THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

WP # 0025MSS-496-2010

Date April 18, 2011

BAYESIAN MODEL SELECTION IN SPATIAL LATTICE MODELS

Joon Jin Song
Center for Statistical Research and Consulting
Department of Mathematical Sciences
University of Arkansas
Fayetteville, AR 72701, USA
jjsong@uark.edu

Victor De Oliveira1
Department of Management Science and Statistics
The University of Texas at San Antonio
San Antonio, TX 78249, USA
victor.deoliveira@utsa.edu

BAYESIAN MODEL SELECTION IN SPATIAL LATTICE MODELS

Joon Jin Song
Center for Statistical Research and Consulting
Department of Mathematical Sciences
University of Arkansas
Fayetteville, AR 72701, USA
jjsong@uark.edu

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

April 13, 2011

Abstract

This work describes a Bayesian approach for model selection in Gaussian conditional autoregressive models and Gaussian simultaneous autoregressive models which are commonly used to describe spatial lattice data. The approach is aimed at situations when all competing models have the same mean structure, but differ on some aspects of their covariance structures. The proposed approach uses as selection criterion the posterior model probabilities computed using some default priors for the model parameters. The proposed methodology is illustrated using two real datasets.

Key words: Bayes factors; CAR models; Jeffreys prior; Neighborhood system; SAR models; Spatial data; Weight matrix.

JEL Classifications: C11, C31

[1]Corresponding author.

1 Introduction

Spatial lattice data (or areal data as they are also known) consist of observations collected at sub-regions (or sites as they are also known) that form a partition of a region of interest. These data often represent aggregates or summaries of a quantity of interest over the sub-regions, e.g., number of cancer cases or crime rates in the counties of a state that occurred over a period of time. Two of the most commonly used classes of models to describe these data are conditionally autoregressive (CAR) models and simultaneously autoregressive (SAR) models; see Anselin (1988), Cressie (1993) or Banerjee, Carlin and Gelfand (2004) for ample descriptions of these models.

A review of the spatial statistics literature reveals that CAR models are the most often used for the analysis of lattice/areal data, a practice that may be due in part to the influential works of Besag (1974) and Cressie (1993) who recommended CAR over SAR models (and related variants). On the other hand, a review of the spatial econometrics and geography literature reveals the opposite situation, namely, that SAR models (and related variants) are the most often used for the analysis of lattice/areal data. This in turn seems to be due in part to the influential works of Ord (1975) and Anselin (1988) who advocated the use of SAR models (and related variants).

In addition, the formulation of both CAR and SAR models requires selecting a neighborhood system that specifies direct relations between sub-regions. A common practice is to specify this using geographic adjacency, which is often the default choice. However, for some datasets the direct relation between sub-regions may be better quantified using distance or other criterion, such as neighborhood systems defined in terms of the values of an explanatory variable. For instance, Case, Rosen and Hines (1993) found that a SAR model based on similarity of racial composition in states, as the criterion to define neighbors, fits U.S. states expenditure data much better than a SAR model that uses neighbors based on geographic adjacency.

All the above choices are important for an adequate fit and interpretation of these models. Nevertheless, it appears that in general there is little or no subject-based knowledge to choose the neighborhood system or between CAR and SAR models, and quite often these choices are made based on tradition or taste. Scientific guidelines to aid choosing neighborhood systems, and between CAR and SAR models are then relevant in applications. Numerous model selection methods have been proposed in the literature (see for instance Claeskens and Hjort, 2008), but it seems that model selection methodology for spatial models has not been explored much, and is somewhat 'underdeveloped' (Zhu, Huang and Reyes, 2010). Previous works on model selection methods for

lattice data include Hepple (1995a,b) on standard Bayesian methods, Florax, Folmer and Rey (2003) on test of hypotheses methods, and Zhu et al. (2010) on penalized likelihood methods.

The goal of this work is to study model selection in CAR and SAR models using standard Bayesian methods, such as those described in Kass and Rafferty (1995) and Berger and Pericchi (2001). Specifically, we would use posterior model probabilities as the criterion for model selection, and compute them based on (modifications of) default priors that have recently been derived for CAR and SAR models. The methods studied here are aimed at situations when all competing models have the same mean structure, and the model differences rely on some aspects of the covariance structure. Similar problems and methods have been considered by Osiewalski and Steel (1993) for the analysis of regression models with elliptically symmetric distributions, and by Hepple (1995a,b) for the analysis of some spatial models, although the Bayes factors proposed in the latter were not not well defined and calibrated. Conditions on the models being compared and the improper priors on the model parameters that guarantee well defined and calibrated Bayes factors were obtained by Berger, Pericchi and Varshavsky (1998). Their model conditions apply to the context considered here where all competing models have the same mean structure, and we modify some previously proposed default priors to fully fit their conditions. We also describe a simple Monte Carlo method to compute the marginal densities of the competing models based on importance sampling, which works well for both small and large datasets. Finally, the proposed model selection approach is illustrated on two real datasets, one on a regular lattice and one on an irregular lattice.

