ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Spatial dependence in Bayesian Analysis

D. Sheeba Singh and **M. Immaculate Mary**
Department of Mathematics, Noorol Islam University, Kumaracoil,
Kanyakumari District, Tamilnadu, India
***Corresponding Author: E-mail: dssheeba@gmail.com**

ABSTRACT

In several situations, the averages are observed and the regression analysis is performed for the geographically defined regions. Generally, it is assumed that the observations on the location that are near each other may approach a similar score on the excluded variable in the regression which will cause the error terms that are spatially auto-correlated. In this paper we have studied the consequences of adopting the conditional auto regressive (CAR) and intrinsic conditional autoregressive (ICAR) prior distributions in Bayesian spatial regression models. Also we have obtained an analytical expression which interprets the effect of the neighborhood structure in the posterior covariance matrix of the spatial random effects.

Keywords: Random effects, autocorrelation, adjacency matrix.

1. INTRODUCION

Spatial dependency guides to the spatial autocorrelation problem in statistics since, like temporal autocorrelation. With the spatial autocorrelation coefficient, we can compute the proximity of location and the resemblance of the characteristics of the places. In Bayesian data Analysis, CAR and ICAR models as prior distributions, are important in studying the covariance behavior for spatial random effects. The CAR model was developed by Besag (1974) as the model on the doubly infinite regular lattice.

2. THE CAR AND ICAR MODELS

Let R be the spatial domain, partitioned into n areal units $R_1, R_2, \ldots, R_n$ such that $R = R_1 \cup R_2 \cup \ldots \cup R_n$ and $R_i \cap R_j = \varphi$, for every $i \neq j$. Let $X = \{x_1, x_2, \ldots, x_n\}^T$, where x_i is the random variable measured at area i. Denote $x_{-i} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ is the (n-1) dimensional vector without the i^{th} coordinate of X.A random field X is called the Conditional Auto Regressive Model, if it is characterized by the conditional distributions;

$x_i / x_{-i} \sim N(\mu_i + \sum_{j \in R} \varphi_{ij}(x_j - \mu_j), \sigma_i^2), i \in R$ (1)where$\varphi_{ii} = 0$, $\sigma_i^2 > 0$ for i = 1,2,…,n.

The measure of spatial proximity between the areal unit i and j is given by the matrix $\varphi = (\varphi_{ij})$. Let $A = (a_{ij})$ be an n×n symmetric weighted adjacency matrix, such that $a_{ij} = 1$ if and only if the location i and j are neighbours, and $a_{ii} = 0$. Define $W = (w_{ij})$ where $w_{ij} = \frac{a_{ij}}{\beta_i}$ and denote $\beta_i = \sum_j a_{ij}$, $\sum_j a_{ij}$ give the number of location close to the region i. Also define $\phi = \rho W$ and $d_i = \frac{\sigma^2}{\beta_i}$ and the restriction of ρ.Under these assumptions, the CAR Model (1) is defined and the vector X follows a multivariate normal distribution,

(i.e) $X \sim N(\mu, (I - \rho W)^{-1} D)$ (2)

where $\mu = (\mu_1, \mu_2, \ldots, \mu_n)^T$, I and D are the identity and diagonal matrices.

(i.e.) $D = dia(d_1, d_2, ..., d_n) = \sigma^2 dia(\frac{1}{\beta_1}, \frac{1}{\beta_2}, ..., \frac{1}{\beta_n})$.

The restriction of ρ ensures that the matrix $(I - \rho W)^{-1} D)$ is positive definite. It is sufficient that the value of ρ lies between $\frac{1}{\min \lambda_i}$ and $\frac{1}{\max \lambda_i}$ where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigen values of W. The equation (2) reduces to

$$x_i / x_{-i} \sim N(\mu_i + \rho(x - \mu)_i, \frac{\sigma^2}{\beta_i}), i \in R$$

where $(x - \mu)_i = \sum_j w_{ij}(x_j - \mu_j)$ is the average of the deviation $x_j - \mu_j$ among the neighbor's areas. The entries of the correlation matrix are the functions of W and ρ only.

