

Comparison of different software implementations for spatial disease mapping.

Abstract

Disease mapping is a scientific field that aims to understand and predict disease risk based on counts of observed cases within small regions of a study area of interest. Hierarchical model-based approaches that borrow information from neighbouring areas via conditional autoregressive (CAR) random effects on the local disease rates have gained a lot of popularity, thanks to the readily implemented Markov chain Monte Carlo methods. Nowadays, many software implementations to model risk distributions exist. Many of these applications differ, to varying degrees, in the underlying methodology. This paper provides an in-depth comparison between analysis results, coming from R-packages CARBayes, R2OpenBUGS, NIMBLE, R2BayesX, R-INLA, and RStan. We investigate CAR models typically used in disease mapping for spatially discrete count data. Data about diabetics in children and young adults in Belgium are used in a case study, while simulation studies are undertaken to assess software performance in different settings.

Keywords Disease mapping, conditional autoregressive models, software packages, relative risks, diabetics

1 Introduction

Disease mapping is a flourishing study in spatial epidemiology (Lawson 2018, Elliott et al. 2001). The main goal of disease mapping is to estimate the spatial pattern of disease risk across a set of non-overlapping areal units and to identify areas that have a high level of disease risk (Anderson et al. 2017). Many different tools for disease mapping are currently available. Bayesian hierarchical spatial models are commonly used for answering public health problems (Waller and Carlin 2010, Lawson 2018). In these models, the local disease risk is modelled using a set of spatial random effects, which borrow information over neighbouring regions to get more reliable region-specific estimates. Depending on the distributional assumption of these random effects, different models can be constructed.

The two most commonly used models in spatial epidemiology are the intrinsic conditional autoregressive (CAR) model and the convolution model. The intrinsic CAR model assumes a spatially autoregressive random effect to deal with the association among neighbouring regions (Besag et al. 1991). Besag et al. (1991) also introduced a spatial model in which the random effects are split into two components: a spatially structured CAR component and an unstructured, independent, random component. This model, called the convolution model or Besag-York-Mollie (BYM) model, allows for additional area-specific noise. In contrast to the intrinsic CAR model,

in which the spatial dependence is pre-determined by the neighbourhood structure, the convolution model can more flexibly capture the underlying spatial dependence via the relative contribution of the two random effects (Eberley and Carlin 2000). Several other models exist. For example, Leroux et al. (1999) proposed the use of one random effect, which includes a parameter to measure the level of spatial correlation among the areas. Diggle et al. (1998) assumed spatial random effects, which are correlated according to the distance between two regions.

A wide range of software implementations to model spatial risk distributions exist, and hierarchical model-based methods have gained a lot of popularity due to the availability of ready-to-apply software. We will focus on software tools that implement the intrinsic CAR and convolution model and which can be used within R software (R Core Team 2018). Many of these implementations, such as OpenBUGS (Lunn et al. 2009), the R packages CARBayes (Lee 2013) and NIMBLE (de Valpine et al. 2017) make use of a Markov chain Monte Carlo (MCMC) method for analyzing the CAR models. Other implementations like BayesX (Brezger et al. 2005, Umlauf et al. 2015) and Stan (Carpenter et al. 2017, Stan Development Team 2018a) estimate CAR models using either MCMC techniques or an approximation method. BayesX allows to analyze the model with a mixed model based methodology. Stan offers, as a fast alternative estimation algorithm, the use of variational inference. Another technique which has gained a lot of popularity is integrated nested Laplace approximations, implemented in the R package INLA (Rue et al. 2009, Lindgren and Rue 2015).

Several one-to-one comparisons of these software packages have been conducted in literature, some of which focused on models in the area of spatial epidemiology. Brezger et al. (2005) compared WinBUGS, a precursor of OpenBUGS, and BayesX in the setting of an additive regression model. They concluded that BayesX is computationally less expensive and has better mixing properties than WinBUGS for the resulting Markov chains. Carroll et al. (2015) performed a comparative study between INLA and OpenBUGS, with the focus on spatial Poisson data models. Computationally, INLA is much more efficient than OpenBUGS. However, major differences were observed with respect to the precision parameters for spatial random effects, when default values in INLA were used. On the other hand, Neyens et al. (2018) concluded that estimates of INLA in its default state and WinBUGS are very similar, based on a simulation study that focused on a negative binomial CAR model for count data. Bivand et al. (2017) recently made a comparison of several software implementations of the spatial intrinsic CAR and SAR (simultaneous autoregressive) model on two Gaussian examples in the context of spatial econometrics. They show that spatially structured data may indeed be estimated readily using a variety of approaches and conclude that robust results across methods of estimation are obtained from INLA and WinBUGS implementation of the intrinsic CAR model for Gaussian responses. To the best of our knowledge, no comparison study has been made with the comparison of all previously mentioned software packages, especially no general comparison study involving the software packages NIMBLE, CARBayes and the approximation method of Stan in the context of disease mapping. Hence, in this manuscript, a simulation study is conducted to compare several of the available software implementation in the context of spatial epidemiology, in order to check whether previous results can be generalized. Focus is on the similarity of the model parameters, the relative risks and computational time. Differences among the default assumptions and default output from the software packages will be highlighted.

Further, it will be looked if high-risk detection is performed alike. The focus is on count data, which are typical for disease mapping.

The remainder of this manuscript is structured as follows. Section 2 describes the used CAR models and software packages, with their characteristics. Section 3 illustrates and compares the different implementations for a case study. Section 4 provides an in-depth comparison of the implementations via a simulation study. Finally, section 5 gives conclusions and recommendations for further research.

2 Methodology

2.1 Models

In disease mapping, a geographic region of interest is partitioned into n small areas such as municipalities, districts, electoral wards, For each area k , we have observed the number of (newly diagnosed) disease or mortality cases, denoted by Y_k . In order to account for population differences (population size, age/gender distribution), it is common to compare the observed number of cases to the expected number of cases in that area, indicated by E_k . This is calculated using an age and/or gender standardization,

$$E_k = \sum_g p_g n_{gk},$$

where g corresponds to the age and/or gender strata, p_g is the rate of having the disease in a group g for a representative standard population (e.g. the country or the continent) and n_{gk} is the number of people in the age/gender category in area k . Commonly, this is modelled using a Poisson distribution,

$$Y_k | E_k, R_k \sim \text{Poisson}(E_k R_k), \quad k \in \{1, \dots, n\}, \quad (1)$$

with R_k the unknown relative risks of interest. The maximum likelihood estimate of R_k is given by the standardized incidence ratio (SIR) or standardized mortality ratio (SMR), which is the ratio of the observed to the expected number of cases in an area. If the SIR (SMR) is bigger than 1, the area has an elevated risk of disease (mortality) as compared to the rate at which individuals acquire the disease in the standard population (Lawson 2018).

However, the SIR (SMR) is unstable if the disease is rare and/or the population is small, and it does not correct for spatial structure of other underlying patterns in disease risk. Spatial models introduce spatial association to the disease risk, borrowing information from neighbouring areas, and are therefore more sensible solutions. The spatial models are characterized by the use of spatial random effects ϕ , while they can also contain a set of p explanatory variables $\mathbf{x}^T = (\mathbf{x}_1, \dots, \mathbf{x}_p)$. A general formulation of the spatial models is

$$\ln(R_k) = \mu + \mathbf{x}_k^T \beta + \phi_k, \quad (2)$$

where the regression parameters are noted by β and the intercept term by μ . Depending on the prior distribution of the random effects ϕ , different spatial models are constructed. The most commonly used spatial models are: the intrinsic CAR model and the convolution model. These are described in the next sections.

Note that CAR models make use of a so-called neighbourhood matrix. The neighbourhood structure can be defined in several ways. In this manuscript, areas are considered to be neighbours when they share at least one border point, a structure better known as the queen method. Two neighbouring areas j and k are denoted by $j \sim k$ and the number of neighbours of area k is denoted by n_k .

2.1.1 The intrinsic model

The intrinsic model was introduced by Besag et al. (1991). The full conditional distribution of the random effects ϕ is defined by

$$\phi_k | \phi_{-k}, \tau^2 \sim N\left(\frac{1}{n_k} \sum_{j \sim k} \phi_j, \frac{\tau^2}{n_k}\right).$$

The more neighbours an area has, the smaller the variance of the conditionally random effects. In case of strong spatial structure, it is a logical consequence of the fact that the more neighbours an area has, the more information there is. A disadvantage of the intrinsic model is that the joint distribution of ϕ is improper. This can however be solved by imposing the restriction $\sum_{k=1}^n \phi_k = 0$.

2.1.2 The convolution model

The convolution model or Besag-York-Mollie (BYM) CAR model was proposed by Besag et al. (1991). The random effects of this model can be formulated as follows

$$\begin{aligned} \phi_k &= \theta_k + \psi_k \\ \theta_k | \sigma^2 &\sim N(0, \sigma^2) \\ \psi_k | \psi_{-k}, \tau^2 &\sim N\left(\frac{1}{n_k} \sum_{j \sim k} \psi_j, \frac{\tau^2}{n_k}\right). \end{aligned}$$

Thus, the random effects are split up in a spatial component, represented by ψ and a random noise component, represented by θ . The advantage is that the relative contributions of the two components allow different strength of spatial association. A drawback of splitting up the random effects is that only the sum $\theta_k + \psi_k$ is identifiable.

2.2 Software packages

Several software packages implement the intrinsic CAR and convolution model. Most packages make use of the Bayesian inference framework using Markov chain Monte Carlo (MCMC) sampling. These methods provide samples from the joint posterior distribution of the unknown quantities. Posterior summary measures are calculated from these samples and reported. However, also alternative estimation methods are available. We will discuss and

Table 1: Software packages. An overview of the implementations of the CAR models.

Name	R Package	Version	Estimation method	Function
CARBayes	CARBayes	5.0	Markov chain Monte Carlo	S.CARleroux/ S.CARbym
R2OpenBUGS	R2OpenBUGS	3.2-3.2	Markov chain Monte Carlo	bugs
NIMBLE	NIMBLE	0.7.0	Markov chain Monte Carlo	nimbleMCMC
R2BayesX-MCMC	R2BayesX	1.1-1	Markov chain Monte Carlo	bayesx
R2BayesX-REML	R2BayesX	1.1-1	Empirical Bayes	bayesx
RStan-MCMC	RStan	2.18.2	Markov chain Monte Carlo	stan
RStan-VI	RStan	2.18.2	Variational inference	vb
INLA	INLA	18.07.12	Integrated nested Laplace approximation	inla

compare software packages that allow estimation of the intrinsic and convolution model in a Bayesian hierarchical framework. An overview of the software packages, the used version, the estimation method and function is given in table 1. R (R Core Team 2018), version 3.5.0, was used to perform the data analysis and simulation study. In appendix A, the implementation code of the software packages listed in table 1 is given.

OpenBUGS is the most recent version of the BUGS (Bayesian Inference Using Gibbs Sampling) project (Lunn et al. 2009), which chooses an appropriate MCMC scheme from either Gibbs, Metropolis Hasting and slice sampling. It is a general software package to perform Bayesian analyses. OpenBUGS has a built-in function to implement the intrinsic model and the convolution model: `car.normal()`. Moreover, it provides the user the possibility to write their own code. The R2OpenBUGS package offers an R interface to OpenBUGS. It was the first developed software among the packages being discussed.

An extension of the BUGS language is available in the R package NIMBLE (Numerical Inference for statistical Models for Bayesian and Likelihood Estimation, de Valpine et al. 2017, NIMBLE Development Team 2019). NIMBLE combines flexible model specification with the possibility to write a variety of algorithms adapted to the defined model. NIMBLE has a similar built-in function as OpenBUGS to implement the intrinsic and the convolution model, namely `dcar_normal()`. Depending on the model definition, several MCMC techniques can be evoked. A specialized sampling scheme is available for the spatial models. This is a random walk Metropolis Hasting algorithm.

CARBayes is specifically designed for the analysis of generalized linear mixed models for areal unit data (Lee 2013). It implements several univariate and multivariate CAR models using MCMC sampling, including the intrinsic and the convolution model. The intrinsic model can be fitted using the function `S.CARleroux()`, a function originally proposed to fit a Leroux model (Leroux et al. 1999), but which allows the user to fit the intrinsic CAR model by setting "rho" (a spatial dependence parameter) equal to one. The function `S.CARbym()`

is the implementation of the convolution model. The MCMC updates for the random effects and the regression parameters use a combination of Gibbs sampling and either random walk Metropolis steps or Metropolis adjusted Langevin algorithm (MALA, Roberts and Rosenthal 1998).

Another software package is Stan (Carpenter et al. 2017, Stan Development Team 2018a). It is a C++-based package to perform Bayesian analyses using fast MCMC algorithms including Hamiltonian Monte Carlo (HMC) and its variant, the No-U-Turn (NUTS) sampler. Stan can be accessed through R via the package RStan (Stan Development Team 2018b). There is no built-in function for CAR models, but the user can directly implement the probability function up to a proportionality constant (Morris 2018). Aside from MCMC techniques, Stan also makes use of other estimation techniques, including variational inference (VI). Here, an approximation of the posterior distribution is created by posing a family of approximate densities and to find the one of the family that minimizes the discrepancy between the approximation and the true posterior distribution. This discrepancy is measured with the Kullback-Liebler divergence. The key is choosing a family flexible enough to capture a density close to the true posterior distribution, but also not too complicated. Due to the difficulty in computing the Kullback-Liebler divergence, minimizing the Kullback-Liebler divergence is translated in optimizing the evidence lower bound (ELBO). The ELBO maximizes the difference between the expectation of the log joint posterior distribution under the approximation and the expectation of the log approximation under the approximation (Blei et al. 2017, Kucukelbir et al. 2015, Nathoo et al. 2013, Teng et al. 2017). Being an approximate method, this can be used to quickly get a rough idea of the parameter distribution. Within the context of disease mapping, the method is particularly useful when the number of areas is large, or when complex spatial models are used.

BayesX is a software package for analyzing the general class of structured additive regression (STAR) models (Brezger et al. 2005). One of the techniques implemented in BayesX uses a full Bayesian approach using MCMC simulation, which consists of a Metropolis Hasting sampling based on iteratively weighted least squares proposals. An alternative estimation method is an empirical Bayesian approach, based on mixed model methodology (Fahrmeir et al., 2004). It can also be interpreted as a penalized likelihood method. BayesX has the build-in function “mrf” or “spatial”, which defines the random effects of the intrinsic model. To include the spatially unstructured random effect for the convolution model, the function “re” or “random” is available.

