A note on intrinsic Conditional Autoregressive models for disconnected graphs

Anna Freni-Sterrantino∗
Massimo Ventrucci†
Håvard Rue‡

May 7th 2018

Abstract

In this note we discuss (Gaussian) intrinsic conditional autoregressive (CAR) models for disconnected graphs, with the aim of providing practical guidelines for how these models should be defined, scaled and implemented. We show how these suggestions can be implemente in two examples, on disease mapping.

Keywords: CAR models, Disease mapping, Disconnected graph, Gaussian Markov Random Fields, Islands, INLA

1 Introduction

Conditional Autoregressive (CAR) models are widely used to represent local dependency between random variables, with numerous applications in spatial statistics, disease mapping [1, 2] and imaging [3]. This paper discusses CAR models for disconnected graphs and provides specific recommendations in order to implement them in practice. Through this note, we will illustrates the proposed methods using a classic example on disease mapping, leaving the straightforward generalisation to the reader.

Disease mapping concerns the study of disease risk over a map of geographical regions. Let assume a study area partitioned in $n$ non overlapping regions, indexed by $i = 1, ..., n$, with $y_i$ the number of cases observed for a given disease in region $i$. If the disease is rare, $y_i$ can be assumed as Poisson distributed with mean $\theta_i = E_i r_i$, where $E_i$ is the expected number of cases (computed on the basis of the demographic characteristics of a reference population) and $r_i$ is the relative risk associated with living in region $i$; $r_i = 1$ means no augmented risk with respect to the average in the whole study area, $r_i > 1$ ($r_i < 1$) indicates higher (lower) risk than average.

As an example, in ecological regression studies investigating the relationship between pollution, say $z$, and health, the risks might be modelled as

$$\log(r_i) = \alpha + \beta z_i + x_i$$

where $\alpha$ is the log baseline risk and $\beta$ is the effect of pollution. The term $x_i$ is a random effect capturing residual extra variability in $i$, possibly due to unobserved risk factors. Some of these unobserved may vary

∗Corresponding author, Small Area Health Statistics Unit, Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom
Email:a.freni-sterrantino@imperial.ac.uk
†Department of Statistics, University of Bologna, Bologna, Italy
‡CEMSE Division, King Abdullah University of Science and Technology, Saudi Arabia
smoothly over space, inducing spatial structure in the residuals \(x_i\)'s. This structure can be modelled using a CAR prior.

The definition of a CAR model starts by specifying a graph. The graph consists of a collection of nodes and edges representing, respectively, regions and neighbouring relationships between them. The number of nodes determines the size of the graph. For the sake of a general definition, in this paper a graph is seen as a collection of nodes belonging to one or more connected components.

Within a connected component, each node is connected to at least another node through a path (i.e. a set of contiguous edges). Therefore, we say a graph is connected if it consists of one connected component of size larger than one (e.g. a 'mainland' component). A graph is disconnected if it is not connected, meaning that it consists of more than one connected component of any size (e.g. a 'mainland' component of size \(n\) plus an 'island' component of size 1).

Within a connected graph, specification of neighbouring relationships is clear, as each node has at least one neighbour. In this case, the definition of a CAR model follows straightforwardly from specification of the full conditionals \[3\]

\[
x_i | x_{-i}, \kappa \sim \mathcal{N} \left( \sum_{j \sim i} x_j / n_i, \left( n_i \kappa \right)^{-1} \right), \quad i = 1, \ldots, n.
\]

where \(x_{-i} = \{x_k, k \neq i\}, i \sim j\) means \(i\) and \(j\) are neighbouring nodes and \(n_i\) is the number of neighbours of \(i\). The precision parameter \(\kappa\) regulates the degree to which \(x_i\) is shrunk to the local mean \(\sum_{j \sim i} x_j / n_i\).

Note that the variance \(\text{Var}(x_i | x_{-i})\) is inversely proportional to the number of neighbours \(n_i\). Therefore, both the prior assigned to \(\kappa\) and the structure of the graph play a role in determining the shrinkage properties of the CAR prior.