$$cor(x_i, x_j) = \frac{\sqrt{\beta_i}}{\beta_j} \frac{(I - \rho W)_{ij}^{-1}}{\sqrt{(I - \rho W)_{ii}^{-1} (I - \rho W)_{jj}^{-1}}}$$

When ρ = 1, this CAR model is reduced to Intrinsic CAR model (ICAR). In ICAR, the covariance matrix is not positive definite. In this case, a set of conditional distributions is obtained given by

$$x_i / x_{-i} \sim N(\mu_i + \rho(x - \mu)_i, \frac{\sigma^2}{\beta_i}), i \in R$$

The set of conditional distributions is proportional to the following when

$$\exp\left(-\frac{1}{2\sigma^2}\sum_{i,j} a_{ij}(x_i-\mu_i-(x_j-\mu_j))^2\right.$$

$$=\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^T D^{-1}(I-W)(x-\mu)'\right. \qquad (3)$$

Equation (3) is integrable if a linear constraint $\sum_i (x_i-\mu_i)=0$ is added.

Hence, a proper n-dimensional vector is obtained which is normally distributed and its probability mass function depends on the (n-1) vector subspace orthogonal to the unit vector 1= (1, 1, 1… 1). Therefore, the inverse of its co-variance matrix does not exist. The major results corresponding to the correlations given by the CAR model are summarized here. The implications that the CAR model involves the covariance between the pairs of areas using the Tamil Nadu map is demonstrated here. Consider the graph formed by the thirty contiguous districts of Tamilnadu. The graph is obtained by joining the adjacent districts of Tamilnadu. If two districts i and j share the border, they are connected by an edge, (i.e. wij> 0) as shown in figure 1.

Figure.1. Model Involves the Covariance between the Pairs of Areas Using the Tamil Nadu Map

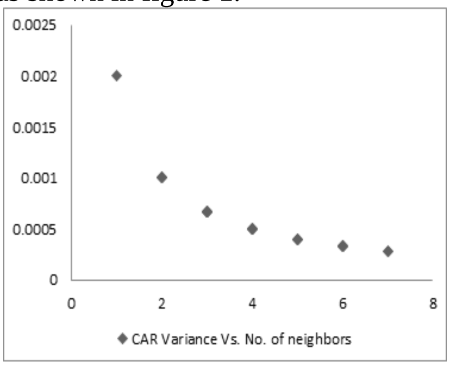

Figure.2.Shows the number of neighbors against the variances ρ = 0.1444

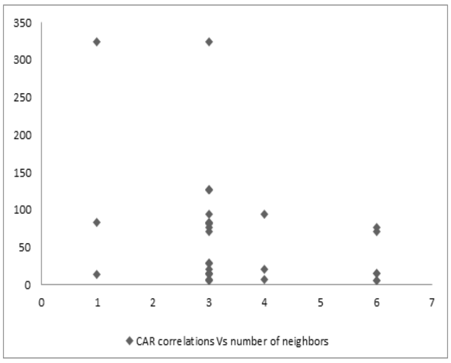

Figure.3.shows the number of neighbors against the correlations between the pairs of adjacent districts

The neighbouring districts denoted as the pair of points (i, j) contribute two values in this graph depending on the number of neighbours in each region. It means that for one pair of districts (i, j), there are two plots such as (βi, cor (i, j)) and (βj, cor (i, j)). Hence, there is a vast variation in the correlations and variances for a given number of neighbours.

From Table.1, it becomes clear that as the value of ρ increases, the correlation also increases. When ρ = -0.7112, the correlation between Tirunelveli and Kanyakumari districts is -0.20575, whereas the correlation between Tirunelveli and Tuticorin districts is -0.19673 and the correlation between Tirunelveli and Virudhunagar districts is -0.17974.When ρ = 0.8556, the correlation between Tirunelveli and Kanyakumari districts is 0.552665, the correlation between Tirunelveli and Tuticorin districts is 0.265858 and the correlation between Tirunelveli and Virudhunagar districts is 0.168249.Thus, as the value of ρ changes, the correlation is also changed. Let us look at the behaviour of correlation between the locations i and j when ρ tends to be at its lower bound -0.7112. The correlation of the pair of locations tends to be either 1 or -1. When the spatial dependence parameter ρ is positive, the correlation is monotonically increasing.

