical Models with Intrinsic Conditional Autoregressive Priors." *Bayesian Analysis*, 14(1): 181–209. MR3910043. doi: https://doi.org/10.1214/18-BA1107. 988, 989, 990, 991, 996, 998, 1007

Lee, D. (2011). "A comparison of conditional autoregressive models used in Bayesian disease mapping." *Spatial and Spatio-temporal Epidemiology*, 2(2): 79–89. URL http://www.sciencedirect.com/science/article/pii/S1877584511000049. 986

Lee, D. and Mitchell, R. (2013). "Locally adaptive spatial smoothing using conditional auto-regressive models." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(4): 593–608. MR3083913. doi: https://doi.org/10.1111/rssc.12009. 986, 987

Leroux, B., Lei, X., and Breslow, N. (1999). *Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence*, 135–178. Statistical Models in Epidemiology, the Environment and Clinical Trials. New York: Springer. MR1731684. doi: https://doi.org/10.1007/978-1-4612-1284-3_4. 986

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(2): 599–625. MR3472004. doi: https://doi.org/10.1214/15-BA972. 986

Logan, J. R., Bauer, C., Ke, J., Xu, H., and Li, F. (2020). "Models for Small Area Estimation for Census Tracts." *Geographical Analysis*, 52(3): 325–350. doi: https://doi.org/10.1111/gean.12215. 986

Martinez-Beneito, M. A., Botella-Rocamora, P., and Banerjee, S. (2017). "Towards a Multidimensional Approach to Bayesian Disease Mapping." *Bayesian Analysis*, 12(1): 239–259. MR3597574. doi: https://doi.org/10.1214/16-BA995. 987

O'Hagan, A. (1995). "Fractional Bayes Factors for Model Comparison." *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1): 99–138. URL http://www.jstor.org/stable/2346088. MR1325379. 987, 993, 994

Penrose, R. (1955). "A generalized inverse for matrices." *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3): 406–413. MR0069793. 989

Porter, E. M., Franck, C. T., Ferreira, M. A. R. (2023). "Supplementary Material for "Objective Bayesian Model Selection for Spatial Hierarchical Models with Intrinsic Conditional Autoregressive Priors"." *Bayesian Analysis*. doi: https://doi.org/10.1214/23-BA1375SUPP. 988

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. 986

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2): 319–392. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 990

Scott, J. G. and Berger, J. O. (2010). "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem." *Annals of Statistics*, 38(5): 2587–2619. MR2722450. doi: https://doi.org/10.1214/10-AOS792. 992

Song, J. J. and De Oliveira, V. (2012). "Bayesian model selection in spatial lattice models." *Statistical Methodology*, 9(1-2): 228–238. MR2863610. doi: https://doi.org/10.1016/j.stamet.2011.01.003. 986, 987

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639. MR1979380. doi: https://doi.org/10.1111/1467-9868.00353. 987, 998

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., and Fortin, M.-J. (2018). "Spatial autoregressive models for statistical inference from ecological data." *Ecological Monographs*, 88(1): 36–59. doi: https://doi.org/10.1002/ecm.1283. 986

Watanabe, S. (2010). "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory." *Journal of Machine Learning Research*, 11: 3571–3594. MR2756194. 998

White, P., Gelfand, A., and Utlaut, T. (2017). "Prediction and model comparison for areal unit data." *Spatial Statistics*, 22: 89–106. MR3726177. doi: https://doi.org/10.1016/j.spasta.2017.09.002. 986, 987

Wu, H.-H., Ferreira, M. A., Elkhouly, M., and Ji, T. (2020). "Hyper nonlocal priors for variable selection in generalized linear models." *Sankhya A*, 82(1): 147–185. MR4155019. doi: https://doi.org/10.1007/s13171-018-0151-9. 1007

Acknowledgments

We are grateful for the comments and suggestions of the Associate Editor and two anonymous reviewers that have led to a greatly improved manuscript.