Finally, integrated nested Laplace approximation (INLA, Rue et al. 2009) can be used as an alternative to the MCMC methods. INLA is a Bayesian method that computes approximations of posterior marginal distributions for latent Gaussian models. Marginal posterior distributions of the parameters in the latent field and hyperparameters are approximated via Laplace approximations. It is implemented in the R package INLA. The intrinsic model is implemented in INLA via the “besag” model and the convolution model, via the “bym” model.

Several manuscripts compare the Bayesian inference framework. For example, Taylor and Diggle (2013) compare the Metropolis adjusted Langevin algorithm with integrated nested Laplace approximation in the context of spatial log-Gaussian Cox processes. In the same context, Teng et al. (2017) made a comparison between Hamiltonian Monte Carlo, integrated nested Laplace approximation and variational inference. We focus on the performance of the different techniques implemented in a specific software package. In the next paragraph,

differences amongst model specification, MCMC characteristics and summary statistics of the different software packages is discussed.

2.2.1 Differences among model specifications

While all investigated software packages implement the intrinsic CAR and convolution models, there are some subtle differences. For example, while R2OpenBUGS, NIMBLE, RStan and INLA allow areas without neighbours, CARBayes and R2BayesX do not allow this.

Another difference among (default) implementations is the specification of priors and hyperpriors (table 2). Some packages provide default prior and hyperprior specifications and these are used in our simulation study, when available. We explicitly choose for this approach since many analysts use default program settings. If there are no defaults, the priors or hyperpriors are chosen such that they are similar for the different software packages. The priors do not differ drastically, but note that even small differences can cause differences in the estimates. The choice of the prior for the spatial and random variance parameter is even more challenging due to the issue of how to choose interpretable hyperpriors. Gelman (2006) states that the inverse-gamma family for variance parameters can cause serious problems. However, several of the software packages chose this prior as default for the variance parameter. Generally, it is not recommended to use the default priors blindly. Sørboye and Rue (2014) propose guidelines to choose a hyperprior for the precision parameter of the spatial random effects.

Besides the recommendation to not blindly use the defaults priors, the user of a software package also has to be aware of possible identifiability problems. Identifiability problems can cause that the statistical models not uniquely estimate the true values of the model parameters. Like mentioned in section 2.1.1, the intrinsic model has an improper joint distribution of the random effects, which implies an implicit overall intercept (Goicoa et al. 2018). The different software packages deal with this identifiability problem in a different way. CARBayes, R2OpenBUGS and R2BayesX automatically imposes the sum-to-zero constraint, namely $\sum_{k=1}^n \phi_k = 0$, which solves the identifiability problem. In the NIMBLE software, it is recommended to not include an additional intercept parameter, but also gives the user the possibility to impose the sum-to-zero constraint by adding the specification `zero_mean=1`. In contrast to the sum-to-zero constraint, RStan offers a soft version by setting a prior on the random effects which constraints the mean of the random effects to be within some epsilon of zero. INLA also has a default version of a sum-to-zero constraint to each connected sub-graph.

2.2.2 Differences among MCMC characteristics

Besides differences among model specifications, as discussed in the previous section, the software packages using MCMC sampling differ in the choice of MCMC features. Some, like R2BayesX-MCMC and RStan-MCMC, provide default settings while others do not. These defaults can also vary between software packages. An overview of the default/recommended characteristics of the MCMC algorithms is shown in table 3. These characteristics are the number of iterations, the burn-in part, thinning and the number of chains. If there are no defaults in the software packages concerning these characteristics, the user-specified values of this study are given. While

Table 2: Software packages. An overview of the priors and hyperpriors used for the simulation study. The variance is used in the formulation of the normal distribution.

Software package		CARBayes	R2OpenBUGS	NIMBLE	R2BayesX	RStan	INLA
Model							
Intrinsic							
μ	Prior	$\sim N(0, 10000)$	An improper (flat) prior	An improper (flat) prior	An improper (flat) prior	$\sim N(0, 102\,400)$	$\sim N(0, 0)$
	Default or user-specific	Default	User-specified	User-specified	Default	User-specified	Default
τ^2	Hyperprior	$\tau^2 \sim \text{Inverse-Gamma}(1, 0.01)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\tau^2 \sim \text{Inverse-Gamma}(0.001, 0.001)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\text{Log}\left(\frac{1}{\tau^2}\right) \sim \text{Log-gamma}(1.5 \times 10^{-5})$
	Default or user-specific	Default	User-specified	User-specified	Default	User-specified	Default
Convolution							
μ	Prior	$\sim N(0, 10000)$	An improper (flat) prior	An improper (flat) prior	An improper (flat) prior	$\sim N(0, 102\,400)$	$\sim N(0, 0)$
	Default or user-specific	Default	User-specified	User-specified	Default	User-specified	Default
τ^2	Hyperprior	$\tau^2 \sim \text{Inverse-Gamma}(1, 0.01)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\tau^2 \sim \text{Inverse-Gamma}(0.001, 0.001)$	$\frac{1}{\tau^2} \sim \text{Gamma}(1, 0.01)$	$\text{Log}\left(\frac{1}{\tau^2}\right) \sim \text{Log-gamma}(1.5 \times 10^{-4})$
	Default or user-specific	Default	User-specified	User-specified	Default	User-specified	Default
σ^2	Hyperprior	$\sigma^2 \sim \text{Inverse-Gamma}(1, 0.01)$	$\frac{1}{\sigma^2} \sim \text{Gamma}(1, 0.01)$	$\frac{1}{\sigma^2} \sim \text{Gamma}(1, 0.01)$	$\sigma^2 \sim \text{Inverse-Gamma}(0.001, 0.001)$	$\frac{1}{\sigma^2} \sim \text{Gamma}(1, 0.01)$	$\text{Log}\left(\frac{1}{\sigma^2}\right) \sim \text{Log-gamma}(1.5 \times 10^{-4})$
	Default or user-specific	Default	User-specified	User-specified	Default	User-specified	Default

Table 3: Software packages. An overview of the Markov chain Monte Carlo characteristics.

Software package	Default or chosen	Number of iterations	Burn-in	Thin	Number of chains
CARBayes	Chosen	20 000	10 000	1	1
R2OpenBUGS	Chosen	20 000	10 000	1	1
NIMBLE	Default	10 000	0	1	1
R2BayesX-MCMC	Default	12 000	2000	10	1
RStan-MCMC	Default	2000	1000	1	4

R2BayesX-MCMC uses thinning, RStan-MCMC uses multiple chains.

Some of the software packages using MCMC provide a convergence diagnostic statistic. CARBayes provides convergence checks with Geweke diagnostic (Geweke 1991). The function GRStats in R2BayesX computes Gelman and Rubin diagnostics for multiple chains (Gelman et Rubin 1992). Rstan returns warning messages when the estimated potential reduction factor, the \hat{R} -statistic (Gelman et Rubin 1992), is substantially greater than 1. Convergence can also be checked using a traceplot, as available in e.g. R2OpenBUGS.

2.2.3 Differences among summary statistics

In table 4, an overview of summary measures, which are presented by default by the software packages, is given. We should be aware of these differences when comparing results of the different packages, especially if the posterior distribution is highly skewed. Results of RStan-VI and INLA are approximations of the posterior. Furthermore, INLA gives the estimates from the marginals instead of the joint posterior distribution. R2BayesX-REML provides empirical Bayes posterior mode estimates. The software packages R2OpenBUGS, NIMBLE, RStan and INLA

Table 4: Software packages. A resume of the summary statistics of the parameters and relative risks, provided by the different software packages.

Estimates Software package	Mean	Mode	Median	Quantiles	Standard deviation
CARBayes			×	×	
R2OpenBUGS	×		×	×	×
NIMBLE	×		×	×	×
R2BayesX-MCMC	×		×	×	×
R2BayesX-REML		×			
RStan-MCMC	×		×	×	×
RStan-VI	×		×	×	×
INLA	×	×	×	×	×

provide estimates of the relative risks, while CARBayes and R2BayesX give estimates of the observed counts. The software packages CARBayes and R2Bayes do not give credible intervals of the relative risks. In this study, the intervals of the relative risks for the MCMC methods are constructed based on the MCMC samples of the parameters.

2.3 Comparison techniques

Bivand et al. (2017) mention that model comparisons using values of deviance information criterion (DIC) are challenging, as values are only comparable within each software implementation, but not among software implementations. With this argument in mind, we choose to focus on parameter and relative risk estimation, and not on model selection in this manuscript. The agreement statistic provides a tool to compare parameter estimates in the Bayesian paradigm (Faes et al. 2011). It gives an indication of the amount of overlap between the two densities of the same parameter coming from two different software packages. The agreement statistic is based on the integrated absolute difference (IAD). Let f be the posterior density proposed by one software package and \tilde{f} the posterior density proposed by another software package, then the IAD can be specified as

$$\text{IAD}(f, \tilde{f}) = \int_{-\infty}^{\infty} |\tilde{f}(x) - f(x)| dx.$$

Then, the agreement statistic can be formulated as follows

$$\text{Agreement}(f, \tilde{f}) = 1 - \frac{\text{IAD}(f, \tilde{f})}{\sup \text{IAD}(f, \tilde{f})}.$$

If $f = \tilde{f}$, the agreement is 1. Values of the agreement range between 0 and 1. The R code to compute the agreement statistic is given in appendix B. Another comparison measure is the relative efficiency. This is the ratio of the standard deviations of the relative risks in an area k compared between two software packages. Values larger than 1 indicate that the second software package provides a more efficient parameter estimate than the first. Further, the exceedance probability, which estimates the chance of having the relative risk exceeding a certain value, is considered.

For the simulation study, two kinds of mean absolute differences are also used as comparison statistics. On the one hand, we calculate the mean absolute difference per area, as

$$\text{MAD}_k = \frac{1}{m} \sum_{l=1}^m |RR_k^{(l)} - \hat{R}R_k^{(l)}|,$$

which represents the difference between the true value $RR_k^{(l)}$ and the estimated value $\hat{R}R_k^{(l)}$. On the other, we calculate the overall absolute difference for simulation (l)

$$\text{OAD}^{(l)} = \frac{1}{n} \sum_{k=1}^n |RR_k^{(l)} - \hat{R}R_k^{(l)}|,$$

which represents the average absolute difference in relative risk estimates over all areas.

3 Data analysis

In this section, the different software implementations listed in table 1 are considered for the analyze of the diabetic data.

3.1 Data description

In this study, we investigate the spatial pattern of the number of persons younger than 25 who received medication or care for diabetes per municipality in Belgium in 2014. The data contain the number of individuals that received diabetics-related medication among the total number of subscribed individuals for reimbursement. Belgium consists of 589 municipalities, with a median population of children and young adults of 3336. The data were collected by the InterMutualische Agentschap (www.ima-aim.be). The number of cases per municipality ranges from 0 to 654. Figure 1 shows the SIR of the diabetic data. An SIR bigger than 1 indicates that there are more observed diabetic cases in a municipality than expected from the overall incidence in the Belgian population. An elevated risk of diabetes around Liège, in the east of the country, can be seen from figure 1. As mentioned before, the use of a spatial model, such as the CAR model, can greatly improve our investigation. As such, this is a typical disease mapping setting. We will now use different software applications to fit the same CAR model. Although the focus is directed towards the comparison of techniques, the epidemiological question at hand is whether there are locations with increased risk.

3.2 Results

The data are analyzed with the intrinsic and the convolution model, whereby no covariates are used. The analyses are conducted with the software implementations listed in table 1. In the section “software packages”, detailed information over the used characteristics, such as e.g. the choice of hyperpriors and the burn-in part of the MCMC methods, can be found. In order to avoid convergence problems, to use the typically non-zero burn-in and in view

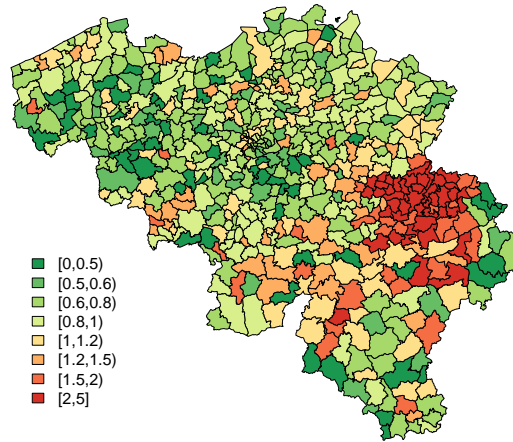


Figure 1: Children and young adult diabetics study. Standardized incidence ratio (SIR) map of the number of persons younger than 25 who are getting medication or care for diabetes in the 589 municipalities of Belgium in 2014.

of similarity with the BUGS language, the MCMC characteristics of NIMBLE are taken similar as R2OpenBUGS. This means 1 chain, 20 000 iterations, a burn-in of 10 000 and no thinning.

The results of the children and young adults diabetics data analysis are given below. The first parts discusses the differences between parameter results on basis of the estimates and the agreement of the densities. The second describes the relative risk results in terms of estimates, relative efficiency and the exceedance probability. The convergence of the different MCMC implementations was checked and no problems were found.

3.2.1 The parameters

Table 5 presents the parameter estimates of the intrinsic model and the convolution model respectively for the implementations with the different software packages. The estimates of the intercept parameter in model (2) and the variance parameters of the random effects are given. For most software packages, the estimates for the intrinsic model are consistent. A slightly higher value can be seen for the intercept parameter achieved from R2BayesX-REML. Further, RStan-VI underestimates the uncertainty and gives more influence to spatial variability. It is

Table 5: Children and young adult diabetics study. Parameter estimates and standard deviations (sd) of the intrinsic and convolution model.