The weighted adjacency matrix will be treated as the transition probability matrix of a Markov Chain defined on a graph.Let the ‘n’ areal units be the vertices of the graph. The areal units R1, R2…Rn is connected by undirected edges.Represent the n areal units as nodes or vertices and the areal units R1, R2,…,Rn are connected by undirected edges. If the areal units i and j are connected by an edge then $W_{ij} \neq 0$.

Let Wnbe the transition probability matrix for the chain after ‘n’ steps. Let P be a square matrix. If the entries of Pn tend to zero as n tends to ∞, then (I-P)-1 will exist and is given by

$$(1-P)^{-1} = 1 + P + P^2 + \ldots$$

Let us take $P = \rho W_T$ where $|\rho| < 1$ and all w_{ij}’s such that $0 \le w_{ij} \le 1$, for all i, j and for all integers n.

$$(1-\rho W_T)^{-1} = 1 + \rho W_T + \rho^2 W_T{}^2 + \rho^3 W_T{}^3 + \ldots.$$

$[W_T^m]_{ij} > 0$ gives the probability of reaching the state j from the state i in m steps and it is positive. It means that there exists a series of ‘m’ edges joining the neighbours whose starting and ending nodes are i and j respectively.

This path is the m^{th} order path connecting the nodes i and j. The value denotes the weighted sum of all the m^{th} order paths between the nodes i and j.
To explain this,

$$[W_T{}^2]_{ij} = \sum_{r=1}^{n} W_{Tir} W_{Trj} = \sum \frac{a_{ir}}{\beta_i}\frac{a_{rj}}{\beta_r} = \frac{1}{\beta_i}\sum \frac{a_{ir}a_{rj}}{\beta_r}$$

The product $a_{ir}\, a_{rj}$= 1, if and only if the nodes i and j are connected by 'r'.Hence, $[W^2]_{ij}\propto$ is a weighted sum of all second- order paths $i\rightarrow r \rightarrow j$.Every path gives a value which is proportional to the number of adjacent regions β_r present in the intervening area of r. In particular, as every region has at least one neighbour, there exists at least one path of the type $i \rightarrow r \rightarrow i$ (ie) $[W_T{}^2]_{ij} > 0$.
In a similar way,

$$[W_T{}^3]_{ij} = \sum_{p=1}^{n}[W_T{}^2]_{ip} W_{Tpj} = \frac{1}{\beta_i}\sum_{p=1}^{n}\sum_{r=1}^{n} a_{ir} a_{rp}\frac{a_{pj}}{\beta_r\beta_p}.$$

Here, every path connecting the nodes i and j is weighted inversely by the number of neighbours connecting the graph at 'r' and 'p'. Also the paths of the type $i \rightarrow j \rightarrow i \rightarrow j$ are taken into account.Using all the above results and assuming $|\rho| < 1$,

We get

$$[(I-\rho W_T)^{-1}]_{ij} = [I]_{ij} + \rho[W_T]_{ij} + \rho^2[W_T{}^2]_{ij} + \rho^3[W_T{}^3]_{ij} + \ldots \quad (4)$$

Each term of the above series represents the probability that lies between 0 and 1. Also $[W^r]_{ij}$ tends to the limit $\frac{\beta_j}{\beta}$ for all i and the speed of convergence of the above series is determined for all i and j and is given by the second largest Eigen value modulus of W_T. Expression (4) is approximately,

$$[(I-\rho W_T)^{-1}]_{ij} \approx [I]_{ij} + \rho[W_T]_{ij} + \rho^2[W_T{}^2]_{ij} + \ldots + \rho^{r-1}[W_T{}^{r-1}]_{ij} + \frac{\beta_j\rho^r}{\beta(1-\rho)}, \text{ for some r}$$

$$\approx [I]_{ij} + \rho[W_T]_{ij} + \rho^2[W_T{}^2]_{ij} + \ldots + \rho^{r-1}[W_T{}^{r-1}]_{ij} \quad (5)$$

If in equation (5), third degree approximation is sufficient, then the covariance between the areal units i and j will approximately be given by

$$\mathrm{Cov}(x_i, x_j) \approx \frac{\sigma^2}{\beta_j}\left(\frac{\rho a_{ij}}{\beta_i} + \frac{\rho^2}{\beta_i}\sum_{r=1}^{n}\frac{a_{ir}a_{rj}}{d_i} + \frac{\rho^3}{\beta_i}\sum_{i=1}^{n}\sum_{r=1}^{n}\frac{a_{ir}a_{rp}a_{pj}}{\beta_p\beta_r}\right)$$

In the equation (4), if $\rho > 0$, then all the coefficients in the given series expansion are non-negative. Hence, the correlation increases monotonically with ρ.Also, in equation (4), series expansion coefficients are given for a specific pair, the correlation of different pairs will increase at different rates.

