Wiley StatsRef:
Statistics Reference Online

Conditional Autoregressive (CAR) Model

By *Alexandra M. Schmidt*[1] *and Widemberg S. Nobre*[2]

Keywords: *areal data level; Bayesian inference; disease mapping; Gaussian Markov random fields; image analysis; lattice data; spatial statistics*

Abstract: Conditional autoregressive (CAR) models are useful to obtain a multivariate **joint distributions** of a random vector based on univariate conditional specifications. These conditional specifications are based on Markovian properties such that the conditional distribution of a component of the random vector depends only on a set of neighbors. Conditional autoregressive models are particular cases of Markov random fields. CAR models have been applied in different areas of science; some examples are image analysis, epidemiology, and agriculture. Typically, Gaussian CAR specifications are used as latent structures in **hierarchical models** for areal level data. In this case, the region of interest is divided into a set of disjoint areas and a CAR random effect is used to account for possible correlation among observations made across the different areas.

1 Introduction

Conditional autoregressive (CAR) models were initially introduced by Besag[1]. They are useful to describe observations that vary over a discrete set of indices, such as the number of cases of a disease across the counties of a state, brain activity over pixels, agricultural field experiments, and **remote sensing**, among others. CAR models have been widely used over the last 30 years. They are commonly used as latent components of **hierarchical models** to capture possible correlations among the observations of interest.

Consider a random vector $U = (U_1, \ldots, U_n)$ and assume that each component is univariate and located at a fixed site $i \in \{1, 2, \ldots, n\}$. Note that the sites might represent points or regions in space, and the components of U might be continuous or discrete. CAR models allow one to obtain the multivariate distribution for U from univariate conditional specifications that are based on *local* properties of the region of interest. More specifically, this is done by assuming that the conditional distribution $p(u_i \mid u_{j\neq i})$ depends only on the sites that are neighbors of site i. Formally, site j is called a *neighbor of site* i if and only if the conditional distribution at site i depends on the value at site j[2]. A system of sites, each with specified neighbors, define a graph. As outlined by Besag[2], establishing the conditional probability distributions is related to the choice of a graph, and the selection of appropriate distributions, consistent with that graph.

[1] McGill University, Montreal, Canada
[2] Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

DOI: 10.1002/9781118445112.stat08048

CAR models are particular cases of **Markov random fields** (MRF). For more discussion on MRF, see for example, Ref. 3[Chapters 6 and 7], and Ref. 4, which provides a modern view on the use of Gaussian MRF.

2 Gaussian Conditional Autoregressions

A random vector $U = (U_1, \ldots, U_n)^T$ is said to follow a CAR model if

$$U_i \mid u_{-i} \sim N\left(\sum_{j\in\partial i} a_{ij}u_j, \tau_i^2\right) \tag{1}$$

where $u_{-i} = \{u_j, j \neq i\}$, and ∂_i denotes the set containing the indices that are neighbors of i. Our aim is to obtain the **joint distributions** $p(u)$ from the conditional specifications in Equation (1). As shown in Ref. 1, this is achieved by using Brook's lemma[5].

2.1 Brook's lemma

Let $p(u)$ be the density for $u \in \mathbb{R}^n$ and define $\Omega = \{u \in \mathbb{R}^n : p(u) > 0\}$. Let $u, u' \in \Omega$, then

$$\frac{p(u)}{p(u')} = \prod_{i=1}^{n} \frac{p(u_i \mid u_1, \ldots, u_{i-1}, u'_{i+1}, \ldots, u'_n)}{p(u'_i \mid u_1, \ldots, u_{i-1}, u'_{i+1}, \ldots, u'_n)} \tag{2}$$

The uniqueness of $p(u)$ can be determined by keeping $p(u')$ fixed; for simplicity assume that $u' = 0$[4]. Substituting the conditional specification in Equation (1) onto Equation (2) with $u' = 0$, we have that

$$\begin{aligned}\frac{p(u)}{p(0)} &= \prod_{i=1}^{n} \frac{\exp\left\{-\frac{1}{2\tau_i^2}\left(u_i - \sum_{j<i} a_{ij}u_j - \sum_{j>i} a_{ij}0_j\right)^2\right\}}{\exp\left\{-\frac{1}{2\tau_i^2}\left(0_i - \sum_{j<i} a_{ij}u_j - \sum_{j>i} a_{ij}0_j\right)^2\right\}} \\ &= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\frac{1}{\tau_i^2}\left(u_i^2 - 2\sum_{j<i} a_{ij}u_j\right)\right]\right\}\end{aligned}$$

and if $\frac{a_{ij}}{\tau_i^2} = \frac{a_{ji}}{\tau_j^2}, \forall i, j \in \{1, 2, \ldots, n\}$, then $2\sum_i\sum_{j<i} u_i\frac{a_{ij}}{\tau_i^2}u_j = \sum_i\sum_j u_i\frac{a_{ij}}{\tau_i^2}u_j$, and we obtain that

$$p(u_1, u_2, \ldots, u_n) \propto \exp\left\{-\frac{1}{2}\ u^T D^{-1}(I - A)u\right\} \tag{3}$$

