



UHASSELT



Maastricht University

KNOWLEDGE IN ACTION

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

Prior specification of spatio-temporal models.

Marie Aimee Uwineza

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Chellafe ENSOY-MUSORO

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.



UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be

Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2017
2018



Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics

Masterthesis

Prior specification of spatio-temporal models.

Marie Aimee Uwineza

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Biostatistics

SUPERVISOR :

dr. Chellafe ENSOY-MUSORO

Acknowledgements

I would like to deeply express my appreciativeness to Almighty God in heaven, who has bestowed upon me this opportunity and ability of completing this research task to fulfil the Master of Sciences degree.

It is my grand chance of expressing my truthful thankfulness to my supervisors Dr. Chel-lafe Ensoy Musoro, for her continual encouragement and precious personal supervision for carrying out my thesis. Her unique assistance and guidance not only in this research but also in this master program studies have been influential contribution in enhancing my knowledge and skills.

It is my honor of utilizing this occasion to sincerely express my gratitude to Vlaamse Interuniversitaire Raad (VLIR) for the ultimate financial support to fulfil this program at Hasselt University and all of my appreciated teachers and staffs of Center for Statistics at Hasselt University, their hard work and commitment to the excellent education are foundation of my current achievements.

I am also thankful to all my family members especially my husband Munezero and daughter Umwiza Munezero for their wellbeing and endless encouragement based on understanding that have been the great motive behind the recent completion of this program.

I also owe my heartfelt gratefulness to my classmates, groupmates, friends and landlords for their assistance during my study at Hasselt University.

Finally, my deepest thanks are given to the University for its' constant inimitable excellent education, that made me a dynamic competent person.

Marie Aimée UWINEZA

Diepenbeek, Belgium, January 22, 2018

Abstract

Introduction: The population sample estimated parameters are key statistical point to make the population inference, predicting and forecasting the response of future subjects. In Bayesian Statistics, prior knowledge have a big impact in estimating parameters and predicting future values. The interest of this study was to investigate how the choice of prior affect the parameters' estimation for the spatio-temporal model which is constructed under the Bayesian Framework. The objective of this study is to investigate the effect of different prior specifications in a spatio-temporal model after simulating data of aborted cattle in different areas of Flanders.

Methodology: A spatio-temporal model which assumed a space-time variation was used to conduct a simulation for 100 different datasets on 306 areas of Flanders for a period of 36 Months. Each dataset has 11016 observations, the response variable was the number of aborted cattle and there were four covariate variables. The same model was then fitted 11 times on these datasets by applying different priors each time, different criterion have been used to propose a better prior among the used priors. All the analysis was done in Bayesian framework using Integrated Nested Laplace Approximation (INLA) package.

Results: Different combinations of priors have been used for hyperparameters components including the structured and unstructured spatial effect, first order autoregressive (AR1) precision and correlation and time space interaction. A model4 fitted using penalised Complexity prior for both spatial effects in combination with default prior for other terms was proposed among 8 (Model1-Model8) models as good in terms of minimum mean bias, minimum mean squared error and a good coverage probability. Additional 3 models (Model9-Model11) were fitted, where PC priors were used for other components of the model. Likewise model4 still being preferable than others. Averaged DIC for all models were compared, the DIC of model4 was smaller than anyone else with a difference larger 10.

Conclusion: It was concluded that Penalised Complexity (PC) prior in combination with default prior can be used while fitting a spatio-temporal model as they provided stability in hyperparameter estimates, and they can be used in real-world applied statistics as they are constructed by the user to avoid "cut and paste" prior from other related research articles.