Differentiate (4) partially with respect to ρ.

$$\frac{\partial}{\partial\rho}[(I-\rho W_T)^{-1}]_{ij} = [W_T]_{ij} + 2\rho[W_T{}^2]_{ij} + 3\rho^2[W_T{}^3]_{ij} + \ldots\ldots$$

When $\rho\in(0,1)$, the derivative is increasing with ρ. If ρ is small and not too close to 1, then the increasing rate of the above derivative will be dependent on $[W_T{}^2]_{ij}$, the second order neighbours. Consider a specific pair of areas (i, j) for finding the first order derivative. So, the derivative is depending on that specific pair (i, j). Approximately taking up to the second degree terms,

$$\frac{\partial}{\partial\rho}[(I-\rho W_T)^{-1}]_{ij} \approx [W_T]_{ij} + 2\rho[W_T{}^2]_{ij}$$

Therefore, if ρ is large, then it will contribute a positive quantity to the correlation of the second order neighbourhoods. If ρ is small, then the pair (i, j) will have a small amount of correlation that may become more in other areas as in the coefficient of the second order coefficient $[W_T{}^2]_{ij}$ which is relatively large. This is the justification for the extraordinary behaviour of changing ranks between the correlations of Tirunelveli and Kanyakumari, Tirunelveli and Tuticorin, and Tirunelveli and Virudhunagar.

3. COMPUTATIONAL ANALYSIS

The neighbourhood adjacency matrix W_T is stationary for r =167.
When ρ = .1444,

For Tirunelveli and Kanyakumari, $(I-\rho W_T)^{-1} = .05\rho + 0.\rho^2 + .005\rho^3 + ...$
For Tirunelveli and Tuticorin, $(I-\rho W_T)^{-1} = .05\rho + 0.001\rho^2 + .0006\rho^3 + ...$
For Tirunelveli and Virudhunagar, $(I-\rho W_T)^{-1} = .05\rho + 0.002\rho^2 + .0008\rho^3 + ...$
When $\rho = .8556$,
For Tirunelveli and Kanyakumari, $(I-\rho W_T)^{-1} = .28\rho + 0.\rho^2 + .006\rho^3 + ...$
For Tirunelveli and Tuticorin, $(I-\rho W_T)^{-1} = .29\rho + 0.04\rho^2 + .03\rho^3 + ...$
For Tirunelveli and Virudhunagar, $(I-\rho W_T)^{-1} = .28\rho + 0.08\rho^2 + .06\rho^3 + ...$

Let us tabulate the elements of $\rho^r W_T{}^r$ and the cumulative sum $\sum_{j=0}^{r} \rho^j W_T{}^j$ for the pairs of neighbours of the district Tirunelveli. Also we calculate the correlation between the areal units. Here, we take the values of ρ as 0.1444 and 0.8556.

From the calculations in TABLE 2, the coefficients decline slowly for the region whose major portion is connected with the neighbours and then the neighbours whose major position is not adjacent with Tirunelveli. Consequently, for small values of r, the first few terms dominate the series. Taking the first order approximation with r, within 99%, 96% and 94% of their limiting values equal to 0.048635 for the pair Tirunelveli and Kanyakumari, equal to 0.049975 for the pair Tirunelveli and Tuticorin and equal to 0.015256 for the pair Tirunelveli and Virudhunagar respectively.

Using third order approximation, with r = 3, only 1% (for the pair Tirunelveli and Kanyakumari), 1.37% (for the pair Tirunelveli and Tuticorin) and 1.57% (for the pair Tirunelveli and Virudhunagar) from their limiting values respectively.