Model	Intrinsic		Convolution		
	μ (sd)	τ^2 (sd)	μ (sd)	τ^2 (sd)	σ^2 (sd)
CARBayes	-0.076 (0.013)	0.168 (0.022)	-0.078 (0.013)	0.155 (0.021)	0.004 (0.002)
R2OpenBUGS	-0.076 (0.012)	0.166 (0.021)	-0.077 (0.013)	0.152 (0.021)	0.004 (0.002)
NIMBLE	-0.076 (0.012)	0.164 (0.022)	-0.078 (0.013)	0.153 (0.022)	0.005 (0.002)
R2BayesX-MCMC	-0.076 (0.013)	0.170 (0.023)	-0.077 (0.013)	0.161 (0.022)	0.002 (0.002)
R2BayesX-REML	-0.064 (0.012)	0.165	-0.066 (0.012)	0.184	1×10^{-4}
RStan-MCMC	-0.076 (0.012)	0.167 (0.021)	-0.077 (0.012)	0.153 (0.022)	0.005 (0.002)
RStan-VI	-0.075 (0.009)	0.231 (0.010)	-0.092 (0.020)	0.062 (0.011)	0.115 (0.002)
INLA	-0.076 (0.012)	0.165 (0.021)	-0.077 (0.013)	0.154 (0.021)	0.004 (0.008)

Table 6: Children and young adult diabetics study. The agreement statistic whereby the densities of the parameters from a software package are compared with the density of this parameter achieved from the software package R2OpenBUGS (in percentage).

Software package	CARBayes	NIMBLE	R2BayesX -MCMC	RStan -MCMC	RStan-VI	INLA
Model						
Intrinsic						
μ	97.63	97.67	97.20	97.16	86.71	98.43
τ^2	94.91	94.91	92.80	97.57	4.87	98.40
Convolution						
μ	95.54	96.43	96.13	97.32	61.71	98.67
τ^2	95.66	96.58	84.74	97.98	0.35	97.48
σ^2	89.59	93.63	62.11	96.43	0.12	96.82

known that variational inference generally underestimates the variance of the posterior distribution (Blei et al. 2017). For the convolution model, there are differences in estimates with deviations observed for the software packages R2BayesX and RStan-VI, with R2BayesX assigning more weight to spatial variability.

According to the similar parameter estimates for R2OpenBUGS, CARBayes, NIMBLE, RStan-MCMC and INLA, the densities obtained from these software packages have a high overlap (table 6). Table 6 shows the agreement statistic calculated as explained in section 2.3. Note a smaller agreement in table 6 between the densities resulting from R2OpenBUGS and R2BayesX-MCMC for the variance of the random effects. As expected from previous results (table 5), the agreement in density between R2OpenBUGS and RStan-VI is very low. For the intercept parameter in the intrinsic model, this is caused by a more peaked density from RStan-VI, while in the convolution model it is caused by a wider posterior density (figure 2).

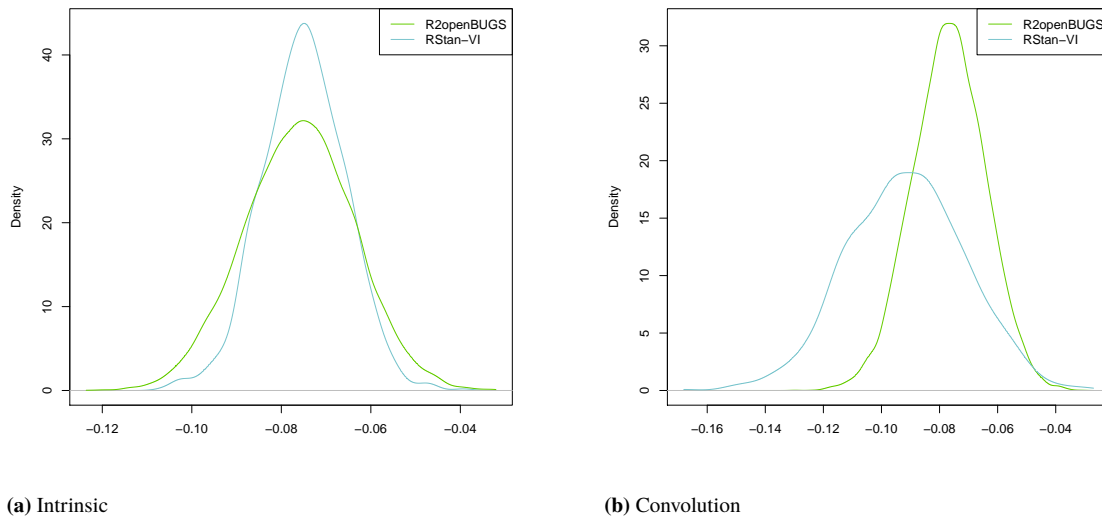
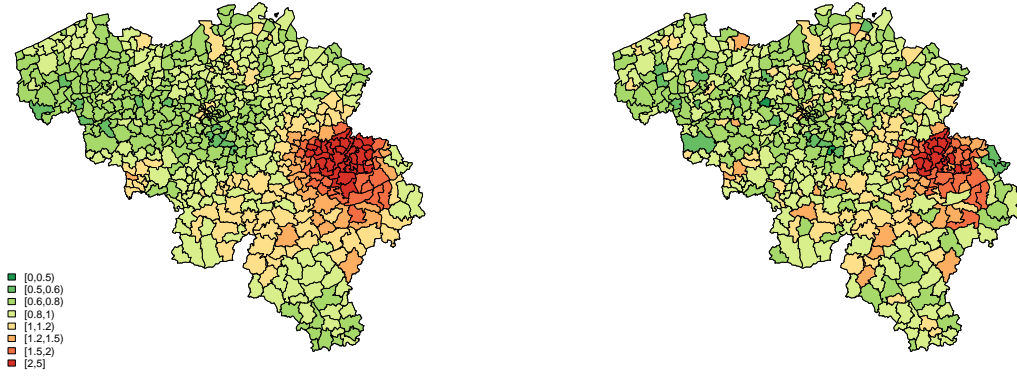


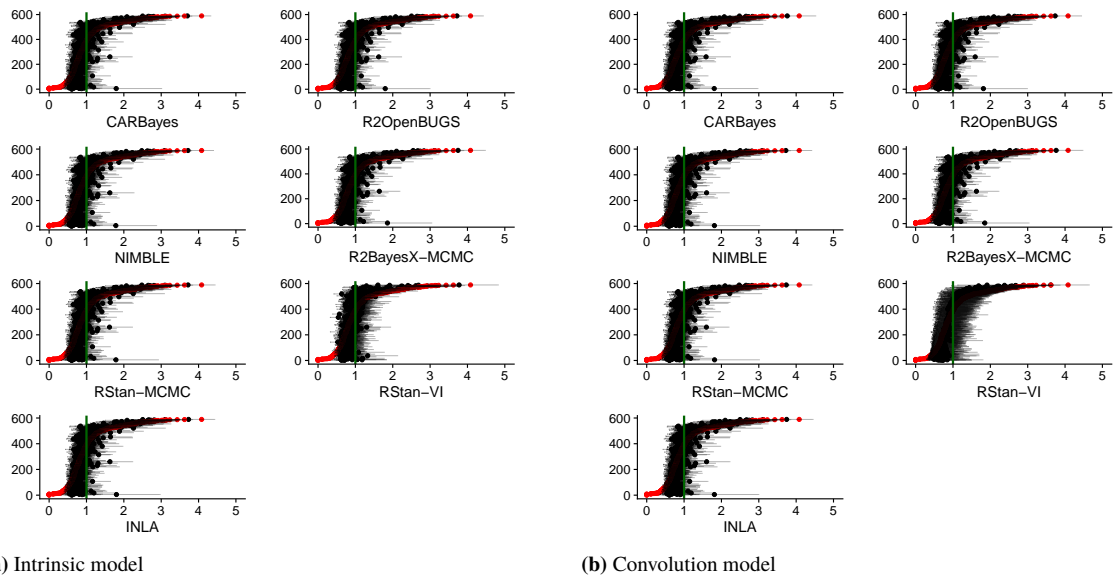
Figure 2: Children and young adult diabetics study. Map of the densities for the intercept term achieved from R2OpenBUGS, respectively RStan-VI.



(a) R2OpenBUGS

(b) RStan-VI

Figure 3: Children and young adult diabetics study. Map of point estimates of the the relative risk estimates of R2OpenBUGS and RStan-VI for the convolution model.



(a) Intrinsic model

(b) Convolution model

Figure 4: Children and young adult diabetics study. Caterpillar plots of the point and interval estimates of the relative risks, sorted by increasing standardized incidence ratio.

3.2.2 The relative risks

The relative risk estimates obtained from R2OpenBUGS and RStan-VI for the convolution model, are displayed in figure 3. The maps of the relative risk estimates coming from CARBayes, NIMBLE, R2BayesX, RStan-MCMC and INLA are similar to those from R2OpenBUGS and are presented in appendix C. Relative risk maps for the intrinsic model can also be found in appendix C. Note that the map resulting from the analysis with R2OpenBUGS is smoother than the SIR map, presented in figure 1. However, one can see an elevated risk of diabetes around Liège, in the east of the country, which was also suggested by the SIR map in figure 1. Furthermore, while the point estimates of the relative risk for most packages lead to similar results, figure 3 shows a difference in relative risk estimates between R2OpenBUGS and RStan-VI caused by the underestimation of the spatial variability by RStan-VI.



Figure 5: Children and young adult diabetics study. Relative efficiency of the relative risks. Results for the intrinsic model are given to the upper right of the diagonal. The lower left half of the table presents the results for the convolution model. The color green indicates that the software packages are about as efficient (relative efficiency between 0.95 and 1.05). The color blue denotes that the software package in the column header is more efficient (relative efficiency above 1.05 or below 0.95). Red indicates that the package denoted in the row name is more efficient than the package given in the column header (relative efficiency above 1.05 or below 0.95).

Figure 4 shows the caterpillar plots of the relative risks for the intrinsic model, the convolution model and different software packages. The caterpillar plots present the estimate of the relative risks together with an 95% equal tail credible interval. The relative risks are sorted according to increasing SIR and the red points indicate the SIR. The plots look similar in shape and extent, except for RStan-VI. For the convolution model, RStan-VI has wider credible intervals. Furthermore, the relative risks are more smoothed towards 1.

While the caterpillar plot of CARBayes is very similar to the one of R2OpenBUGS, NIMBLE, R2BayesX-MCMC, INLA and RStan-MCMC, figure 5 shows a slightly decreased efficiency for CARBayes in some areas. To significantly improve or decrease the relative efficiency, it has to lie outside the interval from 0.95 to 1.05.

Table 7: Children and young adult diabetics study. The number of municipalities in Belgium, with an exceedance probability of one above 0.95.

Software package	CARBayes	R2OpenBUGS	NIMBLE	R2BayesX -MCMC	RStan -MCMC	RStan-VI	INLA
Model							
Intrinsic	76	75	75	76	76	46	76
Convolution	75	75	76	75	75	46	76

Otherwise, the software packages are considered as about as efficient, which is indicated with the color green on figure 5. The software package R2BayesX-MCMC has differing degrees of efficiency. In line with the wider credible intervals of the software package RStan-VI, which was seen on the caterpillar plots, RStan-VI is the least efficient. The software packages R2OpenBUGS, NIMBLE, RStan-MCMC and INLA are about as efficient.

Table 7 shows how many municipalities in Belgium have an exceedance probability of one above or equal to 0.95 according to the analysis with the different software packages. Only minor differences can be seen from table 7, except for RStan-VI. Although R2BayesX-MCMC has a similar number of municipalities with a high probability of elevated disease risk, these are other municipalities. Some exceedance probabilities are higher other lower, but all small differences. The software package RStan-VI has considerably fewer municipalities with exceedance probability of one above 0.95. Overall, this data analysis shows seriously deviant results from RStan-VI for both the parameter and relative risk estimates.

4 Simulation study

In this section, a simulation study that compares the performance of the software packages listed in table 1, is discussed.

4.1 Simulation design

Disease data were simulated using the same study region as in the data analysis, namely for the 589 municipalities in Belgium. The simulated data are generated under three different scenarios: spatial independence, moderate spatial dependence and strong spatial dependence. All the data are created from model (2), with no covariates. In the independent case, the random effects are sampled from an independent normal distribution with mean 0 and standard deviation 0.6. The random effects in the strong spatial dependence scenario are generated from the intrinsic model with a standard deviation τ of 0.7. The R package SUMMER (Martin et al. 2019) was used to create the random effects. It simulates spatial random effects based on algorithm 3.1 of Rue and Held (2005). The moderate spatial dependence scenario is obtained by taking the sum of the random effects of the independent and strong spatial situation per area, similar to the convolution model. In appendix D, the figures of one of the simulated random effects under the three different scenarios is given. The intercept parameter is fixed at 0 for all scenarios. We used the same expected values as in the case study. For each of the three scenarios (independent, moderate

and strong spatial dependence), 200 datasets are created. The intrinsic and convolution model were fitted to each simulated dataset, using the different software packages listed in table 1. Due to time constraints, the analysis of the convolution model with R2BayesX-REML was interrupted if it took longer than 50 minutes. This was the case for 2 datasets in the independent scenario. The characteristics of the software packages in this simulation study are equal to those in the data analysis and are given in the section “software packages”, with deviant MCMC characteristics for NIMBLE like mentioned in the data analysis study. For the software package RStan-VI, three different scenarios were considered. The first scenario uses a single set of default initial values for the parameters. The second presents the results without the datasets for which the analysis gave a warning message. Lastly, the use of multiple initial values for the parameters was examined. Several ways to choose the initial values are possible. In this study, a random grid of initial values was taken. For the analysis with the intrinsic model, the initial values -0.5, -0.25, 0, 0.25 and 0.5 were used for the intercept and the values 0.05, 0.25, 0.45, 0.65 and 0.85 for the variance parameter. In the case of the convolution model, the initial values -0.5, 0 and 0.5 were taken for the intercept and the values 0.05, 0.4 and 0.75 for both variance parameters. The results of another choice of initial values are presented in appendix E. For each data analysis, the estimates from the results with the highest ELBO are taken into consideration.

Two additional simulation studies were conducted to investigate the generalizability of the results obtained from this simulation. While in the above simulation study, the study region (Belgium) is subdivided into 589 municipalities, we consider the subdivision in either (1) districts and (2) statistical sectors. In Belgium, districts are larger study areas as compared to municipalities, with a total of 43 districts. The statistical sectors are defined at a finer resolution as the municipalities, with a total of 19 782 statistical sectors. Data simulation is conducted in the same way as defined before and the results are presented in appendix F.