The graph structure is essentially defined through conditional independence assumptions between the nodes. In particular, prior \[1\] specifies that the \(x_i\)'s are conditionally independent given the information in neighbourhood \(x_j, i \sim j\). These assumptions are reflected in the precision structure of the joint distribution for \(x = (x_1, \ldots, x_n)^T\), derived from the full conditionals in \[1\],

\[
\pi(x | k) = \left( \frac{\kappa}{2 \pi} \right)^{(n-1)/2} |R|^{-1/2} \exp \left( -\frac{\kappa}{2} \sum_{i \sim j} (x_i - x_j)^2 \right), \quad \text{(2)}
\]

where the summation is over the set of all pairs of neighbours, \(i \sim j\), and \(| \cdot |_s\) represents the generalised determinant, calculated as the product of the non zero eigenvalues. Model \[2\] is a multivariate Gaussian with zero mean and \(n \times n\) precision \(Q = \kappa R\), where \(R\) is a matrix representing the neighbourhood structure of the model:

\[
R_{ij} = \begin{cases} 
  n_i & i = j \\
  -1 & i \sim j \\
  0 & \text{otherwise,}
\end{cases}
\]

The conditional independence assumptions between the nodes can be checked immediately looking inside \(R\): if \(R_{ij} = 0\) then \(x_i\) and \(x_j\) are conditionally independent, given all the other variables \(\{x_k : k \neq i, k \neq j\}\) [4]. The structure matrix in \[3\] has rank equal to \(n - 1\), thus density \[2\] is improper. As a consequence of this, the overall mean is unspecified in equation \[2\] and can be identified only when adding a linear constraint, such as \(\sum_i x_i = 0\). For this reason, this type of prior is referred to as intrinsic CAR. The purpose of using an intrinsic CAR prior is to borrow strength of information between neighbours, in the disease mapping example this yields a smoothed map for the risk.

In some applications the graph might be disconnected. A typical example is a graph made of two connected components, one of size larger than 1 (e.g. mainland) and the other of size 1 (e.g. island; components of size 1 will be denoted as singletons hereafter). In this case, a direct application of the intrinsic CAR \[2\] implies that \(n_i = 0\), if \(i\) is a singletons. This yields a constant prior for \(x_i\), i.e. the singleton random effect.
is, at prior, allowed infinite variance. This poses a general issue about interpretation of the prior for \( \kappa \), that may be different in the various connected components of the graph. These issues will be discussed in details in Section 3, were we will introduce the general case of a disconnected graph, with several connected components of any size.

Motivated by the issues involved in a direct application of (2) to a disconnected graph, we propose a solution based on rescaling the precision matrix \( Q \) to have similar shrinkage properties, at prior, in each connected component. The rescaling procedure is drawn by ideas in [5]. With our new definition the effect of the graph on the shrinkage properties (defined by the prior for \( \kappa \)) is marginalized out. Therefore, \( \kappa \) has a clear interpretation as a smoothing parameter, regulating the degree to which \( x_i \) will shrink to a local mean, if \( i \) has neighbours, and to a global mean, if \( i \) has no neighbour. In literature there is a lack of attention [6] on the definition and/or proprieties of a CAR for disconnected graphs, the only reference on this topic is Hodges et al. [7] who discuss the form of the normalizing constant in (2). On the practical side, the GeoBUGS manual [8] offers some guidelines in the case of a graph containing singletons, with a default option which is to set \( x_i \) to zero, if \( i \) is a singleton. Note that this is equivalent to enforcing a sum-to-zero constraint \( (x_i = 0) \) on each singleton random effect.

An alternative approach is to correct the graph, i.e. to remove the singletons by connecting islands to mainland. In our opinion neither of these strategies address the issue in a satisfactory way: the first one adds new constraints, the second one essentially corrupts the graph. We would like to stress that the definition of the graph is part of the modelling process, therefore editing new edges between islands and mainland is only appropriate in cases where borrowing strength of information between them is needed. Changing the graph is inappropriate every time the application at hand requires the original graph, i.e. when borrowing strength of information between islands and mainland is not a sensible choice.

The rest of the paper is organized as follows. Section 2 introduces formally the concept of an intrinsic CAR model defined with respect to a graph. In Sections 3 we revise scaling of the precision matrix of an intrinsic CAR model defined with respect to a connected graph, following [5]. In Section 4 we discuss the issues caused by direct application of (2) in the case of a disconnected graph. We then outline recommendations on how the model should be scaled in this case. Section 5 deals with linear constraints and computation of the normalizing constant. In Section 6 we illustrate the proposed methods in two examples on disease mapping involving two different types of graphs. We conclude with a discussion in Section 7.

## 2 Intrinsic CAR model defined with respect to a graph

In its simple form, the density of an intrinsic CAR model for \( x \) is

\[
\pi(x \mid \kappa) \propto \frac{1}{Z_n(\kappa)} \exp \left( -\frac{\kappa}{2} \sum_{i \sim j} w_{ij}(x_i - x_j)^2 \right)
\]  

(4)

where \( Z_n \) is a normalizing constant that we will return to later on and \( w_{ij} \) is a positive and symmetric weight assigned to the pair of neighbouring nodes \( i \) and \( j \). The definition of a “neighbour” is application dependent and part of the model specification. For example, in many spatial applications, two regions (\( i \) and \( j \), say) are considered to be neighbours if they share a common border. The weight \( w_{ij} \) might be chosen as the inverse Euclidean distance between the centroids of regions \( i \) and \( j \). With no loss of generality, in this work we only discuss the case of unitary weights, \( w_{ij} = 1 \), if \( i \) and \( j \) share a common border, corresponding to equation (2).