When $\rho = 0.8556$, the convergence of $(I-\rho W_T)^{-1}$ is slow because fairly large r^{th} order neighbouring regions provide a reasonable amount to the series sum.
For r =1, we are within 80%, 72% and 52% for the pairs (Tirunelveli and Kanyakumari, Tirunelveli and Tuticorin, Tirunelveli and Virudhunagar) from their limiting values respectively.

For r = 3, the values are much smaller, for r = 10, within 1.3% for the pair Tirunelveli and Kanyakumari , 0.03% for the pair Tirunelveli and Tuticorin and 4.4% for the pair Tirunelveli and Virudhunagar respectively from their limiting values. It presents that the geographically far-away neighbourhood structures revealed in the r-steps paths from i to j in the $W_T{}^r$ entries, have a significant impact on the series limits, since these paths are different for the three pairs of areas. The end result is that the first ordering of correlations when $\rho = 0.1444$ is changed, as ρ increases to 0.8556.

Now, let us find the relationship between the variances var (y_i) and the number of neighbours β_i of first order neighbours.

Wall (2004) commended that there is an exacting negative association between these two quantities but that there is also variation of var (y_i) among the areal units with equal number of neighbour β_i which is given in Fig. 2.

Also using the equation
$(I-\rho W_T)^{-1}_{ij} \approx [I]_{ij} + \rho[W_T]_{ij} + ... + \rho^{r-1}[W_T{}^{r-1}]_{ij} + \frac{\beta_j \rho^r}{D(1-\rho)}$,the above behaviour can be clarified. Suppose $W_T{}^r$ converges very fast such that var (y_i) can be approximated upto the ρ^2 terms. Then,

$$\mathrm{var}(y_i) = \frac{\sigma^2}{\beta_i}\left(1 + \frac{\rho^2}{\beta_i}\sum_r \frac{a_{ir}a_{ri}}{d_r}\right).$$

The declining value of var (y_i) with β_i is obvious. The sum $\sum_r \frac{a_{ir}a_{ri}}{\beta_r}$ depends on its number of terms. That is it depends on the number of first order neighbours β_i. It also depends on the degree of connectedness of these neighbours through this β_r values. Suppose the area 'i' has a single neighbour namely the areal unit r, then

$$\mathrm{Var}(y_i) \approx \sigma^2 (1 + \frac{\rho^2}{\beta_r}).$$

The ICAR model is not a probability distribution without any constraints imposed on it and hence it does not have moments. If $\mu_i = 0$ and the linear constraint $\sum_i x_i = 0$, the vector X follows a multivariate normal distribution with mean $\mu = 0$ and the covariance matrix is non-invertible which is given by $\sum_{i=2}^{n} \frac{1}{\lambda_i} a_i a_i{}'$ where (λ_i, a_i) are the (n-1) pairs of non-zero Eigen values and normalized Eigen vectors of $D^{-1}(I-W_T)$. This normal distribution has its probability density function dependent on the (n-1) vector space orthogonal to the vector **1** = (1,1,…,1). Hence, it is not viable to express the covariance matrix $D^{-1}(I-W_T)$ as a function of the W_T matrix.

Table.1. Co Efficient Value

	$\rho = -0.7112$	$\rho = 0.01$	$\rho = 0.1444$	$\rho = 0.5$	$\rho = 0.8556$
Tirunelveli Vs. Kanyakumari	-0.20575	0.00571	0.08343	0.29606	0.55266
Tirunelveli Vs. Kanyakumari	-0.19673	0.00190	0.02863	0.11386	0.26585
Tirunelveli Vs. Kanyakumari	-0.17974	0.00095	0.01468	0.06632	0.16825