where D is a diagonal matrix with $D_{ii} = \tau_i^2$, and $A_{ij} = a_{ij}, i, j = 1, 2, \ldots, n$. Usually the matrix A is related to a proximity, or adjacency matrix, denoted by W, which describes the neighborhood structure of the region of interest. Besag *et al.*[6] propose an *intrinsic Gaussian autoregression model* (IAR or ICAR) by defining $a_{ij} = \frac{W_{ij}}{W_{i+}}$, with $W_{ii} = 0$, $W_{i+} = \sum_j W_{ij}$, and $\tau_i^2 = \tau^2/W_{i+}$, such that

$$U_i \mid u_{-i} \sim N\left(\frac{\sum_{j\in\partial i} W_{ij}\ u_j}{W_{i+}}, \frac{\tau^2}{W_{i+}}\right) \tag{4}$$

where the mean of each conditional distribution is given by a weighted average of the values of the neighboring areas, and the conditional variance is inversely proportional to the sum of the weights for each area.

Wiley StatsRef: Statistics Reference Online, © 2014–2018 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08048

Following Equation (3), under the ICAR specification the joint density of U is proportional to

$$p(u) \propto \tau^{-g} \exp\left\{-\frac{1}{2\tau^2}\, u^T(D_w - W)\, u\right\} \equiv \tau^{-g} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^{n} \sum_{j\in\partial_i} W_{ij}(u_i - u_j)^2\right\} \tag{5}$$

where D_w is a diagonal matrix with elements $(D_w)_{ii} = W_{i+}$, and the value of g is discussed in Section 4.

Note that the distribution in Equation (5) is improper. We can add any constant to the elements of U and the density does not change. Similarly, the precision matrix $(D_w - W)$ is not of full rank. For this reason, the ICAR model is not used to define the distribution of observed data. Although the joint distribution of U is improper, all the conditional distributions $U_i \mid u_{-i}$ are proper. One way to make this distribution proper is to impose a sum-zero constraint, that is, $\sum_i u_i = 0$. See Ref. 7 for further discussion on different constraints that can be used. Typically, ICAR models are used to assign the **prior distributions** for spatially varying **random effects** in hierarchical models.

2.2 Neighborhood structure

In the ICAR specification, the weighting matrix W is symmetric, and W_{ij} represents nonnegative weights. Note that it plays an important role as the smoothness of the CAR model depends on this weighting matrix. Commonly W is assumed to be known. The most common approach is to assume $W_{ij} = 1$ if i and j are contiguous areas, denoted by $i \sim j$, and $W_{ij} = 0$, otherwise. This is also known as a *first-order neighborhood structure*. In this case, $U_i \mid u_j, j \in \partial_i \sim N\left(\frac{1}{N_i}\sum_{j\in\partial_i} u_j, \frac{\tau^2}{N_i}\right)$, where ∂_i is the set index containing the neighboring areas of i and N_i is the number of neighboring areas. Typically, neighboring matrices based on 0–1 neighborhood structures are sparse. Note further that, for a given value of τ^2, the greater the number of neighbors of an area i, the smaller the conditional variance of $U_i \mid u_{-i}$; therefore, the more concentrated is the mean of the conditional distribution around the average of its neighbors. Another common approach is to assume a distance-based neighborhood matrix, for example, $W_{ij} = 1/d_{ij}$, where d_{ij} is the Euclidean distance between the centroids of areas i and j. Ferreira and Schmidt[8] investigate the use of different neighborhood structures that take into account the particular landscape of the region under study. They noticed that the importance of some covariates in the mean structure changed when different neighborhood structures were assumed. Earnest *et al.*[9] examine the influence of different neighborhood weight matrix structures on the amount of **smoothing** performed by the CAR model and found considerable differences. They recommend an exploratory analysis of the neighborhood weight matrix to guide the choice of a suitable one.

Care must be taken when there are islands in the region under study. In this case, there will be groups of areas that are disconnected from the other areas. A simple solution is to connect these disconnected areas to the main region. Freni-Sterrantino *et al.*[10] provide practical guidelines about how to define these models in this case.

2.3 Proper CAR (PCAR) Specification

As mentioned earlier, the ICAR specification leads to an improper joint distribution for U as $(D_w - W)^{-1}$ is singular. One way to turn it into a proper distribution is to assume a scaled version of the known neighboring matrix, ρW, such that $(D_w - \rho W)$ is nonsingular, and Equation (5) becomes the kernel of a zero mean **multivariate normal** distribution with covariance matrix $\Sigma_u = (D_w - \rho W)^{-1}$. In this case, ρ must be chosen in a way that Σ_u^{-1} is nonsingular. Banerjee *et al.*[11] discuss different approaches. A common one is to assume $\rho \in (1/\gamma_{(1)}, 1/\gamma_{(n)})$, where $\gamma_{(1)}$ and $\gamma_{(n)}$ are, respectively, the minimum and maximum eigenvalues