The interpretation of distribution (4) is that similarity between two neighbours are encouraged, and this induces a smoothing effect between neighbours, and thereby between neighbours of neighbours, and so on. We can formalize this, by defining a undirected graph \( G = (V, E) \), with a set of vertices \( V = \{1, 2, \ldots, n\} \), and edges \( E \), the (unordered) set of neighbours. We then say that the intrinsic CAR is defined with respect to the graph \( G \).

The precision parameter \( \kappa \) determines the amount of smoothing and it is commonly estimated from data. The density (4) is improper or *intrinsic*, in the sense that it is invariant to adding the same constant to all the
$x_i$’s, as an example of a first order polynomial intrinsic CAR. Higher order polynomial (and non-polynomial) intrinsic CAR’s can be defined similarly; see [4, Ch. 3] for a thorough discussion.

We will focus on discussing the simplest cases of intrinsic CAR models for graphs, displayed in equation 4, that are easily generalized to other types of intrinsic CAR models. This approach avoids technicalities which may distract from understanding the underlying ideas.

### 3 Scaling of an intrinsic CAR defined for a connected graph

Intrinsic CAR models have an unresolved issue with scaling, which is not addressed by studying (4). The basic reference is Sørbye et al.[5], on which we will base our arguments in this section.

Let us start by assuming that the graph $\mathcal{G}$ is connected, meaning that it is made of one connected component. An example of a connected graph is shown in Figure 1a. The intrinsic CAR model (4) defined for this graph, has precision matrix (the zeros are not shown):

$$Q = \kappa \begin{bmatrix}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 4 & \end{bmatrix}.$$

The matrix is singular with rank deficiency of 1. The intrinsic density is invariant to adding a constant to $x$, meaning that $x$ and $x + c1$ has the same improper density. However, what is of practical interest here, is how and by how much this model varies around its mean value $\mathbf{E}$, i.e. how $x$ varies if we impose a sum-to-zero constraint $\mathbf{1}^T x = 0$. The critical issue is that this is a complicated function of the graph $\mathcal{G}$, for which we have no good intuition.

![Figure 1: The two graphs used in the discussion: (a) a connected graph, and (b) a disconnected graph.](image)
Let us formalize this by introducing the concept of typical marginal variance. Let \( i \) be a node of a connected graph of size \( n \), we denote as \( s_i^{(k)} \) the (conditional on \( \kappa \)) marginal variance of \( x_i \), which is the \( i^{th} \) entry in the diagonal of the generalised inverse of \( Q \). The typical marginal variance of an intrinsic CAR defined with respect to a connected graph is defined as the geometric mean of the marginal variances,

\[
S^{(\kappa)} = \exp \left( \frac{1}{n} \sum_{i=1}^{n} \log(s_i^{(\kappa)}) \right)
\]

We assume \( S^{(\kappa)} \) as a measure of the deviation of \( \mathbf{x} \) from its null space, i.e. after applying the constraint \( \mathbf{1}^T \mathbf{x} = 0 \). Note that this quantity depends on both \( \kappa \) and the graph, through the structure \( R \).

For our example the marginal variances, \( s_i \)'s, are \( 0.53/\kappa, 0.53/\kappa, 0.19/\kappa, 0.53/\kappa, 0.44/\kappa \) and \( 0.44/\kappa \) for \( x_1, \ldots, x_6 \), which we interpret as follows.

- The issue - that marginal variances are different - is a feature of the intrinsic CAR model, consequently of that the conditional variance, \( \text{Var}(x_i|x_{-i}) \), is inversely proportional to the number of neighbours of node \( i \), see equation 1.

- The typical marginal variance is confounded with our interpretation of the precision parameter \( \kappa \), and we do not know a-priori what \( \kappa = 1 \) means in terms of a typical marginal variance. In the Bayesian framework, this is a crucial issue, since we need to impose a prior distribution for \( \kappa \). We need to address what \( \kappa \) means in terms of its impact on the model.