4.2 Results

In the subsequent part of this manuscript, the results of the simulation study at the level of the municipalities are explained. First, the computation time is discussed. Secondly, differences in parameter estimates are examined. Finally, a comparison of the software packages in terms of the relative risk estimates is made. We will focus on the next questions: “How well are the relative risk estimated compared to the true relative risks?”, and “Do the credible intervals of the relative risks contain the true relative risk, without being too large?”.

Convergence was checked. There were no convergence problems detected, except for RStan. The software package RStan-MCMC displayed a warning message for one analyse of the convolution model in the independent situation. This message stated that there were 2 transitions after warm-up that exceeded the maximum treedepth. It indicates an efficiency concern and does not necessarily mean no convergence. As a solution, Stan Development Team (2018c) has suggested to increase the argument `max_treedepth` above 10. This ensures that the analysis techniques of RStan take more steps in each iteration. However, this does increase the computation time. For 52 of the 1200 analyses with RStan-VI the algorithm may not have converged.

Table 8: Simulation study. Mean time in seconds (standard deviation) to run the model.

Software package	CARBayes	R2OpenBUGS	NIMBLE	R2BayesX-MCMC	R2BayesX-REML	RStan-MCMC	RStan-VI	INLA
Model								
Intrinsic	16 (0.2)	135 (4.2)	120 (2.8)	9 (0.2)	68 (16)	94 (9.5)	50 (1.3)	2 (1.3)
Convolution	21 (0.2)	197 (5.8)	154 (4.3)	18 (0.6)	880 (257)	161 (21.1)	48 (1.5)	10 (2.4)

4.2.1 Computation time

Table 8 shows the mean time, together with the standard deviation, over the 600 simulations, needed to run the model with a specific software package. Note that the computation time for the convolution model with R2BayesX-REML is underestimated since runs with a computation time longer than 50 minutes, were interrupted. The convolution model has longer computation time as compared to the intrinsic model, for all the software packages except RStan-VI (for which computation times amongst the two models are very similar). INLA has the shortest computation time, closely followed by the software packages CARBayes and R2BayesX-MCMC. Next is the package RStan-VI the fastest. It is followed by RStan-MCMC and NIMBLE, which is slightly faster as compared to R2OpenBUGS. The software package with the longest computation time is R2BayesX-REML, especially for the convolution model. While CARBayes, R2BayesX-MCMC and INLA are at least 10 times faster as compared to R2OpenBUGS, RStan-VI has a 3-fold computational gain over R2OpenBUGS. It should be noted that the presented computation time for RStan-VI corresponds to an analysis with a single set of initial values. If multiple initial values are used, the computation time should be multiplied with the number of analyses.

4.2.2 Parameter estimation

Table 9 and 10 present the minimum, mean and maximum of the parameter estimates of the intrinsic and convolution model, respectively, for the three situations (independence, moderate and strong spatial dependence). The software packages CARBayes, R2OpenBUGS, NIMBLE, R2BayesX-MCMC, RStan-MCMC and INLA give consistent results for the intrinsic model, with only a slightly larger range for the variance parameter based on CARBayes in the independence scenario. In the strong spatial dependence scenario, the intrinsic model is the correct underlying model and the parameter estimates of these software packages give on average the true parameter value (true value in the intrinsic model is 0 for μ and 0.49 for τ^2). Some deviations are seen based on R2BayesX-REML, with a tendency to yield larger estimates of the intercept and in general smaller estimates of the variance parameter. The results from the software package RStan-VI deviate much more from the other packages. On average, the estimates of the intercept improved when multiple initial values were used. On the other hand, higher estimates for the variance parameter can be noticed when working with multiple initial values, especially in the strong spatial dependence scenario.

Parameter estimates for the convolution model are found to be very similar for the software packages R2OpenBUGS, NIMBLE, RStan-MCMC and INLA. On average, the estimates are in line with the true parameter values. Results from CARBayes, R2BayesX-REML and R2BayesX-MCMC are somewhat deviating from these

Table 9: Simulation study. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the intrinsic model. Scenario 1 is independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1		2		3	
Software package		μ	τ^2	μ	τ^2	μ	τ^2
CARBayes	Min	-0.084	1.502	-0.054	0.388	-0.039	0.369
	Mean	-0.002	1.932	-9.101x10 ⁻⁴	0.560	-2.212x10 ⁻⁴	0.500
	Max	0.091	2.531	0.046	0.796	0.038	0.620
R2OpenBUGS	Min	-0.086	1.496	-0.054	0.367	-0.038	0.352
	Mean	-0.003	1.905	-8.991x10 ⁻⁴	0.550	-1.616x10 ⁻⁴	0.490
	Max	0.091	2.398	0.046	0.776	0.039	0.614
NIMBLE	Min	-0.085	1.485	-0.054	0.365	-0.038	0.348
	Mean	-0.003	1.905	-9.326x10 ⁻⁴	0.550	-1.608x10 ⁻⁴	0.489
	Max	0.091	2.407	0.047	0.779	0.038	0.613
R2BayesX-MCMC	Min	-0.086	1.509	-0.055	0.370	-0.039	0.355
	Mean	-0.003	1.915	-0.001	0.555	-4.777x10 ⁻⁴	0.494
	Max	0.091	2.410	0.045	0.782	0.039	0.614
R2BayesX-REML	Min	-0.048	1.435	-0.031	0.365	-0.018	0.348
	Mean	0.033	1.824	0.021	0.541	0.021	0.483
	Max	0.125	2.292	0.066	0.758	0.06	0.605
RStan-MCMC	Min	-0.086	1.491	-0.054	0.367	-0.038	0.350
	Mean	-0.003	1.904	-9.130x10 ⁻⁴	0.550	-1.755x10 ⁻⁴	0.489
	Max	0.091	2.399	0.046	0.773	0.039	0.607
RStan-VI	Min	-3.593	0.031	-4.077	0.005	-2.335	0.008
	Mean	0.018	0.783	-0.011	0.369	-0.015	0.453
	Max	0.731	1.862	0.201	4.837	0.185	3.155
Without datasets warning messages	Min	0.032	0.341	-0.099	0.005	-0.080	0.055
	Mean	-0.071	0.779	0.023	0.325	0.026	0.428
	Max	0.131	1.862	0.111	1.255	0.185	1.032
With multiple initial values	Min	-0.089	0.965	-0.066	0.404	-0.071	0.389
	Mean	-0.005	1.447	-0.011	0.631	-0.009	0.684
	Max	0.108	1.964	0.045	1.013	0.075	1.090
INLA	Min	-0.086	1.482	-0.054	0.365	-0.038	0.349
	Mean	-0.003	1.892	-6.682x10 ⁻⁴	0.546	9.158x10 ⁻³	0.487
	Max	0.091	2.383	0.046	0.770	0.039	0.610

packages. It can be observed that again, for the convolution model, the intercept is estimated somewhat higher when estimation is done via R2BayesX-REML. Also some slight differences are seen for the variance parameters, as compared to R2OpenBUGS, RStan-MCMC and INLA. Typically, it assigns more weight to spatial variability. For CARBayes, while in the moderate case the range of the variance parameters is narrower than for R2OpenBUGS, RStan-MCMC and INLA, the opposite occurs in the independent and strongly dependent settings. In the independent setting, there is a tendency to get higher estimates for the spatially correlated random effect, as compared to the other software packages; while in the strongly dependent case, it is the spatially unstructured variance, which is higher. A similar effect is seen for the software package R2BayesX-MCMC, but to a lesser extent, resulting in estimates in between CARBayes and the other MCMC methods. As in the setting of the intrinsic model, the parameter estimates based on RStan-VI are deviating quite a lot from the other packages. The use of multiple initial values improves the intercept estimate and assigns more weight to spatial variability in all three scenario's.

Table 10: Simulation study. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the convolution model. Scenario 1 denotes independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1			2			3		
Software package		μ	τ^2	σ^2	μ	τ^2	σ^2	μ	τ^2	σ^2
CARBayes	Min	-0.084	0.008	0.229	-0.052	0.008	0.050	-0.041	0.226	0.007
	Mean	-0.001	0.228	0.318	-6.210x10 ⁻⁴	0.153	0.085	-0.002	0.387	0.034
	Max	0.094	0.914	0.401	0.044	0.300	0.122	0.036	0.514	0.062
R2OpenBUGS	Min	-0.083	0.006	0.290	-0.052	0.008	0.054	-0.038	0.267	0.005
	Mean	-9.590x10 ⁻⁴	0.010	0.354	7.193x10 ⁻⁴	0.113	0.091	-4.954x10 ⁻⁵	0.439	0.010
	Max	0.095	0.032	0.442	0.045	0.258	0.142	0.039	0.555	0.032
NIMBLE	Min	-0.086	0.004	0.286	-0.050	0.005	0.050	-0.038	0.283	0.004
	Mean	0.002	0.011	0.354	8.739x10 ⁻⁴	0.113	0.091	-1.300x10 ⁻⁵	0.440	0.010
	Max	0.095	0.044	0.443	0.046	0.272	0.140	0.039	0.566	0.034
R2BayesX -MCMC	Min	-0.082	0.002	0.291	-0.053	0.006	0.050	-0.038	0.258	0.001
	Mean	0.001	0.009	0.356	2.937x10 ⁻⁴	0.123	0.090	-3.548x10 ⁻⁴	0.457	0.007
	Max	0.097	0.063	0.441	0.045	0.274	0.140	0.039	0.581	0.044
R2BayesX -REML	Min	-0.046	1.981x10 ⁻⁷	0.280	-0.029	0.0002	0.053	-0.018	0.248	1.028x10 ⁻⁶
	Mean	0.036	0.005	0.344	0.023	0.119	0.088	0.019	0.486	0.005
	Max	0.129	0.077	0.421	0.066	0.247	0.135	0.06	0.734	0.043
RStan -MCMC	Min	-0.083	0.007	0.288	-0.052	0.008	0.051	-0.039	0.267	0.005
	Mean	0.001	0.010	0.354	8.957x10 ⁻⁴	0.113	0.091	-5.395x10 ⁻⁵	0.442	0.009
	Max	0.095	0.035	0.441	0.046	0.263	0.140	0.039	0.564	0.038
RStan-VI	Min	-1.977	0.001	0.011	-5.877	0.002	0.014	-5.164	0.001	0.007
	Mean	-0.116	0.230	0.365	-0.066	0.125	0.215	-0.122	0.132	0.270
	Max	0.601	6.395	3.311	1.355	4.606	5.688	1.242	4.695	4.367
Without datasets warning messages	Min	-1.235	0.001	0.011	-1.284	0.002	0.015	-0.920	0.002	0.007
	Mean	-0.090	0.151	0.336	-0.047	0.061	0.165	-0.050	0.091	0.213
	Max	0.222	1.256	1.495	0.213	0.636	1.666	0.186	0.926	1.730
With multiple initial values	Min	-0.159	0.002	0.001	-0.111	0.002	0.001	-0.126	0.004	0.001
	Mean	-0.019	0.216	0.261	-0.012	0.348	0.043	-0.010	0.590	0.022
	Max	0.087	2.013	0.434	0.056	0.862	0.197	0.084	1.268	0.263
INLA	Min	-0.081	0.007	0.290	-0.052	0.008	0.053	-0.038	0.261	0.005
	Mean	0.002	0.009	0.353	0.001	0.110	0.091	2.132x10 ⁻⁴	0.448	0.009
	Max	0.097	0.034	0.439	0.046	0.252	0.140	0.040	0.568	0.038

4.2.3 The relative risks

Table 11 presents a summary of the overall absolute difference of the relative risk, over the 200 simulation runs. R2OpenBUGS, NIMBLE and RStan-MCMC have in general the smallest absolute bias in the relative risks, closely followed by R2BayesX (both MCMC and REML), INLA and CARBayes. CARBayes has, however, a somewhat larger bias in some of the simulation runs in the independent setting. The software package RStan-VI has high values of the overall absolute difference per created dataset. When leaving out results from datasets for which the analysis gave a warning message, the mean absolute bias in the relative risks is greatly reduced, though still high as compared to the other software packages. This indicates that the estimates from the software package RStan-VI are less reliable, especially the ones with warning messages. By using multiple initial values, the overall absolute difference further decreased, but is still slightly higher compared to the other software packages. Therefore, the use of multiple initial values for the parameters is recommended.

On the other hand, we can look at the mean absolute difference per area. In appendix G, figures presenting a summary of the absolute bias are given for three scenarios: a small, medium and large population. The municipalities with a small, respectively large population, are the 50th municipalities with the smallest, respectively

Table 11: Simulation study. Minimum (min), mean and maximum (max) value of the overall absolute difference per simulated dataset over the analyses of the 200 created simulated runs. Scenario 1 is independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1		2		3	
Software package		Intrinsic	Convolution	Intrinsic	Convolution	Intrinsic	Convolution
CARBayes	Min	0.200	0.189	0.166	0.159	0.158	0.161
	Mean	0.228	0.218	0.182	0.178	0.177	0.177
	Max	0.262	0.251	0.200	0.195	0.203	0.201
R2OpenBUGS	Min	0.192	0.187	0.162	0.160	0.157	0.157
	Mean	0.221	0.215	0.181	0.177	0.175	0.175
	Max	0.245	0.241	0.198	0.195	0.199	0.199
NIMBLE	Min	0.192	0.187	0.163	0.159	0.157	0.157
	Mean	0.221	0.215	0.181	0.177	0.175	0.175
	Max	0.245	0.241	0.198	0.195	0.199	0.200
R2BayesX-MCMC	Min	0.195	0.189	0.164	0.160	0.158	0.158
	Mean	0.222	0.217	0.182	0.178	0.176	0.176
	Max	0.248	0.244	0.199	0.196	0.201	0.201
R2BayesX-REML	Min	0.193	0.189	0.164	0.16	0.157	0.157
	Mean	0.222	0.217	0.182	0.178	0.170	0.176
	Max	0.246	0.242	0.198	0.196	0.201	0.204
RStan-MCMC	Min	0.192	0.187	0.162	0.159	0.157	0.157
	Mean	0.221	0.215	0.181	0.177	0.175	0.175
	Max	0.245	0.241	0.198	0.195	0.199	0.199
RStan-VI	Min	0.206	0.192	0.172	0.163	0.166	0.174
	Mean	4.499x10 ⁹	0.325	0.280	2445.748	0.247	0.330
	Max	8.998x10 ⁹	12.946	4.521	4.890x10 ⁵	1.372	11.268
Without datasets warning messages	Min	0.206	0.192	0.172	0.163	0.166	0.174
	Mean	0.267	0.222	0.231	0.188	0.226	0.206
	Max	0.391	0.254	0.340	0.253	0.407	0.256
With multiple initial values	Min	0.204	0.186	0.163	0.169	0.164	0.167
	Mean	0.227	0.218	0.185	0.187	0.185	0.191
	Max	0.254	0.244	0.205	0.214	0.223	0.236
INLA	Min	0.194	0.188	0.164	0.16	0.157	0.158
	Mean	0.222	0.217	0.182	0.178	0.175	0.175
	Max	0.246	0.243	0.198	0.195	0.200	0.200

largest, population. The median population of the municipalities with smallest, respectively medium and largest, population are 852, 3158 and 16 607 respectively. Since the use of multiple initial values give better relative risk estimates, only these results are presented for RStan-VI. Comparing CARBayes with R2OpenBUGS, there is a tendency of increased bias in the relative risks, in highly populated areas. Results from R2OpenBUGS, NIMBLE and RStan-MCMC are very similar, as well as the results from R2BayesX (MCMC and REML) and INLA. Results from Stan-VI with multiple initial values give for some areas, typically small population areas, increased bias.