2 Spatial Models for Lattice Data

For each site indexed by $i = 1, \ldots, n$, a variable of interest Y_i is observed, usually an aggregate or summary, together with a set of p ($< n$) explanatory variables $x_{i1}, \ldots, x_{ip}$. We consider in this work several classes of CAR models and SAR models that have been extensively used in economics, epidemiology and geography.

2.1 Neighborhood Systems

The collection of sites $\{1, \ldots, n\}$ is assumed to be endowed with a neighborhood system, $\{N_i : i = 1, \ldots, n\}$, where N_i denotes the sites that are neighbors of site i. This neighborhood system is key in determining the covariance structure of the variable of interest, with N_i interpreted as

the sites that have 'direct dependence' with site i. This specification of the dependence structure is natural when modeling lattice data, since similarity between sites often depends on some of the sites' shared features, such as boundaries, proximity or similarity of explanatory variables.

Neighborhood systems can be very general as they only need to satisfy that for any $i, j = 1, \ldots, n$, $j \in N_i$ if and only if $i \in N_j$ and $i \notin N_i$. But they need to be judiciously selected to reflect the presumed direct relations among the responses. An emblematic example commonly used in applications is the neighborhood system defined in terms of geographic adjacency

$$N_i = \{j : \text{site } j \text{ shares a boundary with site } i\}, \qquad i = 1, \ldots, n. \tag{1}$$

But many other options are possible. For regular lattices the most common choices are the first order neighborhood system, where the neighbors of site i are the sites adjacent to the north, south, east and west, and the second order neighborhood system, where the neighbors of site i are its first order neighbors plus their first order neighbors (except for site i). Higher order neighborhood systems are also possible, but less often used. For irregular lattices the most common choices are neighborhood systems based on shared geographic features, such as (1), and those based on distance such as

$$N_i = \{j : 0 < d_{ij} < r\}, \qquad \text{with } r > 0, \qquad i = 1, \ldots, n, \tag{2}$$

where d_{ij} is the distance between sites i and j. Different choices of r result in different neighborhood systems.

2.2 Conditional Autoregressive Models

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ denote the response data. Given a neighborhood system, Gaussian conditional autoregressive (CAR) models are specified by the set of full conditional distributions having the autoregressive structure

$$(Y_i \mid \mathbf{Y}_{(i)}) \sim \mathrm{N}\Big(\mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} c_{ij}(Y_j - \mathbf{x}_j'\boldsymbol{\beta}), \tau_i^2\Big), \qquad i = 1, \ldots, n, \tag{3}$$

where $\mathbf{Y}_{(i)} = \{Y_j, j \neq i\}$, $\mathbf{x}_j' = (x_{j1}, \ldots, x_{jp})$ are the values of the explanatory variables in site j, $\boldsymbol{\beta} \in \mathbb{R}^p$ are regression parameters, and $\tau_i > 0$ and $c_{ij} \geq 0$ are covariance parameters, with $c_{ij} > 0$ if and only if i and j are neighbors (so $c_{ii} = 0$ for all i). For the set of full conditional distributions (3) to determine a well defined joint distribution for $\mathbf{Y}$, the matrices $M = \mathrm{diag}(\tau_1^2, \ldots, \tau_n^2)$ and $C = (c_{ij})$ must satisfy the conditions:

(*a*) $M^{-1}C$ is symmetric, so $c_{ij}\tau_j^2 = c_{ji}\tau_i^2$ for all $i, j = 1, \ldots, n$;

(*b*) $M^{-1}(I_n - C)$ is positive definite.

When (*a*) and (*b*) hold the joint distribution of $\mathbf{Y}$ is given by

$$\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, (I_n - C)^{-1}M), \tag{4}$$

where $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, assumed to have full rank, and I_n is the $n \times n$ identity matrix; see Cressie (1993) or Banerjee et al. (2004). We consider in this work three CAR models in which matrices M and C, in addition to satisfying (*a*) and (*b*), have also the following structure:

(*c*) $M = \sigma^2 G$, with $\sigma^2 > 0$ unknown and G diagonal with known positive diagonal elements;

(*d*) $C = \phi W$, with ϕ an unknown 'spatial parameter' and $W = (w_{ij})$ a known "weight" matrix (not necessarily symmetric) that is nonnegative ($w_{ij} \geq 0$) and satisfies $w_{ij} > 0$ if and only if sites i and j are neighbors (so $w_{ii} = 0$).

In what follows $A = (a_{ij})$ denotes the $n \times n$ symmetric matrix defined by $a_{ij} = 1$ if i and j are neighbors, and $a_{ij} = 0$ otherwise. Several classes of CAR models have been proposed in the literature within the aforementioned structure, that amounts to specify matrices G and W. The most common ones, reviewed by Cressie and Kapat (2008), are described below:

(*i*) The Homogeneous CAR (HCAR) model:
$G = I_n$ and $W = A$.

(*ii*) The Weighted CAR (WCAR) model (Besag, York and Mollie, 1991):
$G = \mathrm{diag}(|N_1|^{-1}, \ldots, |N_n|^{-1})$ and $W = GA$, where $|N_i| = \sum_{j=1}^n a_{ij}$ is the number of neighbors of site i.