Key words: *Spatio-temporal model, Prior specification, INLA*

Contents

| | |
|--|------------|
| Abstract | iii |
| 1 Introduction | 1 |
| 1.1 General Introduction | 1 |
| 1.2 Objectives of the study | 3 |
| 2 Methodology | 5 |
| 2.1 Important Concepts | 5 |
| 2.1.1 Poisson model for count data | 5 |
| 2.1.2 Bayesian method | 6 |
| 2.1.3 Hierarchical models | 6 |
| 2.1.4 Integrated Nested Laplace Approximation | 8 |
| 2.2 Simulation Model | 10 |
| 2.2.1 Parameter specification for simulation model | 11 |
| 2.2.2 Simulation set-up | 12 |
| 2.3 Hyper-prior specification | 13 |
| 2.3.1 Default prior (Gamma prior) | 13 |
| 2.3.2 Equivalent gamma prior | 14 |
| 2.3.3 Halfnormal prior | 14 |
| 2.3.4 Prior according to Bernardinelli | 14 |
| 2.3.5 Penalised Complexity (PC) Prior | 14 |
| 2.4 Prior selection criteria | 15 |
| 2.4.1 Unbiased estimator criteria | 16 |
| 2.4.2 Mean squared Error criteria | 16 |
| 2.4.3 Ratio of Variance criteria | 16 |
| 2.4.4 Coverage criteria | 17 |
| 2.4.5 Deviance information criterion | 17 |
| 2.5 Software | 18 |
| 3 Results | 19 |
| 3.1 Simulated Data | 19 |
| 3.2 Hyperpriors | 22 |
| 3.3 Results from different fitted models | 22 |

| | |
|-----------------------------|----|
| 4 Discussion and Conclusion | 29 |
| References | 33 |
| Appendix | 35 |

List of Tables

| | | |
|---|---|----|
| 1 | <i>True (hyper)parameters for the simulation model</i> | 13 |
| 2 | <i>Comparison of estimated hyperparameters with true hyperparameters by Bias, MSE, coverage and ratio of variance Criterion</i> | 23 |
| 3 | <i>Comparison of estimated hyperparameters with true hyperparameters by Bias, MSE, coverage and ratio of variance Criterion</i> | 26 |
| 4 | <i>Deviance Information Criterion values for all fitted model</i> | 27 |
| 5 | <i>Comparison of estimated parameters with true parameters by Bias, MSE, coverage and ratio of variance Criterion</i> | 35 |
| 6 | <i>Comparison of estimated parameters with true parameters by Bias, MSE, coverage and ratio of variance Criterion</i> | 36 |

List of Figures

| | | |
|---|---|----|
| 1 | <i>Map showing regions of Flanders Belgium</i> | 12 |
| 2 | <i>Map of cattle livestock in Flanders</i> | 19 |
| 3 | <i>Map showing cattle abortions in Flanders for the first simulation.</i> | 20 |
| 4 | <i>Map showing cattle abortions in Flanders for the last simulations.</i> | 20 |
| 5 | <i>Evolution of cattle abortion over time for 2 randomly selected areas</i> | 21 |
| 6 | <i>PC prior density for the mixing parameter ϕ</i> | 37 |

List of Acronyms

| | |
|------|---|
| AR1 | First-order Autoregressive |
| BHM | Bayesian Hierarchical Model |
| BYM | Besag York Mollié |
| DIC | Deviance Information Criteria |
| INLA | Integrated Nested Laplace Approximation |
| MCMC | Markov Chain Monte Carlo |
| PC | Penalised Complexity |

1 Introduction

This chapter gives the general introduction of the study and the motivation for carrying out this specific study.

1.1 General Introduction

The population sample estimated parameters are key statistical point to make the population inference. Inference helps not only in estimating parameters, but also in predicting and forecasting the response of future subjects, however the estimation methods vary depending on the data type as well as model type. Statistical methods for parameter estimation can be broadly classified into the frequentist, likelihood and Bayesian method. In this study, Bayesian method is used, specifically focusing in the prior specification of spatio-temporal model.

Spatio-temporal models are models used to analyze data collected over multiple regions repeatedly and over several periods of time. These models describe space and time variation of disease risk by considering the pattern of disease counts and highlight regions having unusual incidence levels, time trends or both [1]. In order to account for the various sources of variation and often hierarchical or clustered nature of the data, many spatio-temporal models are constructed under the Bayesian framework [2].

Bayesian Statistics is a field of statistics where inference about the true parameters of interest and hypothesis testing are performed by combining data at hand and the beliefs accepted before observing the data. These beliefs are what we call the prior knowledge which is an important component in Bayesian statistics inference. These prior knowledge are obtained from previous studies or from experts' opinions, and should be specified independent of the collected data. These are then combined with the likelihood or observed data to update information on the parameter and come up with a revised probability associated with the parameter, called the posterior probability [3].