DOI: 10.1002/9781118445112.stat08048

of $D_w^{-1/2} W D_w^{-1/2}$. In this case, the full conditional is given by

$$U_i \mid u_{-i} \sim N\left(\rho \sum_{j \in \partial_i} \frac{W_{ij}\, u_j}{W_{i+}}, \frac{\tau^2}{W_{i+}}\right) \tag{6}$$

and the joint distribution of U is $N(0, \tau^2(D_w - \rho W)^{-1})$. As $\Sigma_u^{-1} = \tau^{-2}(D_w - \rho W)$ is the precision matrix of the distribution of U, it follows that $\text{Var}(U_i \mid u_j, j \neq i) = 1/(\Sigma_{u_{ii}}^{-1})$. Therefore, if $W_{ij} = 0$, then U_i and U_j are conditionally independent given $u_k, k \neq i, j$. Usually, CAR models induce sparse precision matrices making the computation of these matrices very fast in the inference procedure. The ICAR model is the limiting case when $\rho = 1$. Also, it can be shown that 0 belongs to $(1/\gamma_{(1)}, 1/\gamma_{(n)})$, leading to an independent model for the U_i's if $\rho = 0$, and in this case $U_i \sim N(0, \tau^2/w_{i+})$.

Besag and Kooperberg[12] call attention to the fact that the marginal variances vary with i, and all neighboring pairs have different covariances. Wall[13] mentions that the resultant correlation structure based on a proper CAR distribution does not seem to follow an intuitive scheme associated with the neighborhood matrix W. This is because $\tau^{-2} D_w(I - \rho W)$ is the precision matrix, whereas the resultant covariance matrix is a function of the inverse of $\tau^{-2} D_w(I - \rho W)$. Assunção and Krainski[14] investigate this further by decomposing the covariance matrix into an infinite sum of terms. Their results show that the entire neighborhood structure of the map affects the covariance of a pair of neighboring areas.

Section 6.4.3.3 of Banerjee *et al.*[11] discusses the choice between a proper and an improper CAR. In ICAR, the mean of the full conditional distribution is given by a weighted average of the process values of the neighboring areas, whereas in PCAR this mean is given by some proportion of this average. As noted by different authors, PCAR models tend not to allow high correlations between neighbors unless ρ is very close to 1.

Under the PCAR specification, some authors have proposed models that allow the neighborhood structure to depend on unknown parameters and be estimated from the data. See for example, Refs 15–17.

3 Non-Gaussian Conditional Autoregression

Besag[1] also discusses conditional autoregression model in the more general case where the conditional distribution $(U_i \mid u_{-i})$ belongs to the **exponential family**. Besag[1] discusses the cases of the autologistic, autobinomial, auto-Poisson, and autoexponential CAR models. Banerjee *et al.*[11] and Held and Rue[18] mention that the autologistic CAR model has gained some popularity in the literature (see e.g., Refs 19–21). In particular, Caragea and Kaiser[22] propose an alternative parametrization of the autologistic model as they show that it presents difficulties in interpreting model parameters across varying levels of statistical dependence. On the other hand, Rue and Held[4] point to the fact that auto-Poisson models can only model negative dependence between neighboring sites, and for this reason have not been much used.

An alternative to account for spatial correlation among observations that are generated from distributions that belong to the exponential family is to include a random (latent) effect in the mean structure of the distribution. And, *a priori*, this latent effect follows a Gaussian CAR distribution following either an ICAR or a PCAR model. More specifically, consider that y_i is an outcome of interest observed at area i, and x_i is a p-dimensional vector of covariates. As the observations are assumed to be generated from a distribution belonging to the exponential family, then

$$f(y_i \mid \theta_i) = \exp\left\{\frac{y_i \theta_i - b(\theta_i)}{\phi_i}\right\} c(y_i, \phi_i) \tag{7}$$

Wiley StatsRef: Statistics Reference Online, © 2014–2018 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08048

The means $\mu_i = E(y_i \mid \theta_i)$ are related to the canonical parameters θ_i via $\mu_i = b'(\theta_i)$, and to the regression coefficients $\beta = (\beta_1, \ldots, \beta_p)'$ via a link function $g(\mu_i) = \eta_i = x_i^T\beta$. To account for correlation among the outcomes y_i, a random effect is included in the linear predictor, such that

$$g(\mu_i) = \mu + x_i^T\beta + U_i \tag{8}$$

where μ is an overall intercept and the U_i's follow a CAR prior specification. Under the specification in Equation (8), the y_i's are assumed to be *conditionally independent* given the U_i's. Clayton and Kaldor[23] propose an **empirical bayes** approach assuming a PCAR prior when modeling the number of cases of lip cancer in Scotland. It is known that CAR prior specifications provide random effects that are too smooth. As mentioned by Leroux *et al.*[24] the parameter τ^2 in the ICAR specification serves to represent both **overdispersion** and spatial dependence in the case of Poisson regression models. To overcome this issue, Besag *et al.*[6] propose a *convolution* prior model (also known as *BYM*) wherein an independent (across areas) random effect is added to Equation (8), such that

$$g(\mu_i) = \mu + x_i^T\beta + U_i + V_i \tag{9}$$

where U_i and V_i are assumed to be independent, *a priori*, V_i follows an independent, zero mean normal distribution with an unknown variance σ^2, and U_i follows an ICAR prior specification. Note that if $\tau^2 \to 0$ then the U_i's are constants, whereas large values of τ^2 imply large but spatially structured variation. On the other hand, if $\sigma^2 \to 0$ then $v_i = 0$, whereas large σ^2 implies large unstructured variability[6]. More importantly, the **likelihood** involves the sum $U_i + V_i$, so it is challenging to identify each of these components from a single realization of the process under study.