The solution proposed in Sørbye and Rue[5] of this apparent dilemma, is simply to scale \( Q \) so that the typical marginal variance is 1 when \( \kappa = 1 \). They also recommend multiplying \( Q \) by the typical marginal variance calculated for \( \kappa = 1 \), i.e. \( S^{(1)} = \exp\left( \frac{1}{n} \sum_{i=1}^{n} \log(s_i^{(1)}) \right) \). In our example, applying this scaling yields the following scaled precision matrix,

\[
Q_{\text{scaled}} = \kappa S^{(1)} \begin{bmatrix}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 4 & -1 & -1 \\
& 2 & -1 & -1 \\
& -1 & -1 & 2 \\
& -1 & -1 & 2
\end{bmatrix}
\]

where \( S^{(1)} = 0.4219 \). The most important consequence of this scaling, is that \( \kappa \) is now the typical precision and not only a precision parameter. This makes it possible to define a meaningful prior distribution and have clear interpretation for \( \kappa \). There is a long tradition to prefer model parameters with a good and clear interpretation.

Our recommendation is clear and unambiguous.

**Recommendation 1** We recommend scaling intrinsic CAR models defined with respect to connected graphs.

Computation of the scaling quantity \( S^{(1)} \) needs the generalised inverse of \( R \) (recall \( Q \) is equal to \( R \), if \( \kappa = 1 \)), which is an \( \mathcal{O}(n^3) \) operation. A better approach, is to make use of the graph of the model and the knowledge of the null-space, and treat \( R \) as a sparse matrix; see Rue and Held [4, Ch. 2.4] for background and details. The scaling \( S^{(1)} \) can then be computed from a rank one correction of the marginal variances from the unconstrained model, see Rue et al.[9] for technical details about the recursions leading to the marginal variances. The computational cost will be be \( \mathcal{O}(n^{3/2}) \) for typical spatial graphs, which is a huge improvement. The `R-function inla.scale.model()` in the R-INLA package (see www.r-inla.org) is an efficient implementation of this.
4 Scaling of an intrinsic CAR defined for a disconnected graph

A graph is disconnected if it is not connected, i.e. it is made of more than one connected component of any size. The practical interpretation of this related to (4), is that there are separated regions and/or nodes with no neighbours. Such cases easily appear if the graph is constructed from a map where we can have islands, or regions that are physically disconnected from the rest of the area. Figure 1b shows a disconnected graph with three connected components of size 3, 2 and 1 (i.e. a singleton). We will use this graph as reference, in this section.

A direct application of the intrinsic CAR model for graph in Figure 1b, gives the precision matrix

\[
Q = \kappa \begin{bmatrix}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 2 \\
\end{bmatrix}
\]

This matrix is singular with rank-deficiency of 3, since the density is invariant when we add a constant to each connected component. There are several unfortunate issues with this intrinsic CAR model, simply because the implicit assumption behind (4) is that the graph is connected. Similar to Section 3, we have the following issues due to the disconnected graph.

- Node 6 has no neighbours so \( Q_{6,6} = 0 \) and we have a constant density for \( x_6 \). This can lead to an improper posterior distribution. To see this, consider the following model for an observed count, \( y_6 \), in node 6,

\[
y_6 | x_6 \sim \text{Po}(\exp(x_6))
\]

\[
\pi(x_6 | \kappa) \propto \text{const}
\]

where \( x_6 = \log(\theta_6) \) is the Poisson mean, \( \theta_6 \), in logarithmic scale. The constant prior (8) implies an improper prior on the Poisson mean, i.e. \( \pi(\theta_6) \propto 1/\theta_6 \). If a zero count is the case, the posterior \( \pi(\theta_6 | y_6 = 0) \propto \exp(-\theta_6)/\theta_6 \) is an improper density. In other words, the constant prior for the singleton makes it difficult for the singleton random effect to shrink to the global mean. Also, this goes against the purpose of using (4) in the first place, which is to do smoothing and borrowing strength.

- The connected components for \( (x_1, x_2, x_3) \) and \( (x_4, x_5) \) are defined as on a connected graph. Even though this is reasonable within each connected component, it is not reasonable when we compare across them. The issue is that the typical marginal variance will in general be different between the connected component. Following Section 3, the marginal variances are \( 0.22/\kappa, 0.22/\kappa, 0.22/\kappa, 0.25/\kappa, 0.25/\kappa \) and \( \infty \) for \( x_1, \ldots, x_6 \). (Note that in this simple case the marginal variances are constant within each connected component.) The typical marginal variance for the first and second component is \( S_{1}^{(1)} = 0.22 \) and \( S_{2}^{(1)} = 0.25 \), while it is \( \infty \) for the singleton \( x_6 \).

To resolve both these issues we propose scaling the intrinsic CAR with respect to the disconnected graph, i.e. scaling (7) similarly to Section 3, but with two minor modifications.

1. We scale each connected component of size larger than one, independently as described in Section 3, (i.e. we apply recommendation 1 to each connected component of size larger than one).

