Comparison of different software implementations for spatial disease mapping

Vranckx M., Neyens T., Faes C.

I-Biostat, Hasselt University, Diepenbeek, Belgium

October 2018

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Outline of the presentation

1. Methodology: models
2. Methodology: software packages
3. Data analysis
4. Simulation study
5. Conclusion
6. Reference

Methodology: models

- Y_k: The observed counts of (newly) diagnosed disease or mortality cases
- E_k: The expected number of disease or mortality cases

in an area k

Methodology: models

- Y_k: The observed counts of (newly) diagnosed disease or mortality cases
- E_k: The expected number of disease or mortality cases

in an area k

Methodology: models

Bayesian hierarchical models in disease mapping

$$Y_k|E_k, R_k \sim Poisson(E_k R_k), \quad k \in \{1, ..., n\}$$
$$ln(R_k) = \mu + \boldsymbol{x_k^T}\boldsymbol{\beta} + \phi_k \ .$$

- Y_k: The observed counts of (newly) diagnosed disease or mortality cases
- E_k: The expected number of disease or mortality cases
- R_k: The relative risk

in an area k

Methodology: models

Bayesian hierarchical models in disease mapping

$$ln(R_k) = \mu + \boldsymbol{x_k^T}\boldsymbol{\beta} + \phi_k$$

- Intrinsic model*

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

- Convolution model*

$$\phi_k = \theta_k + \psi_k$$

$$\theta_k \mid \sigma^2 \sim N(0, \sigma^2)$$

$$\psi_k | \boldsymbol{\psi_{-k}}, \tau^2 \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\psi_j, \frac{\tau^2}{n_k}\right)$$

* Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43, 159, https://doi.org/10.1007/BF00116466.

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	**R2OpenBUGS**	**Markov chain Monte Carlo**
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	**CARBayes**	**Markov chain Monte Carlo**
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	**rstan**	**Markov chain Monte Carlo**
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	**rstan**	**Variational inference**
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	**R2BayesX**	**Markov chain Monte Carlo**
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	**R2BayesX**	**Empirical Bayes**
INLA	INLA	Integrated nested Laplace approximation

Methodology: software packages

An overview of the implementations of the CAR models

Name	R Package	Estimation method
R2OpenBUGS	R2OpenBUGS	Markov chain Monte Carlo
CARBayes	CARBayes	Markov chain Monte Carlo
rstan-MCMC	rstan	Markov chain Monte Carlo
rstan-vi	rstan	Variational inference
R2BayesX-MCMC	R2BayesX	Markov chain Monte Carlo
R2BayesX-REML	R2BayesX	Empirical Bayes
INLA	**INLA**	**Integrated nested Laplace approximation**

An overview of priors and hyperpriors for the intrinsic model

Software package	CARBayes	R2OpenBUGS	R2BayesX	rstan	INLA
μ	$\sim N(0, 10000)$ Default	An improper (flat) prior User-specified	An improper (flat) prior Default	$\sim N(0, 102400)$ User-specified	$\sim N(0, 0)$ Default
τ^2	$\tau^2 \sim$ Inverse-Gamma(1,0.01) Default	$\frac{1}{\tau^2} \sim$ Gamma(1,0.01) User-specified	$\tau^2 \sim$ Inverse-Gamma(0.001, 0.001) Default	$\frac{1}{\tau^2} \sim$ Gamma(1,0.01) User-specified	Log$\left(\frac{1}{\tau^2}\right) \sim$ Log-gamma(1,5x10^{-5}) Default

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Methodology: software packages

Summary statistics

Estimates / Software package	Mean	Mode	Median	Quantiles	Standard deviation
CARBayes			×	×	
R2OpenBUGS	×		×	×	×
R2BayesX-MCMC	×		×	×	×
R2BayesX-REML		×			
rstan-MCMC	×		×	×	×
rstan-vi	×		×	×	×
INLA	×	×	×	×	×

Data analysis: data description

- Data about diabetics in children and young adults
- Belgium
- 2014
- InterMutualistisch Agentschap