These priors can be classified into:

- **Conjugate priors:** If, given any likelihood, the posterior and prior distributions belong to the same family, then we say that the prior and posterior distributions are conjugate to each other. This type of prior is then called a conjugate prior. For example combining a binomial likelihood with a beta prior results into a beta posterior. Then we conclude that our beta prior is a conjugate prior [4].
- **Non informative priors:** These are priors that provide no specific information about the variable of interest. They are some times called vague or flat priors [4]. Ancient studies used the Bayes postulate to construct non informative priors. This postulate states that: “when nothing is known about θ in advance, let the prior $\pi(\theta)$ be a uniform distribution” [5]. Since minimum information is given by these priors, the variance becomes large, leading to a small precision, hence less contribution to the posterior distribution. This means the likelihood dominates the prior in this case.
- **Informative priors:** These are priors based on either historical data or expert knowledge. These priors give precise information about the variables of interest. Because of the value of information given by these priors, they end up dominating the likelihood.

The priors chosen while fitting a model affect significantly the posterior distribution [6], specification of priors of unknown parameters in the model has to be done carefully and at an appropriate scale, to avoid any source of wrong posterior distribution [7].

Priors knowledge have a big impact in estimating parameters and predicting future values, they can be used to solve a problem of multicollinearity, convergence issues, parameter constraints as well as contribute in variables selection [3]. It was shown that prior specification is most suitable and important in cases where data are sparse [8]. Prior specification has been used to specify the variance component related to random effects of a spatio-temporal model like the Besag-York-Mollié model [6].

When fitting Bayesian models or spatio-temporal model, researchers randomly choose or assume different priors distribution of parameter estimation without knowing which can lead to incorrect posterior distribution of estimated parameters, and affect future prediction.

This study is intended to come up with a good proposal priors which can be used when fitting spatio-temporal model and guarantee the researchers to get a correct posterior distribution and hence accuracy prediction for the future values. The interest of this study is to investigate how the choice of priors affect the parameter's estimates in a spatio-temporal model.

1.2 Objectives of the study

This section contains the objectives to fulfill in this study

General objective

The general objective of this study is to investigate the effect of different prior specifications in a spatio-temporal model.

Specific objectives

- To simulate 100 different datasets of aborted cattle in different areas of Flanders.
- To fit a spatio-temporal model using different priors on those simulated data.
- To identify effects of different priors on spatio-temporal model.

This thesis is organized as follows: chapter 2 gives a brief overview of the used methodology. Here we also introduce the nature of the simulated data. In chapter 3, we illustrate the application of the methodology to the simulated data, we give the interpretation of our results from the models and data they are applied to. Lastly in chapter 4, we end up with discussion and conclusion and some recommendations.

2 Methodology

This chapter provides a brief description of the study region, data used in the study, theories about methods of analysis, and other technical aspects of the study.

2.1 Important Concepts

This section covers statistical theories about all methods applied in the analysis to fulfill the objectives of the study.

2.1.1 Poisson model for count data

A Poisson model is the most popular model for count data, especially for small area studies. It is applicable when there is a relatively low count of a disease (disease is rare) and the population is relatively large in each small area [4].

Suppose we have a study region T in disease mapping study, and we split it into n non-intersecting areas, and denote the set of observed disease counts as $y = (y_1, \dots, y_n)$ from the events of disease occur (m_1, \dots, m_n) . These can be assumed as independent realizations of a random sample or realizations of the disease process and observed disease count in each m event is considered as a spatial unit.

Assume a variable y follows a Poisson distribution with mean μ and independent realizations; $y_i \sim Poisson(\mu_i)$, then the Probability Density Function (pdf) of y is given by;

$$f(y_i|\mu_i) = \frac{\mu_i^{y_i} \exp(-\mu_i)}{y_i!}. \quad (1)$$

The likelihood function of Equation (1) is given by;

$$L(Y|\mu) = \frac{\prod_{i=1}^m \mu_i^{y_i} \exp(-\mu_i)}{y_i!}. \quad (2)$$

In Equation (2), μ_i represents the average number of disease cases occurring in region i . Since a Poisson distribution has mean = variance = μ [9], the only parameter needed to determine the probability of disease counts in this case is μ .