2. For connected components of size one, we replace the constant prior with a standard Gaussian with precision \( \kappa \). This assumes the singleton as a non spatial random effect with precision \( \kappa \), as also proposed in [1].
In our example, application of the two instructions above produce the following scaled precision matrix:

\[
Q_{\text{scaled}} = \kappa \left( S_1^{(1)} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} + S_2^{(1)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & -1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)
\]

where \( S_1^{(1)} \) and \( S_2^{(1)} \) are the scaling values for, respectively, the first and second connected component of size larger than one. Note that, since a standard Gaussian prior with precision \( \kappa \) is assigned to the singleton, its marginal variance is the same as for two the connected components. This scaling gives a well defined interpretation of the precision parameter \( \kappa \). The prior for \( \kappa \) has (approximately) the same impact on the different connected components of the graph.

Our recommendation is clear and unambiguous.

**Recommendation 2** We recommend scaling intrinsic CAR models defined with respect to disconnected graphs.

The popular WinBUGS software treats singletons as non-stochastic nodes and sets the associated random effects to zero by default (see GeoBUGS manual[8] page 18). From the perspective of developing user friendly software to fit Bayesian models via Markov Chain Monte Carlo (MCMC) algorithms this seems a safe strategy: if singletons are taken as stochastic nodes, with consequent improper \( \pi(x_i) \propto \text{const} \), this may lead to poor mixing and extremely slow convergence especially when a zero count is observed.

5 Linear constraints and normalising constants

When using intrinsic models we always have to be careful not to introduce confounding [10]. Let \( x \) be the intrinsic CAR defined on a connected graph, then the linear predictor

\[
\eta = x
\]

and

\[
\eta = \mu 1 + x | (1^T x = 0)
\]

are the same. In the first case, there is no intercept as it is implicit in the null-space of the precision matrix for \( x \). In the second case, we explicitly define the intercept and remove it from the intrinsic CAR model. We strongly prefer the second option, since it makes the interpretation of each component explicit and reduces the chance of misunderstanding and misspecifying the intrinsic CAR in more complex scenarios than here.

When the graph is disconnected, we designate one intercept for each connected component with size larger than one. If a connected component specific intercept is needed, we prefer to add it to the model explicitly rather than implicitly.

**Recommendation 3** We recommend to use a sum-to-zero constraint for each connected component of size larger than one.

Rue and Held [4, Ch 3] provide a strong case, to interpret the normalizing constant for the proper part of the model and the improper part as a diffuse Gaussian. Assume the graph has \( n_c \) connected components, each of size \( n_i \). Then the normalizing constant for the scaled intrinsic CAR model (4), will be

\[
Z_n(\kappa) = |R|^{1/2} \prod_{i=1}^{n_c} Z_{n_i}(\kappa)
\]
6 Application

In this section we provide two applications, first we show how the scaling works with a disconnected graph, using lip cancer data from Scotland when we have three islands. In the second application, we show how to include specific component intercepts, using stomach cancer data from Tuscany (Italy) where we have two islands and two connected components.

6.1 Scottish Lip Cancer data: a graph with three singletons

The data are counts of lip cancer cases registered in $n = 56$ Scottish counties during years 1975-1980. Figure 2a shows the standardized mortality ratios (SMRs) over the regions. The graph is displayed in Figure 2b, it has been constructed assuming the counties are the nodes, with edges connecting counties sharing borders. Three counties are island (Orkneys, Shetland and the Outer Hebrides). We apply the model used in Breslow et al. [11]. Let assume vectors $y_i$ and $E_i$ are, respectively, observed and expected lip cancer cases during the study period in regions $i$, covariate $z$ is the “percentage of the population engaged in agriculture, fishing, or forestry” (AFF) and $r_i$ the unknown relative risks, the model is

$$y_i \sim \text{Po}(E_i r_i)$$

$$\log(r_i) = \alpha + \beta z_i/10 + x_i$$

$$\pi(x | \kappa) \propto \frac{1}{Z_n(\kappa)} \exp \left(-\frac{\kappa}{2} x^T R x \right).$$

Breslow et al. [11] analyzed the spatial dependency using an intrinsic CAR model defined on a connected graph, obtained by editing new edges to connect islands to mainland. Following the natural discontinuity
between islands and mainland, we illustrate our definition of an intrinsic CAR with respect to the disconnected graph in Figure 2b. This graph is made of four connected components, three of them are singletons.