Data analysis: parameter results of the intrinsic model

$$ln(R_k) = \mu + \phi_k \qquad \phi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\phi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)
CARBayes	-0.076 (0.013)	0.168 (0.022)
R2OpenBUGS	-0.076 (0.012)	0.166 (0.021)
R2BayesX-MCMC	-0.076 (0.013)	0.170 (0.023)
R2BayesX-REML	-0.064 (0.012)	0.165
rstan-MCMC	-0.076 (0.012)	0.167 (0.021)
rstan-vi	-0.075 (0.009)	0.231 (0.010)
INLA	-0.076 (0.012)	0.165 (0.021)

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Data analysis: parameter results of the intrinsic model

$$ln(R_k) = \mu + \phi_k \qquad \phi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\phi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)
CARBayes	-0.076 (0.013)	0.168 (0.022)
R2OpenBUGS	-0.076 (0.012)	0.166 (0.021)
R2BayesX-MCMC	-0.076 (0.013)	0.170 (0.023)
R2BayesX-REML	**-0.064** (0.012)	0.165
rstan-MCMC	-0.076 (0.012)	0.167 (0.021)
rstan-vi	-0.075 (0.009)	0.231 (0.010)
INLA	-0.076 (0.012)	0.165 (0.021)

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics
and statistical Bioinformatics

Data analysis: parameter results of the intrinsic model

$$ln(R_k) = \mu + \phi_k \qquad \phi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\phi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)
CARBayes	-0.076 (0.013)	0.168 (0.022)
R2OpenBUGS	-0.076 (0.012)	0.166 (0.021)
R2BayesX-MCMC	-0.076 (0.013)	0.170 (0.023)
R2BayesX-REML	-0.064 (0.012)	0.165
rstan-MCMC	-0.076 (0.012)	0.167 (0.021)
rstan-vi	**-0.075 (0.009)**	**0.231 (0.010)**
INLA	-0.076 (0.012)	0.165 (0.021)

Data analysis: parameter results of the convolution model

$$ln(R_k) = \mu + \phi_k$$

$$\phi_k = \theta_k + \psi_k \qquad \theta_k \sim N(0, \sigma^2) \qquad \psi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\psi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)	σ^2 (sd)
CARBayes	-0.0782 (0.013)	0.155 (0.021)	0.004 (0.002)
R2OpenBUGS	-0.077 (0.013)	0.152 (0.021)	0.004 (0.002)
R2BayesX-MCMC	-0.077 (0.013)	0.161 (0.022)	0.002 (0.002)
R2BayesX-REML	-0.066 (0.012)	0.184	1x10^{-4}
rstan-MCMC	-0.077 (0.012)	0.153 (0.022)	0.005 (0.002)
rstan-vi	-0.092 (0.020)	0.062 (0.011)	0.115 (0.002)
INLA	-0.077 (0.013)	0.154 (0.021)	0.004 (0.008)

Data analysis: parameter results of the convolution model

$$ln(R_k) = \mu + \phi_k$$

$$\phi_k = \theta_k + \psi_k \qquad \theta_k \sim N(0, \sigma^2) \qquad \psi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\psi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)	σ^2 (sd)
CARBayes	-0.0782 (0.013)	0.155 (0.021)	0.004 (0.002)
R2OpenBUGS	-0.077 (0.013)	0.152 (0.021)	0.004 (0.002)
R2BayesX-MCMC	-0.077 (0.013)	**0.161 (0.022)**	**0.002 (0.002)**
R2BayesX-REML	-0.066 (0.012)	**0.184**	**1x10^{-4}**
rstan-MCMC	-0.077 (0.012)	0.153 (0.022)	0.005 (0.002)
rstan-vi	-0.092 (0.020)	0.062 (0.011)	0.115 (0.002)
INLA	-0.077 (0.013)	0.154 (0.021)	0.004 (0.008)

Data analysis: parameter results of the convolution model

$$ln(R_k) = \mu + \phi_k$$

$$\phi_k = \theta_k + \psi_k \qquad \theta_k \sim N(0, \sigma^2) \qquad \psi_k \sim N\left(\frac{1}{n_k}\sum_{j\sim k}\psi_j, \frac{\tau^2}{n_k}\right)$$