3.1 Reparametrizations of the BYM model

Because of the identifiability issue with the BYM model, some authors have proposed different reparametrizations of the BYM model. Leroux *et al.*[24] remove V_i from Equation (9) and include a component λ in the CAR prior specification, such that

$$E(U_i \mid u_{-i}) = \left(\frac{\lambda}{1-\lambda q_{ii}}\right)\sum_{j\in\partial_i} u_j, \quad \text{and} \quad V(U_i \mid u_{-i}) = \frac{\tau^2}{1-\lambda+\lambda q_{ii}} \tag{10}$$

where $0 \le \lambda \le 1$ denotes a spatial dependence parameter, such that $\lambda = 0$ defines a nonspatial model, and the level of spatial dependence increases with λ. In this case, the joint prior distribution of the vector U is a zero mean normal distribution, with covariance matrix $\Sigma_u = \tau^2[\lambda Q + (1-\lambda)I_n]^{-1}$, where the diagonal elements of Q, q_{ii}, contain the number of neighbors of region i, and $q_{ij} = -1$ if $i \sim j$, and 0 otherwise. When $\lambda = 1$ the distribution of U is improper, as we have the ICAR specification. Lee[25] performs a simulation study comparing the mostly used CAR prior specifications in **disease mapping**, and conclude that the model proposed by Leroux *et al.*[24] is the best among the fitted ones. Lee[25] claims that this is because, different from the ICAR and the BYM prior specifications, the prior proposed by Leroux *et al.*[24] can represent a range of strong and weak spatial correlation structures with a single set of random effects.

More recently, Simpson *et al.*[26] claim that the components U and V in equation (9) should not be assumed to be independent, suggesting that the priors on τ^2 and σ^2 should be dependent. Simpson *et al.*[26] propose a reparametrization of the model in Equation (9), wherein $g(\mu_i) = \mu + x_i^T\beta + \frac{1}{\tau}(\sqrt{1-\phi}V_i + \sqrt{\phi}U_i^*)$, with $0 \le \phi \le 1$ being a mixing parameter. The component U^* is a scaled spatially structured component where the generalized variance, computed as the geometric mean of the marginal variances, is equal to one. In this case, $1/\tau$ represents the marginal precision contribution from U^* and V, with ϕ representing the fraction of this variance explained by U^*, and $(1-\phi)$ the

DOI: 10.1002/9781118445112.stat08048

fraction explained by V. Riebler *et al.*[27] perform a simulation study comparing the performance of the parametrization proposed by Simpson *et al.*[26] in contrast with the BYM, and the models proposed by Leroux *et al.*[24] and Dean *et al.*[28], who proposed another reparametrization of the BYM model. Their results show that in terms of model comparison criteria, their proposal performs at least equally well as the compared parametrizations. But they claim that their proposal is the only one that allows for interpretation of the parameters and of the hyperpriors[27].

4 Inference Procedure and Software

The normalizing constant of the likelihood function based on CAR models involves unknown parameters. As pointed out by Cressie[3], Besag[2] coined the term **pseudolikelihood** for the function $p(\theta) = \prod_{i=1}^{n} p(u_i \mid u_j, j \neq i; \theta)$, where θ contains the hyperparameters of the model. See Chapter 7 of Cressie[3] for more details on estimation based on pseudolikelihood for CAR models in the exponential family.

Here we focus on the inference procedure for models that follow Equation (9). Since the landmark paper by Besag *et al.*[6], most of the papers that use CAR structures as latent structures to account for spatial correlation in the data follow the Bayesian paradigm to perform inference about the parameters in the model. Considering that the y_i's are conditionally independent realizations from a distribution belonging to the exponential family, whose mean is parametrized as in Equation (9), the parameter vector comprises $\theta = (\mu, \beta, u, v, \tau^2, \sigma^2)$. Under the Bayesian paradigm, the model specification is complete after assigning the prior distribution of θ. Commonly, it is assumed a zero mean normal prior distribution, with some large (fixed) variance, for μ and the components of β; this is to represent prior ignorance about the fixed effects. We assume initially that U follows an ICAR prior specification, and the V_i's follow independent, zero mean normal prior distributions with variance σ^2. The hyperparameters τ^2 and σ^2 are assumed to be independent, *a priori*. An important issue is how to fix the value of g in Equation (5). As this is an improper distribution, there is no correct answer[11]. If all areas are connected, the ICAR distribution is proper in the $n-1$ dimensional space, so one suggestion is to fix $g = (n-1)/2$[29]. Hodges *et al.*[30] discuss the value of g and show that in the presence of islands (disconnected groups of areas) g should be fixed at $(n-I)/2$, where I is the number of islands. See Ref. 7 for a more detailed discussion about this. Assigning a prior distribution for τ^2 is challenging as it is involved in the variance of the *conditional* distribution $U_i \mid u_{-i}$. On the other hand, σ^2 represents the marginal variance of V_i. Therefore, these variances are not comparable. One natural approach is to assign independent Gamma distributions for the precisions, τ^{-2} and σ^{-2}. However, priors of the type $Ga(\epsilon, \epsilon)$ with $\epsilon \to 0$ should be avoided as in this case the probability mass is away from zero[31,32]. In the disease mapping context, Bernardinelli *et al.*[33] propose the use of a prior that is based on an approximation of the distribution of the log relative risks to a normal. More specifically, they assign the prior distribution for τ^2 based on a k-fold variation between the upper (95%) and lower (5%) relative risks. Wakefield[32] proposes an alternative approximation, based on the distribution of the differences $Z_i = U_i - U_n$, $i = 1, 2, \ldots, n-1$.