2.1.2 Bayesian method

Bayesian method is an estimation method that uses Bayes' theorem. The pdf of a Bayesian distribution is given by;

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}. \quad (3)$$

In Equation (3), $p(B)$ is the prior probability which does not depend on the observed data. It indicates belief about the presence of the disease before any observation is made. Event $p(B|A)$ is the posterior probability, obtained from joining the prior information with observed data [3]. Lastly are $p(A|B)$ and $p(A)$ which represent the likelihood functions obtained from the observed data. In cases where a parameter belonging to the prior distribution is not fixed, it is allocated a random distribution. This type of parameter is called a hyperparameter and the prior to which the parameter belongs is called a hyperprior [4].

As shown in Equation (3), the posterior distribution is proportional to the likelihood functions, multiplied by the prior. This is expressed as;

$$p(\theta|y) \propto p(y|\theta)p(\theta),$$

and can be calculated using

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \quad (4)$$

2.1.3 Hierarchical models

A Bayesian hierarchical model is a Bayesian Statistical model used to model data that exhibit a hierarchical structure. These types of data are some times considered as clusters. It is shown that the hierarchical structure in the data allows to account for similarities based on the neighbourhood for area-level data or on distance for point-reference data [10]. For any two areas to be considered as neighbours, they need to share a common boundary with each other. This means that geographically close areas (say, municipalities) have similar incidence rates of the disease. Likewise, they can be considered as clusters where measurements are taken repeatedly over time on the same subject. In this case, we take the area to be the subject, meaning that correlation between measurements should be taken into account [3].

According to [11], this hierarchical structure can be described by two levels. Firstly, we consider the individual level to show the disease status, then the area level to specify the number of cases in one selected area. For the second level, all units in the first level are nested within the second level representing a cluster that contains cases belonging to different areas within the same region.

a hierarchical Bayes model the priors are allocated a given distribution. The distribution of the posteriors in this type of models is written as:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi}). \quad (5)$$

joint posterior distribution is calculated by evaluating double integration of Equation (5) with respect to both parameters and hyperparameters of the model:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi})}{\iint p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi})d\boldsymbol{\theta}d\boldsymbol{\phi}}. \quad (6)$$

In Equation (6), $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y})$ is a posterior distribution, $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\phi})$ is the likelihood and $p(\boldsymbol{\phi})$ is the hyperprior.

There are two methods commonly used to approximate the posterior distribution; Markov Chain Monte Carlo (MCMC) sampling method and the Integrated Nested Laplace Approximation (INLA). The process of MCMC approximation is described as one that involves choosing a parameter as the starting point, say x_0 , and generating a series of other parameters which should converge to give an estimate of the posterior distribution [12]. However, in this study, we focus on the INLA approach to compute the posterior distribution of parameters and hyperparameters.

2.1.4 Integrated Nested Laplace Approximation

The work covered in this section follows studies from Rue *et al.*(2009) [13].

INLA is a computationally efficient method used as an alternative to MCMC. It is used to compute the posterior marginals of all parameters of interest. It is commonly applied when dealing with latent Gaussian models. The advantages of INLA include but are not limited to the fact that it is easy to use and provides precise parameter $p(\boldsymbol{\theta}|\mathbf{y})$ and hyperparameter $p(\boldsymbol{\phi}|\mathbf{y})$ estimates in short computational time periods.

The posterior marginal densities for subject i in relation to parameters and hyperparameters can be written respectively as:

$$p(\theta_i|\mathbf{y}) = \int p(\theta_i|\boldsymbol{\phi}, \mathbf{y})p(\boldsymbol{\phi}|\mathbf{y})d\boldsymbol{\phi}, \quad (7)$$

$$p(\phi_j|\mathbf{y}) = \int p(\boldsymbol{\phi}|\mathbf{y})d\boldsymbol{\phi}_{-j}. \quad (8)$$

Incorporation of these two densities in INLA has been done to compute their nested approximations [13].