Our aim is not to compare models built on different graphs, to evaluate which one fits best. In this first example, our aim is to compare results obtained under the scaled intrinsic CAR model and the unscaled one and emphasize the benefit of scaling. Recall the scaled model is obtained by scaling the precision $Q$ (recommendations 2) as described in detail in Section 4, while the unscaled model has precision $Q = \kappa R$, where $R$ is defined as in (3), i.e. $R[i,i] = 0$ if $i$ is an island. For the sake of comparison, for both scaled and unscaled models, we use the same prior $\pi(\kappa)$ (gamma with shape 1 and rate $5e-5$) and apply the sum-to-zero constraint $\mathbf{1}^T \mathbf{x} = 0$ to the nodes of the connected component.

Figure 3 displays the marginal posterior for six random effects fitted under the scaled (solid line) and unscaled (dashed line) model, using R-INLA. The scaled and unscaled models have different shrinkage properties, this is reflected in the large deviations in the top panels of Figure 3. Note that the island-specific random effects estimated by the unscaled model are less shrunk towards no effect, $x_i = 0$, than those estimated by the scaled model. On the other hand, the posterior for the three random effects belonging to the connected component of the graph (bottom panels of Figure 3) are essentially the same. Results for the other nodes in the connected component are similar and not shown here. The different shrinkage properties of the two models are confirmed by looking at the posterior summaries in Table 1. Though the fixed effects $\alpha$ and $\beta$ are almost unchanged, the relative risks for the singletons are more extreme under the unscaled model than the scaled one.

In the scaled model, the hyper-parameter $\kappa$ has a clear interpretation as a typical precision. The marginal variance within each of the four components of the Scotland graph (the mainland and the three singletons) is $\kappa^{-1}$. In the unscaled model, the marginal variances in the different components are different, in particular it is infinite in the three singletons. For this reason we believe the more extreme estimates obtained under the unscaled model are due to the prior, not a feature of the data (i.e. overfitting). In general, how severe the overfitting will depend on the structure of the graph, the sample size and the prior assigned to $\kappa$. For instance, if data contain little information about the disease risk (e.g. low expected counts, which is not the case with the lip cancer data), the prior $\pi(\kappa)$ may have a large impact on the analysis and larger deviations between scaled and unscaled must be expected.

In table 1, we reported also the Deviance Information Criterion (DIC), the number of effective parameters ($pD$) and the mean deviance $\bar{D}$. All are very close for the two scaled and unscaled models on the disconnected graphs, and they are not much different from those obtained for the scaled model on the connected graph: DIC was 297.1 ($pD = 28.7$ and $\bar{D} = 168$). To allow this comparison we kept the same hyperpriors for all the models fitted.

### 6.2 Tuscany Lung cancer mortality: a graph split in sub-graphs

In this second example, we used data for stomach cancer deaths (9947 cases) in men over 35 years old, collected for each municipality between 1971-79, from the Tuscan Cancer Atlas. Tuscan geography presents two small islands, that have separate municipalities, which we consider as singletons - Capraia and Giglio Isles - and one bigger island Elba composed of 8 municipalities and Tuscan mainland composed of 279 municipalities. Hence, the graphs is composed by 4 connected components, two of which are singletons Figure 4b. In Figure 4a, we plotted the standardised mortality ratio (SMR) and observed that the highest SMRs are concentrated in the North-East.

In table 2, we report statistical summaries for all the study regions, for Elba alone and for all mainland Tuscan municipalities except Elba. In Elba there are 54 cases, hence lower values for observed, expected and SMR values.

The first model we fit is a Besag York Mollié (BYM) [3], disease mapping model with scaled precision matrix on the disconnected graph, as follow:
Figure 3: The effect of scaling the disconnected graph. Upper panels show the marginal posterior (in the linear predictor scale) for the three singletons random effects, $x_6$ (Orkneys), $x_8$ (Shetland) and $x_{11}$ (Outer Hebrides); lower panels show the marginal posterior for three nodes in the connected component of the graph, $x_1$ (Skye-Lochalsh), $x_{45}$ (Edinburgh) and $x_{49}$ (Glasgow). The marginals from the unscaled model (dashed lines) are less shrunk towards $x_i = 0$ than the marginals from the scaled model (solid line).
Parameter | Mean | Standard deviation | 2.5% | Median | 97.5%
---|---|---|---|---|---
Scaled model
\(\kappa\) (Precision) | 3.97 | 1.17 | 2.16 | 3.81 | 6.69
\(\alpha\) (Intercept) | -0.25 | 0.13 | -0.50 | -0.25 | -0.00
\(\beta\) (AFF) | 0.37 | 0.13 | 0.09 | 0.37 | 0.62
\(r_6\) (Orkneys) | 2.87 | 0.90 | 1.42 | 2.77 | 4.93
\(r_8\) (Shetland) | 2.06 | 0.73 | 0.95 | 1.95 | 3.80
\(r_1\) (Outer Hebrides) | 2.32 | 0.63 | 1.28 | 2.26 | 3.75
\(D\text{IC}\) | 299.4
\(pD\) | 30.65
\(\bar{D}\) | 268.75