(*iii*) The Autocorrelation CAR (ACAR) model (Cressie and Chang, 1989):
$G = \mathrm{diag}(|N_1|^{-1}, \ldots, |N_n|^{-1})$ and $W = G^{1/2}AG^{-1/2}$.

For any of the classes of CAR models (*i*)–(*iii*) it can be directly checked that condition (a) holds. For condition (b) to hold the spatial parameter ϕ is required to belong to $(\lambda_n^{-1}, \lambda_1^{-1})$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the ordered eigenvalues of $G^{-1/2}WG^{1/2}$, with $\lambda_n < 0 < \lambda_1$ since $\mathrm{tr}(G^{-1/2}WG^{1/2}) = \mathrm{tr}(W) = 0$. Note that for the HCAR and ACAR models $G^{-1/2}WG^{1/2} = A$, while for the WCAR model $G^{-1/2}WG^{1/2} = G^{1/2}AG^{1/2}$, so for all the three models (*i*)–(*iii*) $G^{-1/2}WG^{1/2}$ is symmetric

and the λ_is are real. The parameter space of $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \phi)$ in any of the above classes of CAR models is then $\Omega = \mathbb{R}^p \times (0, \infty) \times (\lambda_n^{-1}, \lambda_1^{-1})$. The Appendix provides detailed proofs of the above claims for a larger class of CAR models.

2.3 Simultaneous Autoregressive Models

Given a neighborhood system[2], Gaussian simultaneous autoregressive (SAR) models are specified by a set of autoregressive equations on the variables themselves, rather than on their full conditional distributions, given by

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \sum_{j=1}^{n} b_{ij}(Y_j - \mathbf{x}_j'\boldsymbol{\beta}) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{5}$$

where $\mathbf{x}_j'$ and $\boldsymbol{\beta}$ are the same as in CAR models, $\epsilon_i \sim \mathrm{N}(0, \xi_i^2)$ are independent, and $\xi_i^2 > 0$ and $b_{ij} \geq 0$ are covariance parameters, with $b_{ij} > 0$ if and only if i and j are neighbors; let $B = (b_{ij})$ and $M = \mathrm{diag}(\xi_1^2, \ldots, \xi_n^2)$. Provided $I_n - B$ is nonsingular, the n scalar equations in (5) can be written as

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

where X is the same as in CAR models and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$, so

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

We consider SAR models where matrices M and B have a similar structure as, respectively, matrices M and C stated in (c) and (d), but with $G = I_n$ and $W = A$, with A as in the previous section. To guarantee that $I_n - \phi A$ is nonsingular ϕ is required to belong to $\mathbb{R} - \{\lambda_i^{-1}\}_{i=1}^n$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the ordered eigenvalues of A. But the feasible values for ϕ are almost always in practice taken to be the interval $(\lambda_n^{-1}, \lambda_1^{-1})$, so we also do that in this work. Then we consider SAR models of the form

$$\mathbf{Y} \sim \mathrm{N}_n(X\boldsymbol{\beta}, \sigma^2((I_n - \phi A)^2)^{-1}). \tag{6}$$

The parameter space for $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \phi)$ in SAR models is $\Omega = \mathbb{R}^p \times (0, \infty) \times (\lambda_n^{-1}, \lambda_1^{-1})$, which is the same as that of HCAR and ACAR models, but differs from that of WCAR models.

3 Model Selection

Let $M_1, M_2, \ldots, M_k$ be a set of $k \geq 2$ candidate models for the data $\mathbf{Y}$. In principle the differences between the models can be of any nature, but we consider here the case when all candidate models

[2]SAR models admit asymmetric neighboring relations, but these will not be considered here.

are either CAR or SAR models having the same mean structure, but with different covariance structures. For instance, for $i \neq j$ models M_i and M_j may be both CAR (SAR) models having different neighborhood systems, or one of them may be a CAR model while the other is a SAR model, having either the same or different neighborhood systems. The model selection criterion we would use is the posterior model probabilities based on some (modification of) default priors that have been recently derived for CAR and SAR models.

For $j = 1, \ldots, k$, let M_j be either an HCAR, WCAR, ACAR or SAR model, as given in (4) or (6), parameterized by $\boldsymbol{\eta}_j = (\boldsymbol{\beta}, \sigma_j^2, \phi_j) \in \Omega_j$ and having covariance structure depending on matrices G_j and A_j. Recall that for each model $\phi_j \in (1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$, where $\lambda_1^{(j)} \geq \lambda_2^{(j)} \geq \ldots \geq \lambda_n^{(j)}$ are the ordered eigenvalues of A_j, in the case of HCAR, ACAR and SAR models, or of $G_j^{1/2} A_j G_j^{1/2}$ in the case of WCAR models. Then all competing models have similar likelihoods given by

$$L_j(\boldsymbol{\eta}_j; \mathbf{y}) = (2\pi\sigma_j^2)^{-n/2} |\Sigma_{\phi_j}^{-1}|^{1/2} \exp\Big\{ -\frac{1}{2\sigma_j^2} (\mathbf{y} - X\boldsymbol{\beta})' \Sigma_{\phi_j}^{-1} (\mathbf{y} - X\boldsymbol{\beta}) \Big\}, \tag{7}$$

where

$$\Sigma_{\phi_j}^{-1} = \begin{cases} I_n - \phi_j A_j & \text{for HCAR models} \\ G_j^{-1} - \phi_j A_j & \text{for WCAR models} \\ G_j^{-1} - \phi_j G_j^{-1/2} A_j G_j^{-1/2} & \text{for ACAR models} \\ (I_n - \phi_j A_j)^2 & \text{for SAR models.} \end{cases}$$