Software package	μ (sd)	τ^2 (sd)	σ^2 (sd)
CARBayes	-0.0782 (0.013)	0.155 (0.021)	0.004 (0.002)
R2OpenBUGS	-0.077 (0.013)	0.152 (0.021)	0.004 (0.002)
R2BayesX-MCMC	-0.077 (0.013)	0.161 (0.022)	0.002 (0.002)
R2BayesX-REML	-0.066 (0.012)	0.184	1x10^{-4}
rstan-MCMC	-0.077 (0.012)	0.153 (0.022)	0.005 (0.002)
rstan-vi	**-0.092 (0.020)**	**0.062 (0.011)**	**0.115 (0.002)**
INLA	-0.077 (0.013)	0.154 (0.021)	0.004 (0.008)

Simulation study: design

- Model:

$$Y_k|E_k, R_k \sim Poisson(E_k R_k), \quad k \in \{1, ..., n\}$$
$$ln(R_k) = \mu + \phi_k \ .$$

- 3 scenarios $\rightarrow$ depending on ϕ_k:
  - spatial independence
  - moderate spatial dependence
  - strong spatial dependence
- For each scenario: 200 datasets created
- Analysed with the intrinsic and the convolution model

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	5	23
CARBayes	20	28
R2BayesX-MCMC	22	45
rstan-vi	19	23
rstan-MCMC	60	111
R2OpenBUGS	201	293
R2BayesX-REML	212	1880

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	**5**	**23**
CARBayes	20	28
R2BayesX-MCMC	22	45
rstan-vi	19	23
rstan-MCMC	60	111
R2OpenBUGS	201	293
R2BayesX-REML	212	1880

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	5	23
CARBayes	**20**	**28**
R2BayesX-MCMC	**22**	**45**
rstan-vi	**19**	**23**
rstan-MCMC	60	111
R2OpenBUGS	201	293
R2BayesX-REML	212	1880

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	5	23
CARBayes	20	28
R2BayesX-MCMC	22	45
rstan-vi	19	23
rstan-MCMC	**60**	**111**
R2OpenBUGS	201	293
R2BayesX-REML	212	1880

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	5	23
CARBayes	20	28
R2BayesX-MCMC	22	45
rstan-vi	19	23
rstan-MCMC	60	111
R2OpenBUGS	**201**	**293**
R2BayesX-REML	212	1880

Simulation study: computation time

Mean time (seconds) to run the model

Software package	Intrinsic	Convolution
INLA	5	23
CARBayes	20	28
R2BayesX-MCMC	22	45
rstan-vi	19	23
R2OpenBUGS	201	293
rstan-MCMC	60	111
R2BayesX-REML	**212**	**1880**

Simulation study: parameter estimates

- Similar estimates for R2OpenBUGS, rstan-MCMC and INLA
- R2BayesX: estimates between CARBayes and OpenBUGS
- rstan-vi: deviant parameter estimates

The parameter estimates of the convolution model

package		μ	τ^2	σ^2
CARBayes	Min	-0.471	0.601	0.046
	Mean	-0.044	0.879	0.095
	Max	0.311	1.247	0.155
R2OpenBUGS	Min	-0.466	0.779	0.006
	Mean	-0.039	1.028	0.014
	Max	0.316	1.328	0.049
rstan -MCMC	Min	-0.466	0.772	0.007
	Mean	-0.039	1.038	0.012
	Max	0.315	1.356	0.044

The parameter estimates of the convolution model

Software package		μ	τ^2	σ^2
CARBayes	Min	-0.471	0.601	0.046
	Mean	-0.044	0.879	0.095
	Max	0.311	1.247	**0.155**
R2OpenBUGS	Min	-0.466	0.779	0.006
	Mean	-0.039	1.028	0.014
	Max	0.316	1.328	**0.049**
rstan -MCMC	Min	-0.466	0.772	0.007
	Mean	-0.039	1.038	0.012
	Max	0.315	1.356	0.044

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Simulation study: relative risk estimates

$$\text{OAD}^{(l)} = \frac{1}{n}\sum_{k=1}^{n} |RR_k^{(l)} - \hat{RR}_k^{(l)}| \quad \text{simulation } l\text{, area } k$$

Software package			Software package			Software package		
R2Open-BUGS	Min	0.149	CARBayes	Min	0.153	rstan-vi	Min	0.169
	Mean	0.199		Mean	0.210		Mean	0.381
	Max	0.254		Max	0.274		Max	18.658
rstan -MCMC	Min	0.150	R2BayesX -MCMC	Min	0.152			
	Mean	0.199		Mean	0.202			
	Max	0.254		Max	0.256			
			R2BayesX -REML	Min	0.160			
				Mean	0.200			
				Max	0.256			
			INLA	Min	0.151			
				Mean	0.201			
				Max	0.256			