Once the prior distributions are specified, following the Bayes' paradigm, the posterior distribution is proportional to the likelihood function times the prior distribution. Even if the y_i's follow conditionally independent normal distributions, and we assign prior distributions according to the earlier discussion, the resultant posterior distribution does not have a closed form. There are two approaches that have been widely used to obtain samples from the resultant posterior distribution: **Markov chain Monte Carlo** (MCMC) methods (see Ref. 34 for a review) and the integrated nested Laplace approximations (INLA)[35]. Different from MCMC, INLA provides samples of the marginal posterior distributions of the parameters. For more details and examples on the INLA approach, see for example, Ref. 36.

Wiley StatsRef: Statistics Reference Online, © 2014–2018 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08048

One important issue when implementing an MCMC procedure for a mean model that includes an ICAR component together with an overall intercept, as in Equation (9), is to make sure that the resultant posterior distribution is proper. As already discussed, the ICAR distribution is based on pairwise differences between the components of U; therefore, it is not possible to identify the intercept and the random effects. One possible remedy is to impose a linear constraint such that $\sum_i u_i = 0$. In an MCMC algorithm this is done by centering the random effects, that is by replacing u_i by $u_i - \overline{u}$, for all i at each MCMC iteration[11,37]. In the context of disease mapping applications, Knorr-Held and Rue[38] propose different block sampling algorithms to improve the performance of the MCMC.

There are different software packages in `R`[39] that can be used to fit models with CAR components. Four examples that make use of MCMC are `WinBugs`[40], `CARBayes`[41], `NIMBLE`[42], and `STAN`[43]. On the other hand, the software `R-INLA`[35] performs the inference based on the INLA approach. Note that different packages have different parametrizations available. See the respective manuals for further details.

4.1 An Example: Analysis of the Number of Cases of Dengue Fever in Rio de Janeiro

We have available data on the number of cases of dengue fever across the districts of the city of Rio de Janeiro, Brazil, in April 2016. Dengue fever is transmitted to humans through some species of mosquitoes of the genus *Aedes*, being *Aedes aegypti* the most common one. The data are publicly available from the website of the Health Secretary of the city of Rio de Janeiro (SMS-RJ). We fit different specifications of the model described in Equation (9). We assume that $(Y_i \mid e_i, \lambda_i) \sim Poi(e_i\lambda_i)$, where e_i is the expected number of cases in district i, with $e_i = \frac{\sum_i y_i}{\sum_i pop_i} \times pop_i$, and where pop_i is the population size of district i. Rio de Janeiro has $n = 156$ districts. Following the specification in Equation (9), we fit the following models:

- Independent (IND): $\log \lambda_i = \mu + \beta\ HDI_i + V_i$ with $V_i \sim N(0, \sigma^2)$;
- ICAR: $\log \lambda_i = \mu + \beta\ HDI_i + U_i$ with $U_i \sim ICAR(\tau^2)$;
- BYM: $\log \lambda_i = \mu + \beta\ HDI_i + U_i + V_i$, with $U_i \perp V_i$, $U_i \sim ICAR(\tau^2)$, and $V_i \sim N(0, \sigma^2)$;
- Leroux *et al*. (LER): $\log \lambda_i = \mu + \beta\ HDI_i + U_i$ with U_i following the prior specification in Equation (10);
- Simpson *et al*. (BYM2): $\log \lambda_i = \mu + \beta\ HDI_i + \frac{1}{\tau}(\sqrt{1-\phi}V_i + \sqrt{\phi}U_i^*)$.