3.1 Priors

Prior distributions need to be assigned to the parameters of each model M_j, and this would be done using default priors of the form

$$\pi(\boldsymbol{\eta}_j \mid M_j) \propto \frac{\pi(\phi_j \mid M_j)}{\sigma_j^2} \mathbf{1}_{\Omega_j}(\boldsymbol{\eta}_j), \tag{8}$$

where $\pi(\phi_j \mid M_j)$ needs to be specified and $\mathbf{1}_E(\cdot)$ denotes the indicator function of set E. Previous works on Bayesian analysis of CAR and SAR models (e.g. Bell and Broemeling, 2000; Hepple, 1995a,b) have used prior (8) with

$$\pi^U(\phi_j \mid M_j) = \mathbf{1}_{(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})}(\phi_j);$$

we call this the *uniform prior*. In spite of its non-informative appearance, this prior specification might be (arguably) inappropriate in some cases. For many datasets found in practice there is strong correlation between neighboring observations, and such behavior is reproduced in CAR

models only when the spatial parameter ϕ is quite close to one of the boundaries, λ_1^{-1} or λ_n^{-1} (Besag and Kooperberg, 1995). Prior $\pi^U(\phi_j \mid M_j)$ ignores this common behavior as it assigns equal mass to all parameter values. The same issue may also be raised when $\pi^U(\phi_j \mid M_j)$ is used as a default prior in SAR models.

Recently, De Oliveira (2011) and De Oliveira and Song (2008) derived the independence Jeffreys prior for, respectively, CAR and SAR models and studied their main properties. It turns out that this default prior is the same for both CAR and SAR models, and is of the form (8) with

$$\pi^{J1}(\phi_j \mid M_j) = \left\{ \sum_{i=1}^{n} \Big(\frac{\lambda_i^{(j)}}{1-\phi_j \lambda_i^{(j)}} \Big)^2 - \frac{1}{n} \Big[\sum_{i=1}^{n} \frac{\lambda_i^{(j)}}{1-\phi_j \lambda_i^{(j)}} \Big]^2 \right\}^{\frac{1}{2}} \mathbf{1}_{(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})}(\phi_j). \tag{9}$$

This prior is unbounded at $1/\lambda_1^{(j)}$ and $1/\lambda_n^{(j)}$, so it assigns large prior mass to parameter values close to these boundaries that represent strong correlation between neighboring observations. It then represents the common behavior mentioned above. Although $\pi^{J1}(\phi_j \mid M_j)$ is not integrable, it yields a proper posterior in both CAR and SAR models, as long as the eigenvectors corresponding to the extreme eigenvalues $\lambda_1^{(j)}$ and $\lambda_n^{(j)}$ do not belong to the column space of X, a condition likely met in practice; see De Oliveira (2011) and De Oliveira and Song (2008) for details.

3.2 Bayes Factors and Posterior Model Probabilities

For standard Bayesian model selection we also need to assign prior probabilities to all competing models, say $\pi(M_j)$; a sensible default choice is $\pi(M_j) = 1/k$ for all $j = 1, \ldots, k$. These prior model probabilities are updated by the data using Bayes theorem, where for any $i \neq j$ the posterior odds of model M_i against model M_j are given by

$$\begin{aligned} \frac{\pi(M_i \mid \mathbf{y})}{\pi(M_j \mid \mathbf{y})} &= \frac{m(\mathbf{y} \mid M_i)\pi(M_i)}{m(\mathbf{y} \mid M_j)\pi(M_j)} \\ &= B_{ij} \times \text{prior odds}_{ij}, \end{aligned} \tag{10}$$

where $m(\mathbf{y} \mid M_j)$ is the marginal (or prior predictive) density of $\mathbf{Y}$ under model M_j, given by

$$m(\mathbf{y} \mid M_j) = \int_{\Omega_j} L_j(\boldsymbol{\eta}_j \mid \mathbf{y}) \pi(\boldsymbol{\eta}_j \mid M_j) d\boldsymbol{\eta}_j,$$

and

$$B_{ij} = \frac{m(\mathbf{y} \mid M_i)}{m(\mathbf{y} \mid M_j)},$$

is the Bayes factor of M_i against M_j, which represents the evidence provided by the data in favor of M_i, as opposed to M_j. Then the posterior probability of model M_j is

$$\begin{aligned}\pi(M_j \mid \mathbf{y}) &= \left(\sum_{l=1}^{k} \frac{\pi(M_l)}{\pi(M_j)} B_{lj}\right)^{-1}, \qquad j = 1, \ldots, k, \\ &= \frac{m(\mathbf{y} \mid M_j)}{\sum_{l=1}^{k} m(\mathbf{y} \mid M_l)}, \quad \text{when all } \pi(M_j) \text{ are equal.} \end{aligned} \tag{11}$$

The model to be selected is the one for which $\pi(M_j \mid \mathbf{y})$ (or $m(\mathbf{y} \mid M_j)$ in case all $\pi(M_j)$ are equal) is the largest.