Table.2. r Value

ρ = 0.1444	r = 1	r = 2	r = 3	r = 4	r = 5	r = 10	r = 30	r = 50	r = 100
(Tirunelveli and Kanyakumari) $\rho^r W_T^{\ r}$	0.048133	0	0.000502	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cum. Sum	0.048133	0.048133	0.048635	0.048635	0.048635	0.048635	0.048635	0.048635	0.048635
Approximate Correlation	0.083268	0.165819	0.245796	0.322282	0.39436	0.673736	0.95993	0.985388	1.080941
(Tirunelveli and Tuticorin) $\rho^r W_T^{\ r}$	0.048133	0.001159	0.000683	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cum. Sum	0.048133	0.049292	0.049975	0.049975	0.049975	0.049975	0.049975	0.049975	0.049975
Approximate Correlation	0.028636	0.057872	0.087382	0.116792	0.145948	0.278152	0.514592	0.558378	0.57453
(Tirunelveli and Virudhunagar) $\rho^r W_T^{\ r}$	0.048133	0.002317	0.000806	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Cum. Sum	0.048133	0.050450	0.051256	0.051256	0.051256	0.051256	0.051256	0.051256	0.051256
Approximate correlation	0.014685	0.030059	0.045947	0.062207	0.082051	0.15741	0.320951	0.368921	0.398657
ρ= 0.8556									
(Tirunelveli and Kanyakumari) $\rho^r W_T^{\ r}$	0.2852	0	0.014411	0.006592	0.043834	0.004562	0.000104074	0.0000	0.0000
Cum. Sum	0.2852	0.2852	0.299611	0.306203	0.350037	0.354599	0.354703074	0.354703074	0.354703074
Approximate correlation	0.552665	0.857702	0.958873	0.986727	0.994793	0.999954	1.000013	1.00002	1.000019
(Tirunelveli and Tuticorin) $\rho^r W_T^{\ r}$	0.2852	0.040702	0.0342117	0.021725	0.010978	0.0013266	0.000299212	0.0000	0.0000
Cum. Sum	0.2852	0.325902	0.3601137	0.3818387	0.3928167	0.3941433	0.394442512	0.394442512	0.394442512
Approximate Correlation	0.265858	0.466839	0.543609	0.566238	0.573081	0.577153	0.577358	0.57735	0.57735
(Tirunelveli and Virudhunagar) $\rho^r W_T^{\ r}$	0.2852	0.081331	0.061331	0.051703	0.045004	0.024198	0.00057891	0.0000	0.0000
Cum. Sum	0.2852	0.366531	0.427862	0.479565	0.524569	0.548767	0.54934591	0.54934591	0.54934591
Approximate Correlation	0.168249	0.306154	0.365599	0.388639	0.398341	0.407391	0.408248	0.408244	0.408244

4. CONCLUSION

In Spatial Bayesian models, the prior distributions used are ICAR and CAR models. The covariance matrix of the prior distribution is splitted into many terms. This splitting is applicable only when the correlation parameter ρsatisfies the conditions in sensible cases with in the specified intervals. This arrangement explains how the complete neighbourhood formation of the map changes the covariance between two bordering states. The higher degree terms of the infinite sum represents the covariance of two distant neighbours i and j. These higher degree terms describe the paths from i to j connecting the distant neighbour. Overall the CAR and ICAR models are very useful to recognize the influence of the variance matrix and the other parameters to the covariance matrix.

REFERENCES

Alane E. Gelfand, Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis, Biometric, 4(1), 2003, 11–25.

Carsten F. Dormann, Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: a Review, Ecography, 30, 2007, 609-628.

Dongchu Sun, Robert K Tsutakawa and Paul L Speckman, Posterior Distribution of Hierarchical Models Using CAR Distributions, Biometrika, 86(2), 2005, 341-350.

Gentry White and Sujit K. Ghosh, A Stochastic Neighbourhood Conditional Autoregressive Model for Spatial Data, Biostatistics, 4(1), 2003, 11-15.

James's. Hodges, Bradley P. Carlin and Qiao Fan, On the Precision of the Conditionally Autoregressive Prior in Spatial Models, Biometrics, 59(2), 2002, 317-322.

Julian Besag, Spatial Interaction and the Statistical Analysis of Lattice Systems, Journal of the Royal Statistical society, 36(2), 1974, 192-236.

Minjung Kyung and Sujit K. Ghosh, Bayesian Inference for Directional Conditionally Autoregressive Models, Bayesian Analysis, 4(4), 2009, 675-706.

Renato Assunção, Elias Krainski, Neighborhood Dependence inBayesian Spatial Models, Biometrical Journal, 51(5), 2009, 851-869.

Victor De Oliveira, Bayesian Analysis of Conditional Autoregressive Models, Annals of the Institute of Statistical Mathematics, 64, 2012, 107–133.