Table 12 shows the mean coverage of the relative risks, estimated as the percentage of the credible intervals that contain the true relative risk. It also shows the percentage of estimated intervals that are at least 0.1 wider or narrower than the credible intervals coming from the software package R2OpenBUGS. While results

Table 12: Simulation study. For the different software packages, the first line under the specific model gives the coverage of the true relative risk. The second line under the specific model presents the percentage of estimated credible intervals that are at least 0.1 wider than the credible intervals achieved from the software package R2OpenBUGS. The third line under the specific model presents the percentage of estimated credible intervals that are at least 0.1 narrower than the credible intervals achieved from the software package R2OpenBUGS. This is an overview for the independence scenario.

Software package	CARBayes	R2OpenBUGS	NIMBLE	R2BayesX	RStan	RStan-VI	RStan-VI Multiple initial values	INLA
				-MCMC	-MCMC			
Intrinsic model								
Coverage (%)	93.76	94.52	94.56	94.32	94.48	89.63	94.38	94.42
Increased length of the credible interval (%)	9.27		0.46	2.09	0.46	55.28	30.92	1.43
Decreased length of the credible interval (%)	3.69		0.4	2.74	0.49	11.19	50.01	0.04
Convolution								
Coverage (%)	95.31	94.96	94.94	95.17	94.9	95.93	94.43	94.83
Increased length of the credible interval (%)	11.08		0.321	3.52	0.47	45.5	42.28	0.43
Decreased length of the credible interval (%)	0.37		0.3	1.13	0.46	1.51	37.74	0.005

are presented for the independence scenario, the other scenarios lead to similar conclusions; only the independent situation is a little more apparent. Except for one analysis with RStan-VI, all software packages have a high coverage probability, close to the nominal level of 95%. However, CARBayes often has wider credible intervals for some of the areas and thus is estimated with more uncertainty. When working with multiple initial values for RStan-VI, the coverage increases.

5 Discussion

In this manuscript, several of the available software implementations in the context of spatial epidemiology are compared. The investigated software packages were CARBayes, NIMBLE, R2OpenBUGS, R2BayesX, INLA and RStan. Focus was on the most commonly used models in spatial epidemiology for spatially discrete count data, namely the intrinsic and the convolution model. Software packages were examined in terms of differences in implementations characteristics, the parameter and relative risk estimates. They were applied to the data about diabetics in children and young adults in Belgium. Furthermore, simulation studies were undertaken to assess the software performance in different settings. Through the study at the level of the municipalities, the following conclusions can be drawn about the different software packages.

OpenBUGS is the standard (oldest) software package to do Bayesian disease mapping. The R package to use this program from the R environment, R2OpenBUGS, shows good performance in terms of parameter estimates and low mean absolute difference between true and estimated relative risks. On the other hand, it has the largest computation time of all MCMC implementations. As an extension of the BUGS language, the R package NIMBLE has very similar estimates as R2OpenBUGS, but preformances faster. The software package RStan with MCMC methods is also faster as compared to R2OpenBUGS and similar in time as NIMBLE. It gives stable estimates, very similar to R2OpenBUGS.

CARBayes is an R package that also uses MCMC updates for estimates. Although it is a user-friendly software packages, some differences in comparison to the software packages R2OpenBUGS, NIMBLE, RStan with MCMC techniques and INLA occur. From the simulation study, we noticed slight differences in variance estimates for the random effects of the convolution model compared to R2OpenBUGS, NIMBLE, RStan with MCMC methods and INLA. In addition, the software package has slightly more difficulties with estimating the relative risks in large population areas. Moreover, the credible intervals of the relative risks resulting from CARBayes are wider.

The software package R2BayesX with MCMC techniques gives somewhat different results. This could be noticed in the data analysis: R2BayesX with MCMC methods assigns more weight to spatial variability, fluctuates between being less and more efficient and reports some other exceedance probabilities. Further, in the simulation study, the parameter estimates are in between those coming from CARBayes and OpenBUGS.

Besides MCMC implementations, approximation methods were considered. One of these is INLA. Although INLA uses an approximation to estimate parameters and relative risks, it has very good performances and it gives similar results as compared to R2OpenBUGS, NIMBLE and RStan with MCMC techniques. Moreover, it is the fastest implementation technique for the considered disease mapping models.

The software packages R2BayesX and RStan offer, besides MCMC techniques, also approximation methods. We do not recommend the use of the software package R2BayesX based on mixed model methodology and RStan via variational inference in this context of disease mapping models. R2BayesX based on mixed model methodology has the longest running time and shows bias in parameter estimates compared to R2OpenBUGS, RStan with MCMC techniques and INLA. The software package RStan via variational inference uses a fast approximation method. However, it results in deviant parameter estimates and gives poor relative risk estimates as seen for the mean absolute difference comparison technique used in the simulation study. Working with multiple initial values will improve the relative risks estimates, but it will also give loss to time gain compared to the other software packages. These conclusions are based on a geographic region with many subdivided areas. To see if the number of subdivided areas has an impact on the performance of the different software packages, two additional simulation studies were conducted, one with less subdivided areas and one with more. In the situation with less subdivided areas, the computation time seriously decrease. Only the software package RStan has almost no gain in computation time. However, the performance of software packages in terms of computation time depends on the characteristics of the data like for example the number of subdivide, the computer specifications, ... More importantly, the performance in terms of parameters and relative risks estimates leads to similar conclusion. When the number of subdivided areas seriously increase, some problems arise. The computational time, for example, exponential increases. Therefore, further investigation is needed to compare the performance of the software packages for a huge amount of subdivided areas. These results can, of course, not be generalized to other contexts.

The software packages were compared for parameter and relative risk estimates. Future work will involve the examination on how the different software packages do model comparison. This study is also limited to the most commonly used spatial models. It would be interesting to conduct a similar study with spatio-temporal models.

References

- Anderson, C., Lee, D., and Dean, N. (2017) Spatial clustering of average risks and risk trends in Bayesian disease mapping, *Biometrical Journal*, 59, 41–56, doi: 10.1002/bimj.201600018.
- Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43, 1–59, <https://doi.org/10.1007/BF00116466>.
- Bivand, R., Sha, Z., Osland, L., and Thorsen, I.S. (2017) A comparison of estimation methods for multilevel models of spatially structured data, *Spatial Statistics*, 21, 440–459, <http://dx.doi.org/10.1016/j.spasta.2017.01.002>.
- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017) Variational Inference: A Review for Statisticians, *Journal of the American Statistical Association*, 112, 518, doi:10.1080/01621459.2017.1285773.
- Brezger, A., Kneib, T., and Lang, S. (2005) BayesX: Analyzing Bayesian structural additive regression models. *Journal of Statistical Software*, 14 (11), 1–22, doi: 10.18637/jss.v014.i11.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language, *Journal of Statistical Software*, 76(1), <http://dx.doi.org/10.18637/jss.v076.i01>.
- Carroll, R., Lawson, A.B., Faes, C., Kirby, R.S, Aregay, M., and Watjou, K. (2015) Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping, *Elsevier. Spatial and Spatio-temporal Epidemiology*, 14–15, 45–54, doi: 10.1016/j.sste.2015.08.001.
- Diggle, P., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics, *Applied Statistics*, 47, 299–350, <http://dx.doi.org/10.1111/1467-9876.00113>.
- Eberley, L.E., and Carlin, B.P. (2000) Identifiability and convergence issues for Markov Chain Monte Carlo fitting of spatial models, *Statistics in Medicine*, 19(17-18), 2279–2294, [http://dx.doi.org/10.1002/1097-0258\(20000915/30\)19:17/18%3C2279::AID-SIM569%3E3.0.CO;2-R](http://dx.doi.org/10.1002/1097-0258(20000915/30)19:17/18%3C2279::AID-SIM569%3E3.0.CO;2-R).
- Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2001) *Spatial Epidemiology: Methods and applications*. New York: Oxford University Press Inc.
- Faes C., Ormerod, J. T., and Wand, M. P. (2011) Variational Bayesian inference for parametric and nonparametric regression with missing data, *Journal of the American Statistical Association*, 106 (495), <http://dx.doi.org/10.1198/jasa.2011.tm10301>.
- Fahrmeir, L., Kneib, T., and Lang, S. (2004) Penalized Structured Additive Regression for Space-Time Data: a Bayesian Perspective, *Statistica Sinica*, 14 , 731-761.

- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models, *Bayesian Anal*, 1 (3), 515–534, doi:10.1214/06-BA117A.
- Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, 7(4):457–472, <http://dx.doi.org/10.1214/ss/1177011136>.
- Geweke, J. (1991) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, *Bayesian statistics*, University Press, 169-193.
- Goicoa, T., Adin, A., Ugarte, M. D., and Hodges, J. S. (2018). In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results, *Stochastic Environmental Research and Risk Assessment*, 32(3), 749-770, <https://doi.org/10.1007/s00477-017-1405-0>.
- Kucukelbir, A., Ranganath, R., Gelman, A., and Blei, D. M. (2015). Automatic variational inference in Stan, *arXiv: 1506.03431*. 28.
- Lawson, A. B. (2018) Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, *Chapman & Hall/CRC interdisciplinary statistics series*, Third Edition.
- Lee, D. (2013) CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors, *Journal of Statistical Software*, 55(13), 1-24, <http://dx.doi.org/10.18637/jss.v055.i13>.
- Leroux, B.G., Lei, X., and Breslow, N. (1999) Estimation of disease rates in small areas: a new mixed model for spatial dependence, *Statistical Models in Epidemiology; the Environment and Clinical Trials*, 179-192.
- Lindgren, F., and Rue, H. (2015). Bayesian Spatial Modelling with R-INLA, *Journal of Statistical Software*, 63(19), 1-25, doi: 10.18637/jss.v063.i19.
- Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009) The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine*, 28, 3049–3082, <http://dx.doi.org/10.1002/sim.3680>.
- Martin, B. D., Li, Z. R., Hsiao, Y., Godwin, J., Wakefield, J., and Clark, S. J. (2019) Spatio-Temporal Under-Five Mortality Methods for Estimation, R package version 0.2.2., URL <https://CRAN.R-project.org/package=SUMMER>.
- Morris, M. (2018). Spatial Models in Stan: Intrinsic Auto-Regressive Models for Areal Data. Retrieved April 2018 from URL http://mc-stan.org/users/documentation/case-studies/icar_stan.html.
- Nathoo, F., Lesperance, M.L., Lawson, A. B. and Dean C. B. (2013) Comparing variational Bayes with Markov chain Monte Carlo for Bayesian computation in neuroimaging, *Stat Methods Med Res.*, 22(4), 398-423. doi: 10.1177/0962280212448973.
- Neyens, T., Faes, C., and Molenberghs G. (2018) Integrated nested Laplace approximation for the analysis of count data via the combined model: a simulation study, *Communications in Statistics - Simulation and Computation*, <http://dx.doi.org/10.1080/03610918.2017.1400053>.

- NIMBLE Development Team (2019) NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling, R package version 0.7.0, URL <https://cran.r-project.org/package=nimble>, doi: 10.5281/zenodo.1211190.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Roberts, G., and Rosenthal, J. (1998) Optimal scaling of discrete approximations to the Langevin diffusions, *Journal of the Royal Statistical Society Series B*, 60, 225-268, <http://dx.doi.org/10.1111/1467-9868.00123>.
- Rue, H., and Held, L. (2005) Gaussian Markov Random Fields Theory and Applications, *Chapmann & Hall/CRC*, 1st Edition.
- Rue, H., Martino, S., and Chopin, N. (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society, Series B*, 71(2), 319-392, URL www.r-inla.org.
- Stan Development Team (2018a) Stan Modeling Language: User's Guide, Version 2.18, URL <http://mc-stan.org/>.
- Stan Development Team (2018b) RStan: the R interface to Stan. R package version 2.18.2., URL <http://mc-stan.org/>.
- Stan Development Team (2018c) Brief Guide to Stan's Warnings. Retrieved April 2018 from URL <http://mc-stan.org/misc/warnings.html>.
- Sørbye, S. H., and Rue, H. (2014) Scaling intrinsic Gaussian Markov random field priors in spatial modelling, *Spatial statistics*, 8, 39-51, doi: <https://doi.org/10.1016/j.spasta.2013.06.004>.
- Taylor, B. M., and Diggle, P. J., (2013) INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes, *Journal of Statistical Computation and Simulation*, 84, 10 2266-2284, doi: <https://doi.org/10.1080/00949655.2013.788653>.
- Teng, M., Nathoo, F. S., and Johnson, T., D. (2017) Bayesian Computation for Log-Gaussian Cox Processes: A Comparative Analysis of Methods. *Journal of Statistical Computation and Simulation*, 87(11): 2227-2252, doi:10.1080/00949655.2017.1326117.
- Umlauf, N., Adler, D., Kneib, T., Lang, S., and Zeileis, A. (2015). Structured Additive Regression Models: An R Interface to BayesX, *Journal of Statistical Software*, 63(21), 1-46. URL <http://www.jstatsoft.org/v63/i21/>.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE, *Journal of Computational and Graphical Statistics*, 26, 2, 403-413, doi: 10.1080/10618600.2016.1172487.