In general, standard Bayesian model selection can not be done with improper priors since these are specified only up to an arbitrary multiplicative constant, which make the resulting Bayes factors and posterior model probabilities undetermined. But an important exception occurs when all the competing models have the same invariance structure, up to individual model parameters that have proper priors (Berger et al., 1998). The CAR and SAR models we consider here fit this important exception when all the competing models have the same mean structure and prior (8) is used with $\pi(\phi_j \mid M_j)$ proper. The latter is certainly fulfilled by $\pi^U(\phi_j \mid M_j)$, but not by $\pi^{J1}(\phi_j \mid M_j)$ since

$$\pi^{J1}(\phi_j \mid M_j) = O((1 - \phi_j \lambda_i^{(j)})^{-1}) \quad \text{as} \;\; \phi_j \to 1/\lambda_i^{(j)}; \;\; i = 1 \text{ or } n;$$

see De Oliveira and Song (2008). Alternatively we use instead $(\pi^{J1}(\phi_j \mid M_j))^r$, with some $r < 1$, which is proper and has the same "shape" and desirable behavior of assigning large prior mass to parameter values close to the parameter space boundaries.

For the computation of marginal densities in CAR and SAR models with prior (8), standard calculations show that integration with respect to $\boldsymbol{\beta}$ and σ_j^2 can be done analytically, resulting in

$$m(\mathbf{y} \mid M_j) = K c_j \int_{1/\lambda_n^{(j)}}^{1/\lambda_1^{(j)}} h(\phi_j, M_j, \mathbf{y}) d\phi_j, \qquad j = 1, \ldots, k, \tag{12}$$

where

$$h(\phi_j, M_j, \mathbf{y}) = |\Sigma_{\phi_j}^{-1}|^{1/2} |X' \Sigma_{\phi_j}^{-1} X|^{-1/2} (S_{\phi_j}^2)^{-(n-p)/2} \pi(\phi_j \mid M_j),$$

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

and

$$K = \frac{\Gamma(\frac{n-p}{2})}{\pi^{\frac{n-p}{2}}} \quad , \quad c_j = \left(\int_{1/\lambda_n^{(j)}}^{1/\lambda_1^{(j)}} \pi(\phi_j \mid M_j) d\phi_j \right)^{-1}.$$

It is important to note that for the posterior model probabilities (11) to be well defined and calibrated, the proportionality constants in the likelihood and prior of all competing models should be retained (unless a constant is the same across all competing models, like K above). This was not done in Hepple (1995a,b), so some of the posterior model probabilities reported in that article are mistaken. In this work the default choices to be used for $\pi(\phi_j \mid M_j)$ are $\pi^U(\phi_j \mid M_j)$ and $(\pi^{J1}(\phi_j \mid M_j))^{1/2}$. Hence computation of $m(\mathbf{y} \mid M_j)$ involves one-dimensional integration over a bounded interval, which we consider next.

3.3 Computation

The computation of the proportionality constants c_j is straightforward. For $\pi(\phi_j \mid M_j) = \pi^U(\phi_j \mid M_j)$, $c_j = (1/\lambda_1^{(j)} - 1/\lambda_n^{(j)})^{-1}$, while for $\pi(\phi_j \mid M_j) = (\pi^{J1}(\phi_j \mid M_j))^{1/2}$, c_j can be computed either by numerical quadrature, such as the adaptive quadrature algorithm implemented in the `R` function `integrate`, or by the simple Monte Carlo estimate

$$\hat{c}_j = \left(\left(\frac{1}{\lambda_1^{(j)}} - \frac{1}{\lambda_n^{(j)}} \right) \frac{1}{m} \sum_{l=1}^{m} (\pi^{J1}(\phi_j^{(l)} \mid M_j))^{1/2} \right)^{-1}, \tag{13}$$

with $\phi_j^{(1)}, \ldots, \phi_j^{(m)} \overset{\text{iid}}{\sim} \text{unif}(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$ and m large. The computation of marginal densities requires more care. In principle $m(\mathbf{y} \mid M_j)$ could also be computed by numerical quadrature, but in our experience this is likely to fail in datasets with moderate or large sample sizes. The reason is that for such datasets the integrand in (12) is highly peaked (and often concentrated near the right boundary $1/\lambda_1^{(j)}$), so it is almost constant and very close to zero over most of the integration region. Because of that estimates obtained by numerical quadrature for such datasets are often zero or nearly so, and Monte Carlo estimates similar to (13) would also be highly inefficient because of the same reason. We propose Monte Carlo estimation with an importance sampling density tailored to this particular situation.