Unscaled model
\(\kappa\) (Precision) | 2.26 | 0.70 | 1.18 | 2.15 | 3.91
\(\alpha\) (Intercept) | -0.26 | 0.12 | -0.50 | -0.27 | -0.02
\(\beta\) (AFF) | 0.36 | 0.13 | 0.09 | 0.37 | 0.62
\(r_6\) (Orkneys) | 3.54 | 1.20 | 1.58 | 3.40 | 6.27
\(r_8\) (Shetland) | 3.26 | 1.18 | 1.36 | 3.11 | 5.96
\(r_1\) (Outer Hebrides) | 3.07 | 0.83 | 1.66 | 2.99 | 4.89
\(D\text{IC}\) | 298.5
\(pD\) | 30.5
\(\bar{D}\) | 267.98

Table 1: Posterior summaries for scaled (top) and unscaled (bottom) versions of model (9)-(11) and Deviance Information Criterion (DIC), number of effective parameters \(pD\) and mean deviance \(\bar{D}\).

| Observe | Expected | SMR |
|---|---|---|
| All(n=287) | 5th percentile | 2.3 | 1.93 | 0.69 |
| | Median | 18.0 | 9.30 | 1.68 |
| | 95th percentile | 107.0 | 65.31 | 3.42 |
| Elba(n=8) | 5th percentile | 1.35 | 2.87 | 0.34 |
| | Median | 4.90 | 4.90 | 0.82 |
| | 95th percentile | 18.80 | 14.81 | 1.82 |
| Without Elba (n=279) | 5th percentile | 3 | 1.90 | 0.72 |
| | Median | 18 | 9.50 | 1.71 |
| | 95th percentile | 108 | 66.46 | 3.42 |

Table 2: Statistical summaries: stomach cancer mortality for males over 35 years olds, in Tuscany (Italy). Observed and expected, standardised mortality ratio (SMR) by census municipality 1971-79.
Table 3: Posterior summaries for scaled one intercept (top) and scaled two intercepts, and Deviance Information Criterion (DIC), number of effective parameters \( pD \) and mean deviance \( \bar{D} \).

| Parameter | Mean  | Standard deviation | 2.5%  | Median | 97.5% |
|-----------|-------|--------------------|-------|--------|-------|
| \( \tau_v \) (Heterogeneity) | 10.350 | 2.150 | 6.399 | 10.296 | 14.767 |
| \( \kappa \) (Precision) | 122.574 | 224.899 | 15.536 | 61.987 | 606.157 |
| \( \alpha \) (Intercept) | 0.547 | 0.024 | 0.499 | 0.547 | 0.594 |
| \( DIC \) | 1818.21 |
| \( pD \) | 182.24 |
| \( \bar{D} \) | 1636.01 |

| Scaled Model-Two intercepts |
|-----------------------------|
| \( \tau_v \) (Heterogeneity) | 10.264 | 2.167 | 6.192 | 10.253 | 14.576 |
| \( \kappa \) (Precision) | 157.446 | 343.495 | 17.07491 | 71.17663 | 832.5374 |
| \( \alpha_1 \) (Intercept Elba) | -0.048 | 0.193 | -0.435 | -0.045 | 0.324 |
| \( \alpha_2 \) (Intercept Tuscany) | 0.558 | 0.025 | 0.510 | 0.558 | 0.606 |
| \( DIC \) | 1813.21 |
| \( pD \) | 180.96 |
| \( \bar{D} \) | 1632.22 |

**y_i \sim Po(E_ir_i)\)**

\[
\log(r_i) = \alpha + x_i + v_i
\]

Where \( x \) is modelled as a scaled intrinsic CAR with \( \kappa \) representing the precision as in equation (11) and \( v \) is the unstructured spatial residual \( v_i \sim N(0, \tau_v^{-1}) \), with \( \tau_v \) representing the precision of the unstructured random effects. Both precisions have as prior a gamma with shape 1 and rate \( 5e^{-5} \).

In the second model, we include two separate intercepts one for the island of Elba and one for the Tuscan mainland (including the two singletons), to distinguish the lower SMRs observed in Elba versus the higher SMRs the rest of the region.

\[
\log(r_i) = \alpha_j + x_i + v_i
\]

Where \( j = (1, 2) \) indicating the two main connected components of size \( > 1 \).