Appendix A. Software codes

- **CARBayes**

```
formula <- Observed ~ offset(log(Expected))
adjn <- poly2nb(map.shp, queen=TRUE)
W <- nb2mat(adjn, style="B")
```

The code of the intrinsic model

```
S.CARleroux(formula=formula, family="poisson", W=W, burnin=burnin, n.sample=n.sample,
fix.rho=TRUE, rho=1)
```

The code of the convolution model

```
S.CARbym(formula=formula, family="poisson", W=W, burnin=burnin, n.sample=n.sample)
```

- **R2OpenBUGS**

```
adjn <- poly2nb(map.shp, queen=TRUE)
adj=unlist(adjn)
num=card(adjn)
sumNumNeigh=sum(card(adjn))
```

The code of the intrinsic model

```
model {
  for (k in 1 : n_area) {
    Y[k] ~ dpois(b[k])
    log(b[k]) <- log(E[k]) + mu + phi[k]
    RR[k] <- exp(mu+phi[k]) # Area-specific relative risks

    # Exeedance probability
    PP1[k] <- step(RR[k]-1+eps)
    PP2[k] <- step(RR[k]-2+eps)  }
  eps <- 1.0E-6
```

```

# CAR prior distribution for random effects
phi[1:n_area] ~ car.normal(adj[],weights[],num[],pr)
for(i in 1:sumNumNeigh)
{  weights[i] <- 1  }

# Other priors and hyperpriors:
mu ~ dflat()
pr ~ dgamma(1,0.01) # prior on precision
tau2 <- 1/pr  }

```

The code of the convolution model

```

model {
  for (k in 1 : n_area) {
    Y[k] ~ dpois(b[k])
    log(b[k]) <- log(E[k]) + mu + phi[k] + psi[k]
    RR[k] <- exp(mu+phi[k]+psi[k])      # Area-specific relative risks
    psi[k] ~ dnorm(0,sigma.pr)        # Unstructured random effects

    # Exceedance probability
    PP1[k] <- step(RR[k]-1+eps)
    PP2[k] <- step(RR[k]-2+eps)  }
  eps <- 1.0E-6

  # CAR prior distribution for spatial random effects
  phi[1:n_area] ~ car.normal(adj[],weights[],num[],pr.tau)
  for(i in 1:sumNumNeigh)
  {  weights[i] <- 1  }

  # Other priors and hyperpriors:
  mu ~ dflat()
  pr.tau ~ dgamma(1,0.01) # prior on precision of unstructured random effects
  tau2 <- 1/pr.tau
  sigma.pr ~ dgamma(1,0.01) # prior on precision of spatial random effects
  sigma2 <- 1/sigma.pr  }

```

- **NIMBLE**

```
adjn <- poly2nb(map.shp, queen=TRUE)
adj=unlist(adjn)
num=card(adjn)
sumNumNeigh=sum(card(adjn))
```

The code of the intrinsic model

```
nimbleCode({
  for (k in 1 : n_area) {
    Y[k] ~ dpois(b[k])
    log(b[k]) <- log(E[k]) + mu + phi[k]
    RR[k] <- exp(mu+phi[k]) # Area-specific relative risks

    # Exceedance RR
    PP1[k] <- step(RR[k]-1+eps)
    PP2[k] <- step(RR[k]-2+eps) }
  eps <- 1.0E-6

  # CAR prior distribution for random effects
  phi[1:n_area] ~ dcar_normal(adj[1:sumNumNeigh],weights[1:sumNumNeigh],
                              num[1:n_area], pr, zero_mean=1)

  for(i in 1:sumNumNeigh)
  { weights[i] <- 1 }

  #other priors:
  mu ~ dflat()
  pr ~ dgamma(1,0.01) # prior on precision
  tau2 <- 1/pr })
```

The code of the convolution model

```
nimbleCode({
  for (k in 1 : n_area) {
    Y[k] ~ dpois(b[k])
    log(b[k]) <- log(E[k]) + mu + phi[k] + psi[k]
```

```

RR[k] <- exp(mu+phi[k]+psi[k]) # Area-specific relative risks
psi[k] ~ dnorm(0,sigma.pr) # Unstructured random effects

# Exceedance RR
PP1[k] <- step(RR[k]-1+eps)
PP2[k] <- step(RR[k]-2+eps) }
eps <- 1.0E-6

# CAR prior distribution for spatial random effects
phi[1:n_area] ~ dcar_normal(adj[1:sumNumNeigh],weights[1:sumNumNeigh],
                             num[1:n_area], tau.pr, zero_mean=1)
for(i in 1:sumNumNeigh)
{ weights[i] <- 1 }

#other priors:
mu ~ dflat()
tau.pr ~ dgamma(1,0.01) # prior on precision of unstructured random effects
sigma.pr ~ dgamma(1,0.01) # prior on precision of spatial random effects
tau2 <- 1/tau.pr
sigma2 <- 1/sigma.pr })

```

- **R2BayesX**

```

adjn <- poly2nb(map.shp, queen=TRUE)
reg <- nb2gra(adjn)
reg.struct <- as.numeric(rownames(reg))
data <- data.frame(reg.struct,Y,E,reg.struct)

```

The code of the intrinsic model

```

formula_intr <- Y ~ sx(reg.struct,bs="spatial",map=reg)
bayesx(formula_intr,family="poisson",method="..", data=data,offset=log(data$E))

```

The code of the convolution model

```

formula_conv <- Y ~ sx(reg.struct,bs="random") + sx(reg.struct,bs="spatial",map=reg)
bayesx(formula_conv,family="poisson",method="..", data=data,offset=log(data$E))

```

method="MCMC" or "REML" depending on the estimation technique

- **RStan**

```
adjn <- poly2nb(map.shp, queen=TRUE)
adj=unlist(adjn)
num=card(adjn)

n_edges=length(adj)/2
node1=rep(NA,n_edges)
node2=rep(NA,n_edges)

iAdj=0
Edge=0
for (i in 1:n_area) {
  for (j in 1:num[i]) {
    iAdj = iAdj + 1
    if (i < adj[iAdj]) {
      Edge = Edge + 1
      node1[Edge] = i
      node2[Edge] = adj[iAdj] } } }
```

The code of the intrinsic model

```
data {
  int<lower=0> n_area; // The number of areas
  int<lower=0> n_edges;
  int<lower=1, upper=n_area> node1[n_edges]; // node1[i] adjacent to node2[i]
  int<lower=1, upper=n_area> node2[n_edges]; // and node1[i] < node2[i]

  int<lower=0> Y[n_area]; // count outcomes
  vector<lower=0>[n_area] E; // expected numbers
}

transformed data {
  vector[n_area] log_E = log(E); }

parameters {
  real mu; // intercept
  real<lower=0> pr; // precision of spatial effects
```

```

vector[n_area] phi; }

transformed parameters {
  real<lower=0> tau=inv(sqrt(pr)); // convert precision to standard deviation
}

model {
  Y ~ poisson_log(log_E + mu + phi * tau);
  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  mu ~ normal(0.0, 320);
  pr ~ gamma(1.0, 0.01);
  // soft sum-to-zero constraint on phi
  sum(phi) ~ normal(0, 0.001 * n_area); // equivalent to mean(phi) ~ normal(0, 0.001)
}

generated quantities {
  vector[n_area] eta = mu + phi * tau;
  vector[n_area] RE = phi * tau; // Relative risks
  vector[n_area] RR = exp(eta);
  real<lower=0> tau2 = square(tau); }

```

The code of the convolution model

```

data {
  int<lower=0> n_area; // The number of areas
  int<lower=0> n_edges;
  int<lower=1, upper=n_area> node1[n_edges]; // node1[i] adjacent to node2[i]
  int<lower=1, upper=n_area> node2[n_edges]; // and node1[i] < node2[i]

  int<lower=0> Y[n_area]; // count outcomes
  vector<lower=0>[n_area] E; // expected numbers
}

transformed data {
  vector[n_area] log_E = log(E); }

```

```

parameters {
  real mu;          // intercept
  real<lower=0> pr_psi; // precision of heterogeneous effects
  real<lower=0> pr_phi; // precision of spatial effects

  vector[n_area] psi; // heterogeneous effects
  vector[n_area] phi; // spatial effects
}

transformed parameters {
  real<lower=0> sigma = inv(sqrt(pr_psi)); // convert precision to standard deviation
  real<lower=0> tau = inv(sqrt(pr_phi)); // convert precision to standard deviation
}

model {
  Y ~ poisson_log(log_E + mu + phi * tau + psi * sigma);
  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  mu ~ normal(0.0, 320);
  // soft sum-to-zero constraint on phi
  sum(phi) ~ normal(0, 0.001 * n_area); // equivalent to mean(phi) ~ normal(0,0.001)

  pr_psi ~ gamma(1.0, 0.01);
  pr_phi ~ gamma(1, 0.01);

  psi ~ normal(0,1); }

generated quantities {
  vector[n_area] eta = mu + phi * tau + psi * sigma;
  vector[n_area] RR = exp(eta); // Relative risks
  vector[n_area] RE = phi * tau + psi * sigma;
  real<lower=0> tau2=1/pr_phi;
  real<lower=0> sigma2=1/pr_psi; }

```

- INLA

```

reg.struct.inla <- seq(1,n_area)
adjn <- poly2nb(map.shp, queen=TRUE)

```



```
nb2INLA("adjn.inla", adjn)
data <- data.frame(reg.struct.inla,Y,E,reg.struct.inla)
```

The code of the intrinsic model

```
formula_intr=Y~f(reg.struct.inla, model="besag",graph="adjn.inla")
```

```
inla(formula_intr,control.inla = list(strategy = "simplified.laplace"),family="poisson",
E=data$E, data=data)
```

The code of the convolution model

```
formula_conv=Y~f(reg.struct.inla, model="bym",graph="adjn.inla")
```

```
inla(formula_conv,control.inla = list(strategy = "simplified.laplace"),family="poisson",
E=data$E, data=data)
```

Appendix B. Agreement statistic

```
DensParamSoftw1 <- density(MCMCsamplesParamSoftw1,
from=min(MCMCsamplesParamSoftw1, MCMCsamplesParamSoftw2),
to=max(MCMCsamplesParamSoftw1, MCMCsamplesParamSoftw2),n=2^(16))
```

```
DensParamSoftw2 <- density(MCMCsamplesParamSoftw2,
from=min(MCMCsamplesParamSoftw1, MCMCsamplesParamSoftw2),
to=max(MCMCsamplesParamSoftw1, MCMCsamplesParamSoftw2),n=2^(16))
```

```
AbsDiffDens = abs(DensParamSoftw1$y- DensParamSoftw2$y)
```

```
StepsizeDens=(max(DensParamSoftw1$x)-min(DensParamSoftw1$x))/(length(DensParamSoftw1$x)-1)
```

```
AgreementParam=1-0.5*sum(AbsDiffDens*StepsizeDens)
```

Appendix C. Relative risks children and young adult diabetics study

Figure 6: Children and young adult diabetics study. Map of the relative risks estimates for the different software packages and for the intrinsic model.