Let $\tilde{\phi}_j$ be the value that maximizes the integrand in (12), to be computed numerically, and $t \in [3, 4]$. We propose estimating $m(\mathbf{y} \mid M_j)$ using as importance sampling density the normal density with mean $\tilde{\phi}_j$ and standard deviation $\omega_j = (1/\lambda_1^{(j)} - \tilde{\phi}_j)/t$, truncated to the interval $(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$. The resulting estimate is

$$\hat{m}(\mathbf{y} \mid M_j) = \left(\Phi(t) - \Phi\left(t \frac{1/\lambda_n^{(j)} - \tilde{\phi}_j)}{1/\lambda_1^{(j)} - \tilde{\phi}_j)} \right) \right) \frac{\sqrt{2\pi} K c_j \omega_j}{m} \sum_{l=1}^{m} \left(\frac{h(\phi_j^{(l)}, M_j, \mathbf{y})}{\exp\{-(\phi_j^{(l)} - \tilde{\phi}_j)^2 / 2\omega_j^2\}} \right),$$

where $h(\phi_j, M_j, \mathbf{y})$ was defined in the previous section and $\phi_j^{(1)}, \ldots, \phi_j^{(m)} \overset{\text{iid}}{\sim} N(\tilde{\phi}_j, \omega_j^2)$ truncated to $(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$. Since the probability of the truncation set is very close to one (due to the choice of t), the sampling from the truncated normal distribution can be done in the obvious way: get a draw from the $N(\tilde{\phi}_j, \omega_j^2)$ distribution and accept it, unless it falls outside $(1/\lambda_n^{(j)}, 1/\lambda_1^{(j)})$. Our experience suggests that the algorithm works well regardless of how picked the integrand in (12) is, i.e., regardless of the sample size.

4 Examples

In this section we illustrate the proposed model selection methodology using two datasets, one over a regular lattice and one over an irregular lattice.

4.1 Phosphate Dataset

The first dataset we consider was analyzed by Cressie and Kapat (2008) and De Oliveira (2011). It consists of raw phosphate concentrations (in mg P/100 g of soil) collected over several years in an archaeological region of Laconia across the Evrotas river in Greece. The original observations were collected over a 16 by 16 regular lattice, and were transformed to obtain a response with distribution close to Gaussian; see the above references for further details.

As competing models for this (transformed) dataset, Cressie and Kapat (2008) entertained the HCAR, WCAR and ACAR models described in Section 2.2, each with either the first or second order neighborhood systems. In addition, we also entertain the SAR model (6) with $W = A$, having either first or second order neighborhood systems. Let $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_{256})'$ denote the transformed data. Cressie and Kapat (2008) assumed for all competing models that $E\{\tilde{Y}_i\} = \beta_1 + \beta_2 s_{i1} + \beta_3 s_{i2}$ (so $p = 3$), with (s_{i1}, s_{i2}) the coordinates of site i, while De Oliveira (2011) assumed that $E\{\tilde{Y}_i\} = \beta_1$ (so $p = 1$). We consider here both scenarios for the mean structure.

We assign to the eight competing models equal prior probabilities, and compute the posterior probabilities of these models assuming prior (8) with $\pi(\phi_j \mid M_j)$ equal to $\pi^U(\phi_j \mid M_j)$ or $(\pi^{J1}(\phi_j \mid M_j))^{1/2}$; Table 4.1 displays the results. According to this criterion, the ACAR model with first order neighborhood system displays the best fit, for both the $p = 1$ and $p = 3$ scenarios and regardless of the default prior. Also, the posterior model probabilities display little sensitivity to the two default priors. The WCAR model with first order neighborhood system ranks at a not so distant second place.

Table 1: Posterior model probabilities for the phosphate dataset. For each competing model, the results are for two mean structures and two default priors.

models	HCAR-1*	HCAR-2	WCAR-1	WCAR-2	ACAR-1	ACAR-2	SAR-1	SAR-2
				modified independence Jeffreys prior				
$p = 1$	0.099	2.2×10^{-8}	0.321	4.0×10^{-8}	0.443	5.1×10^{-8}	0.136	1.3×10^{-5}
$p = 3$	0.130	7.6×10^{-8}	0.249	9.2×10^{-8}	0.488	1.2×10^{-7}	0.132	1.9×10^{-5}
				uniform prior				
$p = 1$	0.085	4.3×10^{-7}	0.295	6.6×10^{-7}	0.416	6.6×10^{-7}	0.203	1.5×10^{-5}
$p = 3$	0.148	6.3×10^{-7}	0.221	1.6×10^{-9}	0.443	8.7×10^{-7}	0.186	2.1×10^{-5}

*HCAR-1 and HCAR-2 stand for homogeneous CAR models with, respectively, first and second order neighborhood systems. A similar convention holds for the other models.

It is worth noting that Cressie and Kapat (2008) selected the ACAR model with second order neighborhood system as the best model for this dataset by using some graphical and numerical diagnostics they developed. The above analysis does not support this finding since the posterior model probability of the ACAR-2 model is negligible for all the considered mean structures and default priors. In addition, for the models with constant and non-constant mean the maximized log-likelihoods of the ACAR-2 model are, respectively, -167.038 and -165.770, which are substantially smaller than the maximized log-likelihoods of the ACAR-1 model which are -154.686 and -153.533.