In all models above, HDI_i represents the human development index (HDI) of district i. Whenever a CAR specification is included in the fitted model, we assume a 0−1 neighborhood structure. For the fixed effects, we assume independent zero mean normal prior distributions with precision 10^{-3}. Except for the BYM2 model, we assigned gamma prior distributions to the precision parameters of the random effects (models IND, ICAR, BYM, and LER) with parameters 1 and 0.001. For the parameter λ under model LER, we assigned a uniform prior distribution on the interval (0,1). For model BYM2, the prior specification for ϕ and τ followed the prior complexity priors proposed by Simpson *et al.*[26], and also followed by Riebler *et al.*[27]. As `R-INLA` is the only package that has implemented the reparametrization of the BYM model proposed by Simpson *et al.*[26], the results shown here are based on the `R-INLA` outputs. Table 1 presents the values of DIC[44] under each fitted model. The results are very similar, with BYM2 performing slightly better. The estimated effect of HDI on the log-relative risk is estimated at −0.25 with 95% posterior credible interval equaling (−0.45 ; −0.06) under model BYM2. Figure 1 shows the posterior mean of the relative risks. We note some trend of the relative risk in the east−west direction, and a small district in the east side of the city that has a relative risk much greater than the others. Under model BYM2, the precision parameter is estimated at 0.95 (0.68; 1.27), and the parameter ϕ is estimated at 0.61 (0.34; 0.85), suggesting that for this data set the spatial component is stronger than the independent one.

DOI: 10.1002/9781118445112.stat08048

Table 1. DIC values under each of the fitted models for the dengue fever data set.

Model	IND	ICAR	BYM	LER	BYM2
DIC	1056.85	1049.83	1049.11	1050.30	1048.87

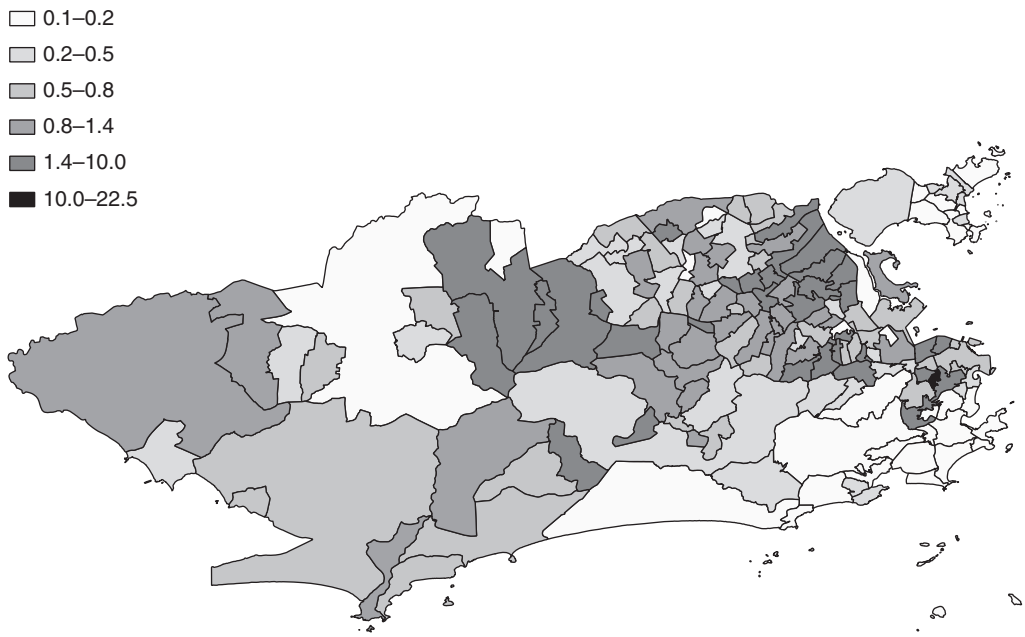

Figure 1. Posterior mean of the relative risk of dengue fever in Rio de Janeiro in April 2016.

5 Further Topics

Some recent developments related to CAR models, and more generally to GMRF, show an explicit connection between Gaussian processes (GP) and GMRF. Different from GMRF, GPs are continuously indexed and are commonly used to describe processes that vary continuously over a region. Lindgren *et al.*[45] used a formulation based on stochastic partial differential equations (SPDEs) that provide an explicit link between GPs and GMRF. In particular, they show that the ICAR model is an approximation to a GP with a Matérn covariance function. This is important because the inference procedure under CAR models gains in speed as the precision matrix is usually sparse. See Refs 45–47 for more details. Some further topics that involve CAR models and have also been explored in the literature are as follows.

5.1 Spatial confounding

The presence of a CAR latent effect in Equation (8) might affect the estimation of the fixed effects, especially if the vector x contains covariates that vary spatially across the region. This is known in the literature as spatial confounding and has been discussed by many authors. Clayton *et al.*[48] were probably the first

Wiley StatsRef: Statistics Reference Online, © 2014–2018 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08048

to point to the fact that the effect of a covariate changes in the presence of a CAR random effect. More recently, Wakefield[32] and Reich *et al.*[49] noted similar behavior of the fixed effects in the presence of correlated latent effects. A remedy proposed by Hodges and Reich[50] is to assume the spatial latent effects to be orthogonal to the fixed effects. They call this restricted spatial regression. An alternative approach has been proposed by Hughes and Haran[51] and is implemented in the `R` package `ngspatial`[52]. For further discussion about spatial confounding see for example, Refs 53–55.