In table 3, we report posterior model estimate summaries, with DIC, number of effective parameters and mean deviance. The results show that we should prefer the second model - the two intercepts- as the DIC is lower, and the intercept for Elba is 0.95 (CI 95% 0.64,1.38) while for the rest of the region the baseline risk is 1.74 (CI 95% 1.66,1.83). The unstructured effects remain unchanged between the two models, while precision \( \kappa \), for the second model is higher, as assigning two intercepts help to explain better the spatial variability in the data. We compared these results with scaled and unscaled modes on a connected graph (scaled \( DIC = 1816.39 \) (\( pD = 180.7 \) and \( \bar{D} = 1635.6 \) and unscaled \( DIC = 1816.63 \) (\( pD = 181.0 \) and \( \bar{D} = 1635.6 \)). For this data, the model with a separated intercept for each connected component was the best according DIC.

7 Summary

We motivated the definition of intrinsic CAR models for disconnected graphs under two main recommendations: scaling the precision structure and applying sum-to-zero constraints on the connected components of
the graph. Scaling the precision structure sets the typical marginal variance to $\kappa^{-1}$ in each component of the graph, where $\kappa$ is the precision of the intrinsic CAR model. This immediately suggests a fair prior for random effects associated with the singletons in a disconnected graph in terms of a normal with zero mean and variance $\kappa^{-1}$. The advantage is that the prior assigned to $\kappa$ has the same interpretation regardless of the particular structure of the graph.

We applied this strategy to a disease mapping example on lip cancer in Scotland, using the natural disconnected graph with three singletons. In this example we observed for the scaled model compared to the unscaled one, the random effect for the three islands were closer to the average mean.

In general, the extent to which the unscaled intrinsic CAR leads to overfitting should depend on the on the prior assigned on $\kappa$. When data contain little information about the disease risk for people living in the area (e.g. small regions with low expected counts, which is not the case with the lip cancer data), the prior $\pi(\kappa)$ may have a large impact on the analysis. In these situations larger deviations between scaled and unscaled intrinsic CAR models must be expected, as the unscaled one has no control on the marginal variance, hence no control on the impact of $\pi(\kappa)$, whereas the scaled one provides good intuition of $\kappa$.

Finally, in the second example, we show, how to distinguish the effect of each connected component by introducing separate intercepts for graph of size $> 1$. The effect of jointly scaling the precision matrix and allowing specific separate baseline log risks for each component leaves the precision prior specification unchanged, as the scaling will guarantee that the connected components have the same precision assigned that regulates the degree to which the $x_i$ are shrunk toward the local mean. When there is a belief that the component of the graph should not be linked jointly, the introduction of multiple intercept offers an elegant solution. An immediate application of this approach could be beneficial in epidemiological studies where geographical discontinuity is an issue.
Acknowledgement

We thank Dr M. A. Vigotti (University of Pisa) for having made available the dataset from the Tuscany Atlas of Mortality 1971-1994. Massimo Ventrucci is supported by the Environmental processes and human activities: capturing their interactions via statistical methods (EPHASTAT) founded by the Italian Ministry for Education, University and Research. We can provide the code for the examples.

References

[1] J. Wakefield. Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–183, 2007.

[2] A. B. Lawson. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman & Hall/CRC Interdisciplinary Statistics. Chapmann & Hall/CRC, 2nd edition, 2013.

[3] J. Besag, J. York, and A. Mollié. Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43(1):1–59, 1991.

[4] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

[5] S. H. Sørbye and H. Rue. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8(3):39–51, 2014.

[6] L. Knorr-Held. Some remarks on gaussian markov random field models for disease mapping. In P. Green, N. Hjort, and S. Richardson, editors, *Highly structured stochastic systems*, pages 260–264. Oxford University Press, Oxford, 2002.

[7] J. S. Hodges, B. P. Carlin, and Q. Fan. On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, 59(2):317–322, 2003. ISSN 1541-0420. doi: 10.1111/1541-0420.00038. URL http://dx.doi.org/10.1111/1541-0420.00038.

[8] D. J. Spiegelhalter, A. Thomas, N. Best, and D. Lunn. *WinBUGS User Manual, Version 1.4*, 2002.

[9] H. Rue and S. Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*, 137(10):3177–3192, 2007. Special Issue: Bayesian Inference for Stochastic Processes.

[10] T. Goicoa, A. Adin, M. D. Ugarte, and J. S. Hodges. In spatio-temporal disease mapping models, identifiability constraints affect pql and inla results. *Stochastic Environmental Research and Risk Assessment*, 32(3):749–770, 3 2018. ISSN 1436-3240. doi: 10.1007/s00477-017-1405-0.

[11] D. G. Clayton N. E. Breslow. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993. ISSN 01621459. URL http://www.jstor.org/stable/2290687.