4.2 Crime Dataset

The second dataset we consider was analyzed by Baller et al. (2001) and De Oliveira and Song (2008). It consists of homicide rates per $100,000$ habitants for the year 1980 in the south of the United States, a region that forms an irregular lattice containing $n = 1412$ counties within 16 states and the District of Columbia; see the above references for further details.

Previous analyzes of this dataset showed the response mean structure to be well described by a linear model on the following explanatory variables (socio-economic factors): an index of resource deprivation, an index of population structure, median age, divorce rate and unemployment rate; we assume this to be the response mean structure.

Baller et al. (2001) and De Oliveira and Song (2008) considered SAR models and spatial lag

Table 2: Posterior model probabilities for the south crime dataset. For each competing model, the results are for two mean structures and two default priors.

models	HCAR	WCAR	ACAR	SAR
	modified independence Jeffreys prior			
AC*	4.2×10^{-6}	0**	0	0
D70	0.857	0	0	0.065
D100	3.0×10^{-3}	0	0	0.074
	uniform prior			
AC	3.6×10^{-6}	0	0	0
D70	0.822	0	0	0.074
D100	3.4×10^{-3}	0	0	0.100

*AC, D70 and D100 indicate the neighborhood systems formed by, respectively, adjacent counties, counties within a 70 miles radius and counties within a 100 miles radius.
**A value of zero indicates the estimated posterior model probability is less that 10^{-15}.

models SLM[3] with different neighborhood systems, although the analysis in De Oliveira and Song (2008) suggested that SAR models fit this dataset better than SLMs (so the latter will not be considered here). The competing models we entertain for this dataset are the HCAR, WCAR and ACAR models described in Section 2.2 and the SAR model (6) with $W = A$; all models would have the aforementioned mean structure. As for the neighborhood system, we consider the neighborhood system (1) (AC), and the neighborhood systems (2) with $r = 70$ miles (D70) and $r = 100$ miles (D100), where distance is measured between county centroids. The average number of neighbors in the three neighborhood systems are about, 5, 21 and 42, respectively.

We assign to the twelve competing models equal prior probabilities, and again compute the posterior probabilities of these models assuming prior (8) with $\pi(\phi_j \mid M_j)$ equal to $\pi^U(\phi_j \mid M_j)$ or $(\pi^{J1}(\phi_j \mid M_j))^{1/2}$; Table 4.2 displays the results. The HCAR model with the D70 neighborhood system is the one that fits the data best, regardless of the default prior that is used. Also, the posterior model probabilities display little sensitivity to the default priors. A SAR model and a SLM with ten nearest neighbors system were selected by, respectively, De Oliveira and Song (2008) and Baller et al. (2001), but no CAR models were entertained in these analyzes. The present analysis

[3]The spatial lag model is defined by the spatial auto-regressions

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} \; + \; \sum_{j=1}^{n} b_{ij} Y_j \; + \; \epsilon_i \; , \qquad i = 1, \ldots, n, \qquad \epsilon_1, \ldots, \epsilon_1 \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2).$$

shows that an HCAR model with D70 neighborhood system fits this dataset much better than any of the considered SAR models, since the posterior probabilities of all the other competing models are either very small or negligible. This suggests that the common practice among econometricians of not entertaining CAR models for the description of lattice data is unfounded.

5 Conclusions

This work describes a Bayesian model selection approach for spatial lattice models, applicable to both regular and irregular lattices. The methodology has several attractive features that make it compare favorably against other model selection approaches. First, the method does not require the competing models to be nested and can be used to compare models from different classes. Second, the approach provides an easily interpretable measure of how strongly the data support each of the competing models, through their posterior model probabilities, which is an important issue in cases when several models fit the data about equally well. Finally, the approach does not require the specification of personal prior distributions for the parameters of all competing models, and instead uses default priors. This point is relevant in practice since it is often difficult to subjectively assess these prior distributions, and posterior model probabilities are also often sensitive to the priors on the model parameters. The results on the two data analyzes suggest the latter is not the case when the default priors proposed here are used. On the other hand, the main limitation of the proposed approach is that requires all competing models to have the same mean structure, so its applicability is bounded to situations when the important explanatory variables have already been determined. Finally, it is worth noting that the proposed methodology could also be used for model selection in more complex hierarchical and/or multivariate models for lattice data, such as those proposed in Jin, Banerjee and Carlin (2007) and Song, Ghosh, Miaou and Mallick (2006).

Acknowledgements

We thank the referee for useful remarks that motivated us to write the Appendix. The second author was partly supported by a summer grant from the College of Business of The University of Texas at San Antonio and NSF grant HRD-0932339 (Drs. Demetris Kazakos and Richard Smith, project managers).

Appendix

Proposition. Let $M = \sigma^2 G$ and $C = \phi W$ be matrices satisfying conditions (c)–(d) in Section 2.2 and such that $G^{-1}W$ is symmetric. Then:

(1) The matrices $G^{-1/2}WG^{1/2}$ and W have the same non-zero eigenvalues, which are all real.

(2) The matrices M and C determine a CAR model if and only if $\sigma^2 > 0$ and $\phi \in (\lambda_n^{-1}, \lambda_1^{-1})$, where $\lambda_1 \geq \ldots \geq \lambda_n$ are the ordered eigenvalues of $G^{-1/2}WG^{1/2}$.