5.2 Spatiotemporal models

Commonly, observations available in different areas of a region of interest are also available across time. For example, in the disease mapping context, the interest might be on estimating the temporal trend of a disease in different areas. The challenge in this case is to capture the underlying spatiotemporal correlation structure present in the data. The CAR structure can also be used in this context. See for example, Refs 56–58. Other examples, and further discussion about spatio-temporal models using the CAR distribution, can be seen, for example, in Sections 11.7 and 11.8 of Banerjee *et al.*[11], Chapter 11 of Lawson[59], and Chapter 7 of Blangiardo and Cameletti[36].

5.3 Multivariate CAR models

CAR models have also been extended to the case wherein a p-dimensional vector of observations is available at each region i, such that $y_i = (y_{i1}, \ldots, y_{ip})T$. For example, one might be interested in modeling different diseases across a region of interest. Mardia[60] extends the univariate CAR model to the multivariate framework. See Chapter 10 of Banerjee *et al.*[11], and references therein, for a review and extensions of the model proposed by Mardia[60].

Acknowledgments

The authors thank Professor Walter W. Piegorsch for the invitation to write this review. A. M. Schmidt was partially supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. W. S. Nobre is a PhD student in the Graduate Program in Statistics of the Federal University of Rio de Janeiro, Brazil, and thanks *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), for the financial support of his studies.

End Notes

1. *http://www.rio.rj.gov.br/web/sms/dengue-casos-bairro-periodo*
2. Available from *https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_por_IDH*

References

[1] Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems. *J. R. Stat. Soc., Ser. B (Methodological)* **36** (2), 192–236.

DOI: 10.1002/9781118445112.stat08048

[2] Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician* **24** (3), 179–195.
[3] Cressie, N. (1993) *Statistics for Spatial Data, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*, John Wiley & Sons, Inc., New York.
[4] Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*, CRC Press, Boca Raton, FL.
[5] Brook, D. (1964) On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika* **51**, 481–483.
[6] Besag, J., York, J., and Mollié, A. (1991) Bayesian image restoration, with two applications on spatial statistics. *Ann. Inst. Stat. Math.* **43**, 1–59.
[7] Lavine, M.L. and Hodges, J.S. (2012) On rigorous specification of ICAR models. *Am. Stat.* **66**, 42–49.
[8] Ferreira, G.S. and Schmidt, A.M. (2006) Spatial modelling of the relative risk of dengue fever in Rio de Janeiro for the epidemic period between 2001 and 2002. *Braz. J. Probab. Stat.* **20**, 47–63.
[9] Earnest, A., Morgan, G., Mengersen, K., Ryan, L., Summerhayes, R., and Beard, J. (2007) Evaluating the effect of neighbourhood weight matrices on smoothing properties of conditional autoregressive (CAR) models. *Int. J. Health Geograph.* **6**, 54.
[10] Freni-Sterrantino, A., Ventrucci, M., and Rue, H. (2017) A note on intrinsic conditional autoregressive models for disconnected graphs. arXiv preprint arXiv:1705.04854.
[11] Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2014) *Hierarchical Modeling and Analysis for Spatial Data*, CRC Press/Chapman Hall, Boca Raton, FL.
[12] Besag, J. and Kooperberg, C. (1995) On conditional and intrinsic autoregressions. *Biometrika* **82**, 733–746.
[13] Wall, M.M. (2004) A close look at the spatial structure implied by the CAR and SAR models. *J. Stat. Plann. Inference* **121**, 311–324.
[14] Assunção, R. and Krainski, E. (2009) Neighborhood dependence in Bayesian spatial models. *Biom. J.* **51**, 851–869.
[15] White, G. and Ghosh, S.K. (2009) A stochastic neighborhood conditional autoregressive model for spatial data. *Comput. Stat. Data Anal.* **53**, 3033–3046.
[16] Rodrigues, E.C. and Assunção, R. (2012) Bayesian spatial models with a mixture neighborhood structure. *J. Multivariate Anal.* **109**, 88–102.
[17] Lawson, A., Rotejanaprasert, C., Moraga, P., and Choi, J. (2015) A shared neighbor conditional autoregressive model for small area spatial data. *Environmetrics* **26**, 383–392.
[18] Held, L. and Rue, H. (2010) Conditional and intrinsic autoregressions, in *Handbook of Spatial Statistics* (eds A.E. Gelfand, P.J. Diggle, M. Fuentes, and P. Guttorp), Chapman & Hall, CRC Press, Boca Raton, FL, pp. 201–216.
[19] Heikkinen, J. and Hogmander, H. (1994) Fully Bayesian approach to image restoration with an application in biogeography. *J. R. Stat. Soc., Ser. C (Appl. Stat.)* **43**, 569–582.
[20] Hogmander, H. and Møller, J. (1995) Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics* **51**, 393–404.
[21] Hoeting, J.A., Leecaster, M., and Bowden, D. (2000) An improved model for spatially correlated binary responses. *J. Agric. Biol. Environ. Stat.* **5**, 102–114.
[22] Caragea, P.C. and Kaiser, M.S. (2009) Autologistic models with interpretable parameters. *J. Agric. Biol. Environ. Stat.* **14**, 281.
[23] Clayton, D. and Kaldor, J. (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–681.
[24] Leroux, B.G., Lei, X., and Breslow, N. (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence, in *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (ed. M.E. Halloran and Berry, D.), Springer, New York, pp. 135–178.
[25] Lee, D. (2011) A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat. Spatio-Temporal Epidemiol.* **2**, 79–89.
[26] Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H. *et al.* (2017) Penalising model component complexity: a principled, practical approach to constructing priors. *Stat. Sci.* **32**, 1–28.
[27] Riebler, A., Sørbye, S.H., Simpson, D., and Rue, H. (2016) An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* **25**, 1145–1165.
[28] Dean, C., Ugarte, M., and Militino, A. (2001) Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics* **57**, 197–202.
[29] Knorr-Held, L. (2003) Some remarks on Gaussian Markov random field models for disease mapping, in *Highly Structured Stochastic Systems* (eds P. Green, N. Hjort, and S. Richardson), Oxford University Press, Oxford, pp. 260–263.
[30] Hodges, J.S., Carlin, B.P., and Fan, Q. (2003) On the precision of the conditionally autoregressive prior in spatial models. *Biometrics* **59**, 317–322.
[31] Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1**, 515–534.