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Simulation study: relative risk estimates

$$\text{OAD}^{(l)} = \frac{1}{n}\sum_{k=1}^{n} |RR_k^{(l)} - \hat{RR}_k^{(l)}| \quad \text{simulation } l\text{, area } k$$

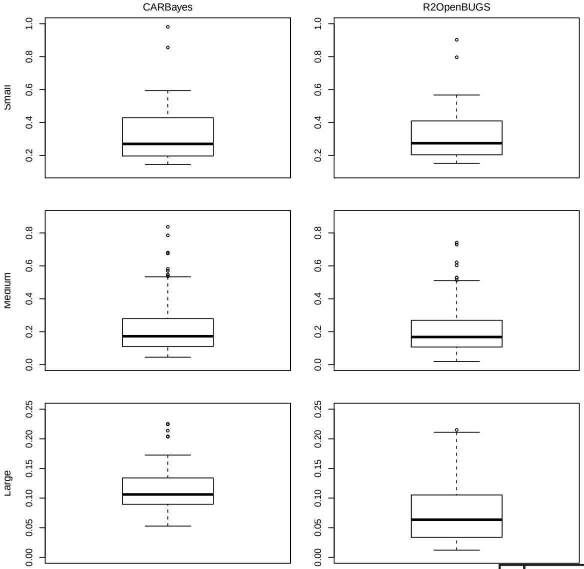
Software package			Software package			Software package		
R2Open-BUGS	Min	0.149	**CARBayes**	Min	0.153	rstan-vi	Min	0.169
	Mean	0.199		Mean	**0.210**		Mean	0.381
	Max	0.254		Max	**0.274**		Max	18.658
rstan -MCMC	Min	0.150	R2BayesX -MCMC	Min	0.152			
	Mean	0.199		Mean	0.202			
	Max	0.254		Max	0.256			
			R2BayesX -REML	Min	0.160			
				Mean	0.200			
				Max	0.256			
			INLA	Min	0.151			
				Mean	0.201			
				Max	0.256			

Simulation study: relative risk estimates

$$\text{OAD}^{(l)} = \frac{1}{n}\sum_{k=1}^{n} |RR_k^{(l)} - \hat{RR}_k^{(l)}| \quad \text{simulation } l \text{, area } k$$

Software package			Software package			Software package		
R2Open-BUGS	Min	0.149	CARBayes	Min	0.153	rstan-vi	Min	0.169
	Mean	0.199		Mean	0.210		Mean	0.381
	Max	0.254		Max	0.274		Max	18.658
rstan -MCMC	Min	0.150	R2BayesX -MCMC	Min	0.152			
	Mean	0.199		Mean	0.202			
	Max	0.254		Max	0.256			
			R2BayesX -REML	Min	0.160			
				Mean	0.200			
				Max	0.256			
			INLA	Min	0.151			
				Mean	0.201			
				Max	0.256			

Simulation study: relative risk estimates



Conclusion

$$\left.\begin{array}{l}\text{1. INLA}\\\text{2. rstan-MCMC}\\\text{3. OpenBUGS}\end{array}\right\}\rightarrow \text{Best performance}$$

$$\left.\begin{array}{l}\text{CARBayes}\\\text{R2BayesX-MCMC}\end{array}\right\}\rightarrow \text{Good performance,}$$

slight differences compared to 1, 2, 3

R2BayesX-REML $\rightarrow$ Deviant parameter estimates and highest computation time

rstan-vi $\rightarrow$ Deviant parameter and relative risk estimates

Reference

- Vranckx M., Neyens T., and Faes C. (2018) Comparison of different software implementations for spatial disease mapping. [Submitted]

- Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, 43, 159, https://doi.org/10.1007/BF00116466.
- Brezger, A., Kneib, T., and Lang, S. (2005) BayesX: Analyzing Bayesian structural additive regression models. *Journal of Statistical Software*, 14 (11), 122, 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.
- 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.

Reference

- 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.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- 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 (2017) Stan Modeling Language: Users Guide and Reference Manual, Version 2.17.0., URL http://mc-stan.org/.