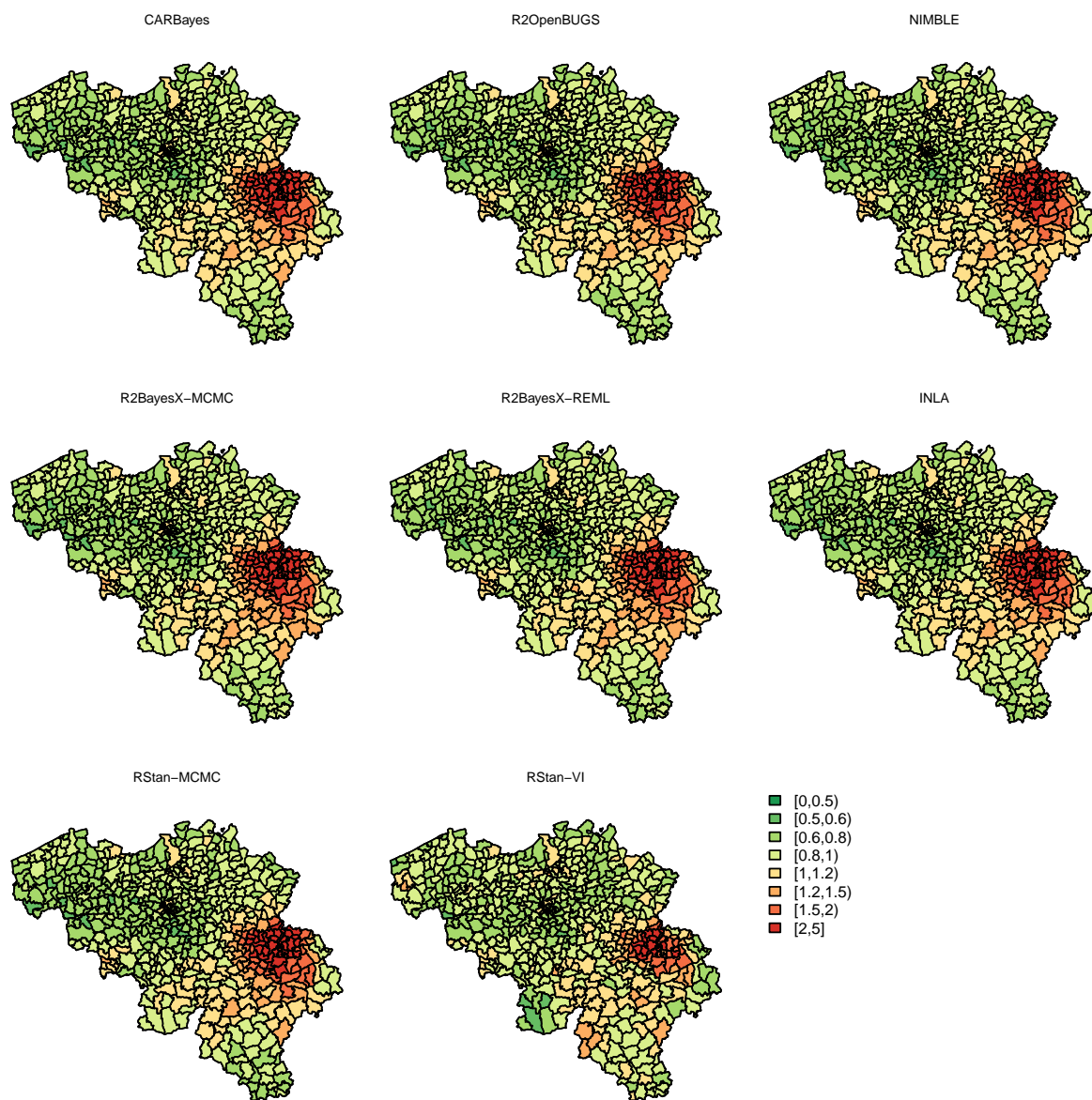
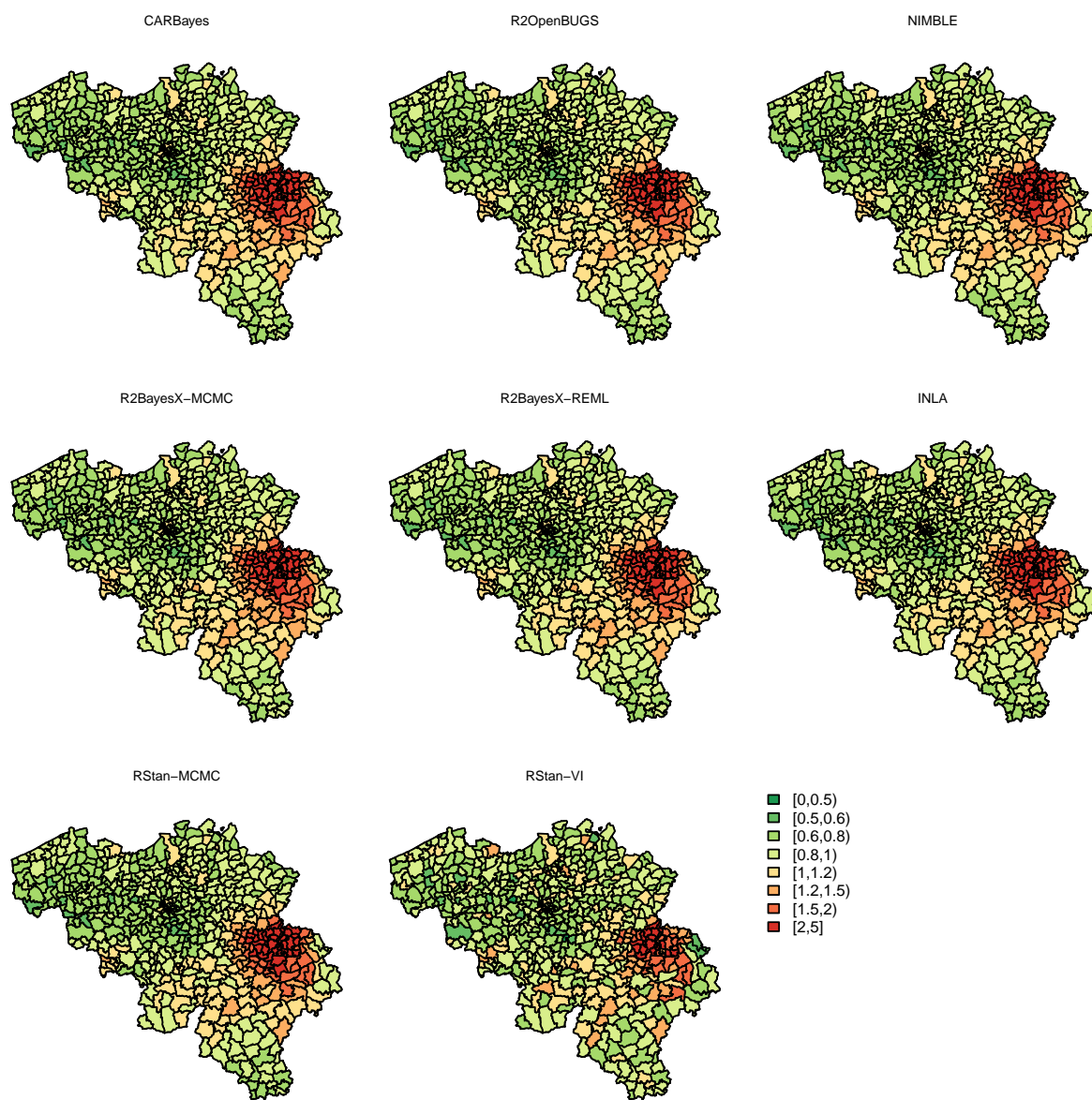
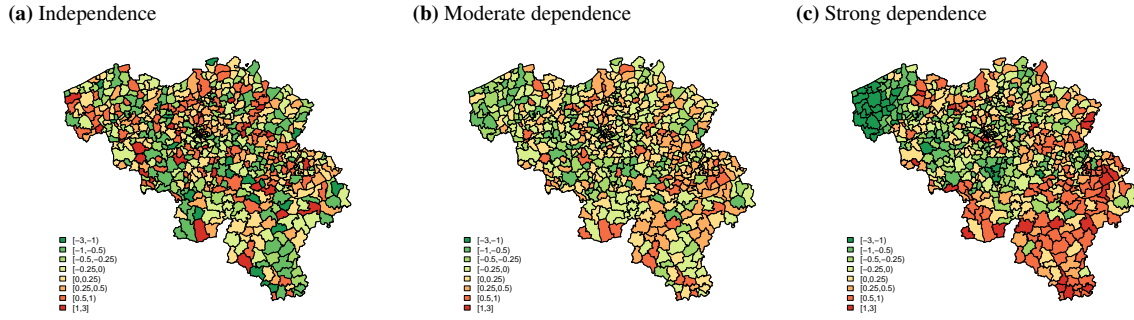


Figure 7: Children and young adult diabetics study. Map of the relative risks estimates for the different software packages and for the convolution model.



Appendix D. Random effects simulation study.

Figure 8: Simulation study. Map of one of the simulated random effects under the three different scenarios: spatial independence, moderate spatial dependence and strong spatial dependence.



Appendix E. Multiple initial values RStan-VI

In this appendix, the results of the use of particular multiple initial values for the parameters in the analysis with the software package RStan-VI are given. The parameter estimates in the previous analysis for which the ELBO was maximized were chosen as initial values. Table 13 and 14 present the minimum, mean and maximum of the parameter estimates of the intrinsic and convolution model, respectively, for the three situations (independence, moderate and strong spatial dependence). A summary of the overall absolute difference of the relative risks, over the 200 simulation runs, is given in table 15.

Table 13: Simulation study. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the intrinsic model. Scenario 1 is independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario	1		2		3	
	μ	τ^2	μ	τ^2	μ	τ^2
Min	-0.110	0.880	-0.086	0.389	-0.078	0.319
Mean	-0.007	1.424	-0.017	0.626	-0.009	0.668
Max	0.088	1.921	0.058	1.053	0.069	1.421

Table 14: Simulation study. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the convolution model. Scenario 1 denotes independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario	1			2			3		
	μ	τ^2	σ^2	μ	τ^2	σ^2	μ	τ^2	σ^2
Min	-0.133	0.002	0.008	-0.087	0.003	0.001	-0.081	0.018	0.001
Mean	-0.027	0.124	0.282	-0.020	0.386	0.037	-0.014	0.551	0.022
Max	0.076	1.507	0.450	0.045	0.828	0.169	0.072	1.095	0.257

Table 15: Simulation study. Minimum (min), mean and maximum (max) value of the overall absolute difference per simulated dataset over the analyses of the 200 created simulated runs. Scenario 1 is independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario	1		2		3	
	Intrinsic	Convolution	Intrinsic	Convolution	Intrinsic	Convolution
Min	0.2	0.185	0.165	0.163	0.164	0.163
Mean	0.229	0.218	0.186	0.187	0.185	0.191
Max	0.282	0.247	0.216	0.225	0.22	0.227

Appendix F. Simulation study districts.

As a means to investigate software performance for datasets of small sizes, we present a simulation study in which 43 Belgian districts rather than 589 municipalities are the spatial entities of interest. The expected values are calculated based on the age, gender, population size of a district in comparison to the rate of people having diabetic in the respectively age, gender category of Belgium. The study design is identical to the simulation study with municipalities in Belgium. Figure 9 shows one of simulated random effects under the three different scenarios, which were defined as independence, moderate spatial dependence and strong spatial dependence. No convergence problems were detected, except for the software package RStan. For 456 of the 600 simulation runs, RStan-MCMC stated that a number of transitions after warm-up had exceeded the maximum treedepth. For two analyses of the convolution model in the independent scenario and one analysis in the strong spatial dependence scenario, RStan-MCMC reported that there were a number of divergent transitions after warm-up. As a solution, Stan Development (2018c) suggested to increase the argument `adapt_delta` above 0.8. This ensures that RStan takes smaller steps, which could possible avoid divergence problems, at the cost of computational efficiency. For 1024 of the 1200 performed analyses with RStan-VI, the algorithm may not have converged.

F.1. Computation time

Table 16 presents the mean time, together with the standard deviation, over the 600 simulations needed to run the model with a specific software package. Except for RStan, there is a huge time gain when the number of subdivided

Figure 9: Simulation study districts. Map of one of the simulated random effects under the three different scenarios: spatial independence, moderate spatial dependence and strong spatial dependence.

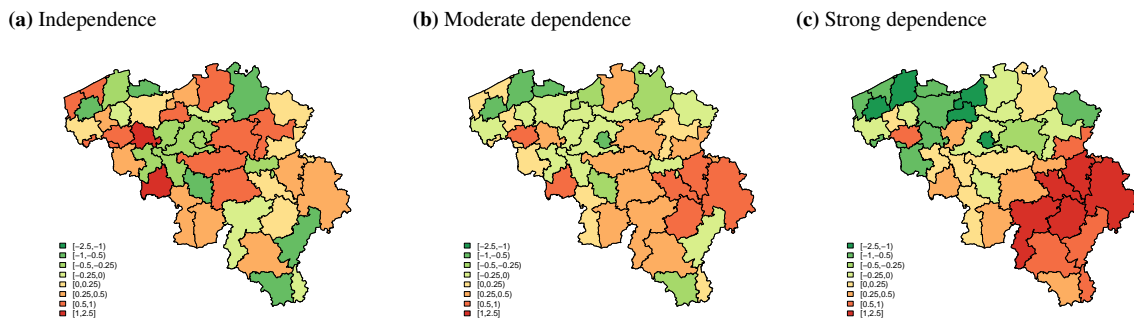


Table 16: Simulation study districts. Mean time in seconds (standard deviation) to run the model.

Software package	CARBayes	R2OpenBUGS	NIMBLE	R2BayesX-MCMC	R2BayesX-REML	RStan-MCMC	RStan-VI	INLA
Model								
Intrinsic	2 (0.1)	13 (0.9)	39 (8.4)	1 (0.1)	1 (0.1)	125 (25.3)	107 (29.0)	3 (1.0)
Convolution	3 (0.1)	18 (1.2)	32 (2.8)	1 (0.1)	1 (1.1)	144 (33.6)	92 (28.8)	3 (2.1)

areas decreases. Similarly as for the simulation study of the municipalities in Belgium, the software packages CARBayes, R2BayesX-MCMC and INLA have the shortest computation time. For this simulation study, the approximation method R2BayesX-REML also has one of the smallest computation times. In contrast, the software package R2OpenBUGS is faster than NIMBLE and RStan (both MCMC as VI) has the longest computation time.

F.2. Parameter estimation

The minimum, mean, and maximum of the parameter estimates of the intrinsic and convolution model, for the three scenarios, are shown in table 17 and 18. Due to an infinite precision for the random noise component of the random effects of the convolution model from INLA in the strong spatial dependence scenario, one analysis of the convolution model from INLA is not taken into account. In line with the simulation study on the municipality scale, the software packages R2OpenBUGS, NIMBLE, RStan-MCMC and INLA show similar results and the estimates are in line with the expected values for the corresponding scenario. The software package CARBayes has slightly different estimates for the variance parameters as compared to R2OpenBUGS, NIMBLE, RStan-MCMC and INLA. For example, it assigns more weight to spatial variability in the independence and the moderate spatial dependence scenario for the convolution parameter estimates. R2BayesX-MCMC also gives somewhat deviant estimates, as it assigns more weight to spatial variability for the intrinsic parameter estimates in the independence and moderate spatial dependence scenario. Bias in parameter estimates can also be noticed for the approximation method R2BayesX-REML. Especially in the strong spatial dependence scenario, very large spatial variance estimates for the convolution model are reported. Furthermore, the parameter estimates based on RStan-VI differ considerably from the other packages.

F.3. Relative risks

A summary of the overall absolute difference of relative risks, over the 200 simulation runs, is presented in table 19. Again, similar conclusions as seen in the simulation study at the level of the municipalities can be made for the relative risk estimates. The software package R2OpenBUGS, NIMBLE, RStan-MCMC, R2BayesX (both MCMC and REML) and INLA have the smallest absolute bias in the relative risks. Large bias is seen in some of the simulation runs from CARBayes. Furthermore, RStan-VI has higher values of the overall absolute difference per created dataset.