Wiley StatsRef: Statistics Reference Online, © 2014–2018 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08048

[32] Wakefield, J. (2007) Disease mapping and spatial regression with count data. *Biostatistics* **8**, 158–183.

[33] Bernardinelli, L., Clayton, D., and Montomoli, C. (1995) Bayesian estimates of disease maps: how important are priors? *Stat. Med.* **14**, 2411–2431.

[34] Gamerman, D. and Lopes, H.F. (2006) *Markov Chain Monte Carlo – Stochastic Simulation for Bayesian Inference*, 2nd edn, Chapman & Hall/CRC, Boca Raton, FL.

[35] Rue, H., Martino, S., and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc., Ser. B (Stat. Methodol.)* **71**, 319–392.

[36] Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-Temporal Bayesian Models with R-INLA*, John Wiley & Sons, Ltd, Chichester, West Sussex.

[37] Gelfand, A.E. and Sahu, S.K. (1999) Identifiability, improper priors, and Gibbs sampling for generalized linear models. *J. Am. Stat. Assoc.* **94**, 247–253.

[38] Knorr-Held, L. and Rue, H. (2002) On block updating in Markov random field models for disease mapping. *Scand. J. Stat.* **29**, 597–614.

[39] R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/ (accessed 07 November 2017).

[40] Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat. Comput.* **10**, 325–337.

[41] Lee, D. (2013) CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors. *J. Stat. Software* **55**, 1–24.

[42] NIMBLE Development Team (2017) NIMBLE: An R Package for Programming with BUGS Models, Version 0.6-6, http://r-nimble.org (accessed 07 November 2017).

[43] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A. (2016) Stan: a probabilistic programming language. *J. Stat. Software* **20**, 1–37.

[44] Spiegelhalter, D., Best, N., Carlin, B., and Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc., Ser. B* **64**, 583–639.

[45] Lindgren, F., Rue, H., and Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc., Ser. B (Stat. Methodol.)* **73**, 423–498.

[46] Simpson, D., Lindgren, F., and Rue, H. (2012) Think continuous: Markovian Gaussian models in spatial statistics. *Spat. Stat.* **1**, 16–29.

[47] Simpson, D., Lindgren, F., and Rue, H. (2012) In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* **23**, 65–74.

[48] Clayton, D.G., Bernardinelli, L., and Montomoli, C. (1993) Spatial correlation in ecological analysis. *Int. J. Epidemiol.* **22**, 1193–1202.

[49] 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**, 1197–1206.

[50] Hodges, J.S. and Reich, B.J. (2010) Adding spatially-correlated errors can mess up the fixed effect you love. *Am. Stat.* **64**, 325–334.

[51] Hughes, J. and Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Stat. Soc., Ser. B (Stat. Methodol.)* **75**, 139–159.

[52] Hughes, J. and Cui, X. (2017) ngspatial: Fitting the centered autologistic and sparse spatial generalized linear mixed models for areal data, Denver, CO. R package version 1.2.

[53] Paciorek, C.J. (2010) The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Stat. Sci. Rev. J. Inst. Math. Stat.* **25**, 107–125.

[54] Hanks, E.M., Schliep, E.M., Hooten, M.B., and Hoeting, J.A. (2015) Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics* **26**, 243–254.

[55] Page, G.L., Liu, Y., He, Z., and Sun, D. (2017) Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scand. J. Stat.* **44**, 780–797.

[56] Schmid, V. and Held, L. (2004) Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* **60**, 1034–1042.

[57] Nobre, A.A., Schmidt, A.M., and Lopes, H.F. (2005) Spatio-temporal models for mapping the incidence of malaria in Pará. *Environmetrics* **16**, 291–304.

[58] Fernandes, M.V.M., Schmidt, A.M., and Migon, H.S. (2009) Modelling zero-inflated spatio-temporal processes. *Stat. Modell.* **9**, 3–25.

[59] Lawson, A.B. (2013) *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, CRC Press, A Taylor & Francis Group, New York.

[60] Mardia, K.V. (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivariate Anal.* **24**, 265–284.

DOI: 10.1002/9781118445112.stat08048