Proof. (1) $G^{-1}W$ symmetric implies that $G^{1/2}(G^{-1}W)G^{1/2} = G^{-1/2}WG^{1/2}$ is also symmetric, so the eigenvalues of the latter are all real. Since for any two matrices, say A and B, AB and BA have the same non-zero eigenvalues (Schott, 2005 p. 130), the eigenvalues of W are all real (even when W is not symmetric).

(2) Matrices M and C determine a CAR model iff conditions (a)–(b) in Section 2.2 are satisfied (Cressie, 1993 p. 413). $G^{-1}W$ symmetric implies condition (a) holds. From the proof of (1), all the eigenvalues of $G^{-1/2}WG^{1/2}$ are real and $\mathrm{tr}(G^{-1/2}WG^{1/2}) = \mathrm{tr}(W) = 0$, so $\lambda_n < 0 < \lambda_1$. To check condition (b) note that $M^{-1}(I_n - C) = 1/\sigma^2(G^{-1} - \phi W)$, so this matrix is positive definite iff $\sigma^2 > 0$ and $G^{-1} - \phi W$ is positive definite. Now

$$G^{-1} - \phi W = G^{-1/2}(I_n - \phi G^{-1/2}WG^{1/2})G^{-1/2},$$

so $G^{-1} - \phi W$ is positive definite iff $(I_n - \phi G^{-1/2}WG^{1/2})$ is positive definite. Since the latter is symmetric, it is positive definite iff all its eigenvalues are positive, which are given by $\{1 - \phi\lambda_i\}_{i=1}^n$. If $\phi > 0$ we have $1 - \phi\lambda_1 \geq \ldots \geq 1 - \phi\lambda_n$, so these eigenvalues are positive iff $1 - \phi\lambda_n > 0$, which is equivalent to $\phi > \lambda_n^{-1}$. On the other hand, if $\phi < 0$ we have $1 - \phi\lambda_1 \leq \ldots \leq 1 - \phi\lambda_n$, so these eigenvalues are positive iff $1 - \phi\lambda_1 > 0$, which is equivalent to $\phi < \lambda_1^{-1}$. In summary, $G^{-1} - \phi W$ is positive definite iff $\phi \in (\lambda_n^{-1}, \lambda_1^{-1})$, which proves the result.

References

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Baller, R.B., Anselin, L., Messner, S.F., Deane, G. and Hawkins, D.F. (2001), Structural Covariates of U.S. County Homicide Rates: Incorporating Spatial Effects. *Criminology*, 39, 561-590.

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

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

Berger, J.O. and Pericchi, L.R. (2001), Objective Bayesian Methods for Model Selection: Introduction and Comparisons, In *'Model Selection', Institute of Mathematical Statistics Lecture Notes–Monograph Series* volume 38 (P. Lahiri editor), Beachwood Ohio, pp 135-207.

Berger, J.O., Pericchi, L.R. and Varshavsky, J.A. (1998), Bayes Factors and Marginal Distributions in Invariant Situations, *Sankhyā A*, 86, 79-92.

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

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

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

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

Claeskens, G. and Hjort, N.L. (2008), *Model Selection and Model Averaging*. Cambridge: Cambridge University Press.

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

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

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

De Oliveira, V. (2011), Bayesian Analysis of Conditional Autoregressive Models, *Annals of the Institute of Statistical Mathematics*, to appear.

De Oliveira, V. and Song, J.J. (2008), Bayesian Analysis of Simultaneous Autoregressive Models, *Sankhyā*, 70-B, 323-350.

Florax, R.J.G.M., Folmer, H. and Rey, S.J. (2003), Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology, *Regional Science and Urban Economics*, 33, 557-579.

Hepple, L.W. (1995a), Bayesian Techniques in Spatial and Network Econometrics: 1. Model Comparison and Posterior Odds, *Environment and Planning A*, 27, 447-469.

Hepple, L.W. (1995b), Bayesian Techniques in Spatial and Network Econometrics: 2. Computational Methods and Algorithms, *Environment and Planning A*, 27, 615-644.

Jin, X., Banerjee, S. and Carlin, B.P. (2007), Order-free Coregionalized Lattice Models With Application to Multiple Disease Mapping, *Journal of the Royal Statistical Society Series B*, 69, 817-838.

Kass, R.E. and Raftery, A.E. (1995), Bayes Factors, *Journal of the American Statistical Association*, 90, 773-795.

Ord, J.K. (1975), Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, 70, 120-126.

Osiewalski, J. and Steel, M.F.J. (1993), Regression Models Under Competing Covariance Structures: A Bayesian Perspective, *Annales d'Economie et de Statistiques*, 32, 65-79.

Song, J.J., Ghosh, M., Miaou, S. and Mallick, B (2006), Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping, *Journal of Multivariate Analysis*, 97, 246-273.

Zhu, J., Huang, H-C. and Reyes, P.E. (2010), On Selection of Spatial Linear Models for Lattice Data, *Journal of the Royal Statistical Society B*, 72, 389-402.