Table 17: Simulation study districts. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the intrinsic model. Scenario 1 denotes independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1		2		3	
Software package		μ	τ^2	μ	τ^2	μ	τ^2
CARBayes	Min	-0.264	0.712	-0.132	0.251	-0.0349	0.215
	Mean	0.008	1.664	-0.003	0.528	0.012	0.485
	Max	0.242	3.045	0.1019	0.997	0.064	0.784
R2OpenBUGS	Min	-0.264	0.696	-0.131	0.239	0.003	0.201
	Mean	0.003	1.626	-0.004	0.512	0.011	0.466
	Max	0.202	3.013	0.1013	0.927	0.023	0.754
NIMBLE	Min	-0.265	0.694	-0.132	0.239	0.003	0.202
	Mean	0.003	1.625	-0.004	0.512	0.011	0.466
	Max	0.202	3.009	0.101	0.935	0.023	0.761
R2BayesX-MCMC	Min	-0.264	0.726	-0.132	0.255	0.002	0.213
	Mean	0.002	1.706	-0.005	0.540	0.011	0.492
	Max	0.202	3.104	0.102	0.957	0.0232	0.795
R2BayesX-REML	Min	-0.259	0.717	-0.128	0.249	0.006	0.209
	Mean	0.006	1.672	-0.0015	0.529	0.014	0.481
	Max	0.205	3.089	0.104	0.955	0.026	0.779
RStan-MCMC	Min	-0.265	0.695	-0.131	0.237	0.003	0.203
	Mean	0.003	1.629	-0.004	0.512	0.011	0.465
	Max	0.202	2.995	0.101	0.912	0.022	0.742
RStan-VI	Min	-6.861	4.099 x10 ⁻⁴	-1.421	3.424 x10 ⁻⁴	-2.857	5.897x10 ⁻⁴
	Mean	-0.012	0.596	0.009	0.250	0.026	0.299
	Max	1.5621	3.861	0.790	2.047	1.179	4.559
INLA	Min	-0.264	0.694	-0.131	0.238	0.003	0.201
	Mean	0.003	1.624	-0.004	0.511	0.011	0.465
	Max	0.202	3.008	0.101	0.927	0.023	0.753

Table 18: Simulation study districts. Minimum (min), mean and maximum (max) value of the parameter estimates over the analyses of the 200 created datasets with the convolution model. Scenario 1 denotes independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1			2			3		
Software package		μ	τ^2	σ^2	μ	τ^2	σ^2	μ	τ^2	σ^2
CARBayes	Min	-0.269	0.543	0.062	-0.136	0.139	0.031	-0.049	0.174	0.025
	Mean	0.004	1.228	0.210	-0.006	0.4921	0.091	0.009	0.554	0.090
	Max	0.230	2.750	0.447	0.098	1.514	0.419	0.033	3.396	0.623
R2OpenBUGS	Min	-0.247	0.009	0.049	-0.129	0.009	0.021	0.003	0.010	0.008
	Mean	0.005	0.034	0.314	-0.004	0.084	0.087	0.012	0.304	0.033
	Max	0.2056	1.04	0.542	0.104	0.477	0.181	0.021	0.655	0.150
NIMBLE	Min	-0.265	0.008	0.021	-0.126	0.008	0.016	-4.708x10 ⁻⁴	0.009	0.008
	Mean	0.007	0.037	0.316	-0.004	0.081	0.089	0.011	0.304	0.035
	Max	0.259	1.226	0.540	0.096	0.492	0.198	0.031	0.664	0.166
R2BayesX-MCMC	Min	-0.263	0.005	0.107	-0.132	0.005	0.005	0.003	0.004	0.003
	Mean	0.003	0.057	0.326	-0.004	0.118	0.086	0.011	0.345	0.031
	Max	0.204	0.895	0.566	0.100	0.538	0.173	0.022	0.675	0.180
R2BayesX-REML	Min	-0.257	1.870x10 ⁻⁷	5.574x10 ⁻⁴	-0.128	1.240x10 ⁻⁶	5.212x10 ⁻⁴	-0.012	1.404x10 ⁻⁴	1.870x10 ⁻³
	Mean	0.007	0.058	0.337	-8.404x10 ⁻⁴	0.121	0.087	0.014	25.337	0.032
	Max	0.207	1.173	0.621	0.103	0.546	0.210	0.023	4929.460	1.565
RStan-MCMC	Min	-0.263	0.011	0.067	-0.132	0.010	0.013	0.003	0.011	0.007
	Mean	0.004	0.024	0.318	-0.004	0.084	0.087	0.011	0.318	0.029
	Max	0.203	0.650	0.534	0.100	0.479	0.179	0.023	0.657	0.146
RStan-VI	Min	-5.836	2.620x10 ⁻⁴	0.002	-6.022	0.002	0.002	-5.202	0.003	0.001
	Mean	-0.258	0.076	0.491	-0.256	0.041	0.726	-0.214	0.112	0.279
	Max	1.059	3.419	12.927	0.250	2.189	89.430	0.995	5.211	4.381
INLA	Min	-0.261	0.011	0.154	-0.131	0.010	0.011	0.003	0.011	0.007
	Mean	0.004	0.016	0.329	-0.003	0.075	0.093	0.012	0.322	0.033
	Max	0.204	0.114	0.541	0.100	0.558	0.190	0.023	0.678	0.181

Table 19: Simulation study districts. Minimum (min), mean and maximum (max) value of the overall absolute difference per simulated dataset over the analyses of the 200 created simulated runs. Scenario 1 is independence, scenario 2 moderate spatial dependence and scenario 3 strong spatial dependence.

Scenario		1		2		3	
Software package		Intrinsic	Convolution	Intrinsic	Convolution	Intrinsic	Convolution
CARBayes	Min	0.051	0.057	0.046	0.044	0.051	0.056
	Mean	0.118	0.105	0.060	0.063	0.073	0.074
	Max	0.281	0.323	0.134	0.184	0.249	0.194
R2OpenBUGS	Min	0.049	0.048	0.044	0.043	0.049	0.049
	Mean	0.063	0.063	0.054	0.054	0.060	0.060
	Max	0.080	0.079	0.071	0.070	0.074	0.073
NIMBLE	Min	0.048	0.048	0.044	0.043	0.049	0.049
	Mean	0.063	0.063	0.054	0.054	0.060	0.060
	Max	0.080	0.080	0.071	0.069	0.074	0.073
R2BayesX-MCMC	Min	0.049	0.048	0.043	0.043	0.050	0.048
	Mean	0.063	0.063	0.054	0.054	0.060	0.060
	Max	0.080	0.080	0.069	0.069	0.073	0.073
R2BayesX-REML	Min	0.049	0.048	0.043	0.043	0.050	0.049
	Mean	0.063	0.063	0.054	0.054	0.060	0.062
	Max	0.080	0.080	0.070	0.068	0.073	0.479
RStan-MCMC	Min	0.049	0.048	0.044	0.043	0.049	0.049
	Mean	0.063	0.063	0.054	0.054	0.060	0.060
	Max	0.080	0.079	0.071	0.070	0.074	0.074
Rstan-VI	Min	0.053	0.048	0.048	0.046	0.057	0.053
	Mean	0.367	0.224	0.222	0.096	0.347	3.700×10^4
	Max	24.534	16.803	1.370	1.498	21.513	7.399×10^6
INLA	Min	0.049	0.048	0.043	0.043	0.050	0.049
	Mean	0.063	0.063	0.054	0.054	0.060	0.060
	Max	0.081	0.080	0.071	0.069	0.074	0.073

Appendix G. Absolute bias per area.

Figure 10: Simulation study. Box plots of the absolute difference per area for a small, medium and large population in the independence situation. The top three rows correspond to the intrinsic model, the bottom three to the convolution model.

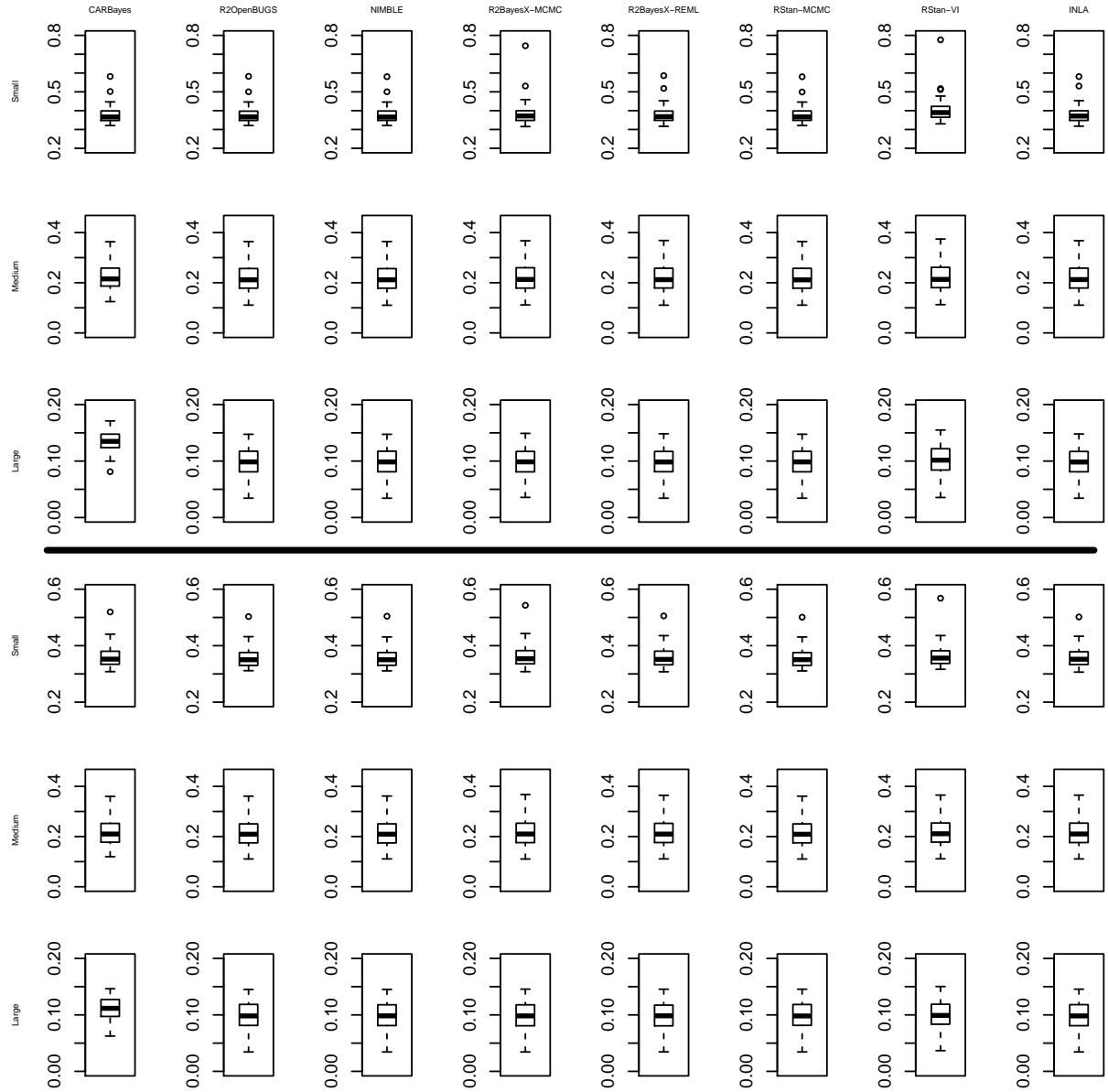


Figure 11: Simulation study. Box plots of the absolute difference per area for a small, medium and large population in the moderate dependence situation. The top three rows correspond to the intrinsic model, the bottom three to the convolution model.

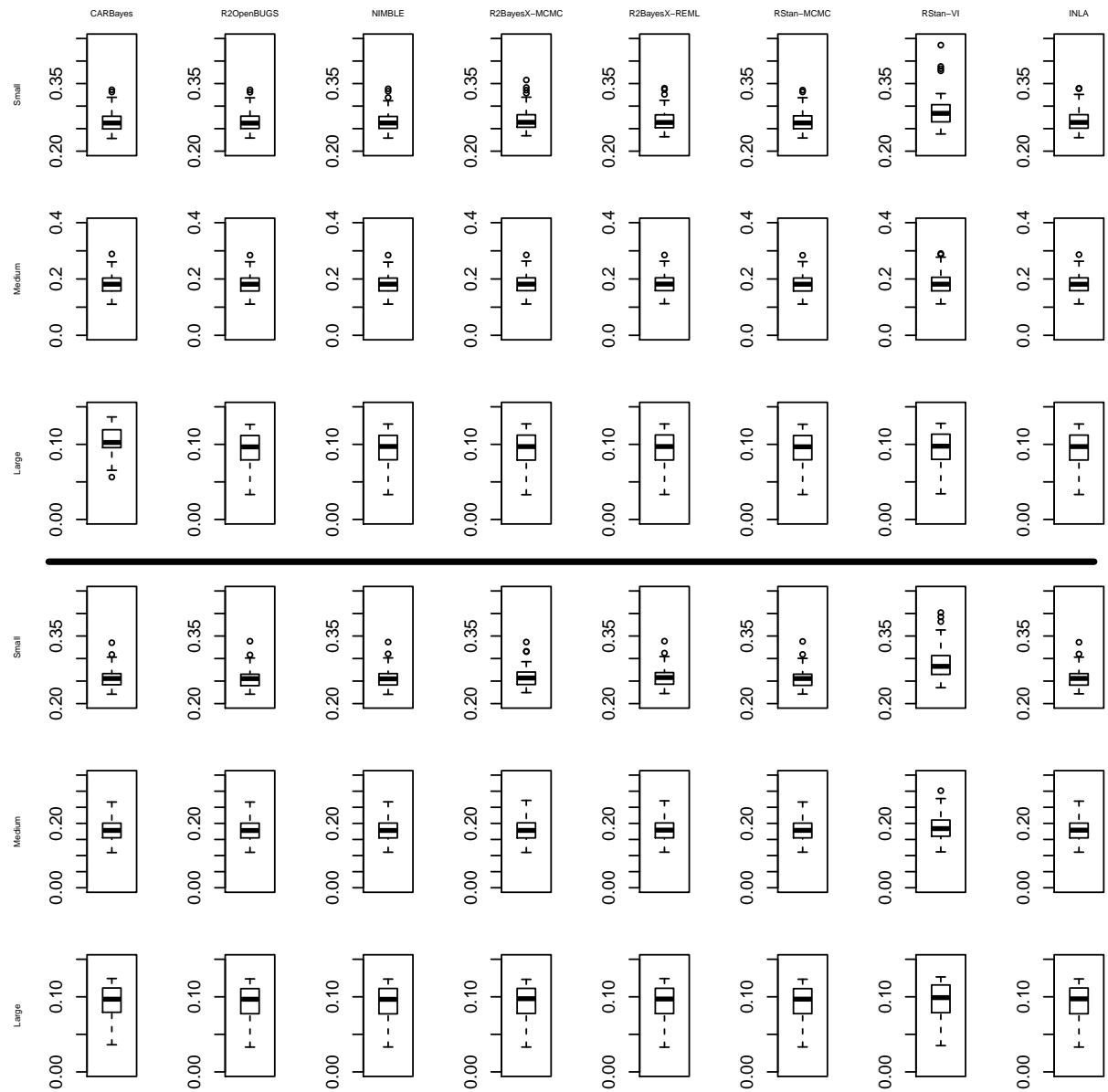


Figure 12: Simulation study. Box plots of the absolute difference per area for a small, medium and large population in the strong dependence situation. The top three rows correspond to the intrinsic model, the bottom three to the convolution model.