To approximate $p(\boldsymbol{\phi}_j|\mathbf{y})$ in Equation (8), posterior marginals for $\boldsymbol{\phi}_j$ can be obtained directly from $\tilde{p}(\boldsymbol{\phi}|\mathbf{y})$ using numerical integration. It also suggested that the approximate marginal posterior density for hyperparameter $\boldsymbol{\phi}$ can be obtained using Laplace approximation [14], given by:

$$\tilde{p}(\boldsymbol{\phi}|\mathbf{y}) \propto \frac{p(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y})}{\tilde{p}_G(\boldsymbol{\theta}|\boldsymbol{\phi}, \mathbf{y})} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*(\boldsymbol{\phi})} \quad (9)$$

In Equation (9), $\tilde{p}_G(\boldsymbol{\theta}|\boldsymbol{\phi}, \mathbf{y})$ is the Gaussian approximation for the parameter $\boldsymbol{\theta}$, while $\boldsymbol{\theta}^*(\boldsymbol{\phi})$ is the mode for a given hyperparameter $\boldsymbol{\phi}$. The approximation $\tilde{p}(\boldsymbol{\phi}|\mathbf{y})$ of the marginal posterior for $\boldsymbol{\phi}$ can also be computed in three steps:

- Using Quasi-Newton method, we find the mode $\boldsymbol{\phi}^*$ of $\tilde{p}(\boldsymbol{\phi}|\mathbf{y})$ by optimizing $\log \tilde{p}(\boldsymbol{\phi}|\mathbf{y})$ with respect to $\boldsymbol{\phi}$. This Quasi-Newton method works by generating an approximation to the second derivative of $\log \tilde{p}(\boldsymbol{\phi}|\mathbf{y})$. This is done by calculating the difference between successive gradient vectors, which can be approximated using definite differ-

ences.

- When configuring the mode ϕ^* , we compute the negative Hessian matrix $\mathbf{H} > 0$, using differences.
- Lastly, we explore $\log \tilde{p}(\phi|\mathbf{y})$ using the z - parameterization.

To approximate $p(\theta_i|\phi, \mathbf{y})$ in Equation (7), we use three approaches;

- Gaussian approximation: This is the easiest method used to approximate $p(\theta_i|\phi, \mathbf{y})$. The Gaussian (often called the normal) approximation of the distribution of parameter θ is given by:

$$p_G(\theta_i|\phi, \mathbf{y}) = N(\theta_i; \mu_i(\phi), \sigma_i^2(\phi)) \quad (10)$$

In Equation(10), $\mu_i(\phi)$ is the mean and $\sigma_i^2(\phi)$ is marginal variance. With the assumption of normality, both mean and marginal variance can be computed efficiently using recursions as expressed in Section 2.1 of [13]. Results from the Gaussian approximation method are often considered correct, however, some errors arise due to different factors, mainly, location and lack of skewness [13]. As an improvement to the Gaussian approximation, we can compute the Laplace approximation.

- Laplace approximation is given by

$$\tilde{p}_{LA}(\theta_i|\phi, \mathbf{y}) \propto \frac{p(\theta, \phi, \mathbf{y})}{\tilde{p}_{GG}(\theta_{-i}|\theta_i, \phi, \mathbf{y})} |_{\theta_{-i}=\theta_{-i}^*} (\theta_i, \phi) \quad (11)$$

In Equation (11), the denominator \tilde{p}_{GG} is the Gaussian approximation to $\theta_{-i}|\theta_i, \phi, \mathbf{y}$. The precision matrix needed to calculate \tilde{p}_{GG} depends on both θ_i and ϕ . This means that \tilde{p}_{GG} has to be computed for every value of θ_i and ϕ , which becomes both cumbersome and time consuming since we need to factorize the precision matrix n times. Two main approaches are employed to solve this problem which change the Laplace approximation into:

$$\tilde{p}_{SLA}(\theta_i|\phi, \mathbf{y}) \propto N(\theta_i; \mu_i(\phi), \sigma_i^2(\phi))\exp(\text{cubic spline}(\theta_i)). \quad (12)$$

While using the modification in Equation (12), we fit the cubic spline to the difference of the log-density of $\tilde{p}_{SLA}(\theta_i|\boldsymbol{\phi}, \mathbf{y})$ and $\tilde{p}_G(\theta_i, \boldsymbol{\phi}, \mathbf{y})$ at selected points.

- We can also use the Simplified Laplace Approximation, which was derived through a series expansion of the Laplace approximation $\tilde{p}_{LA}(\theta_i|\boldsymbol{\phi}, \mathbf{y})$ around $\theta_i = \mu_i(\boldsymbol{\phi})$. This method helps solve the Gaussian problems of location and skewness, and is sufficient enough to give correct posterior marginals when dealing with models related to observations, like, Poisson and Binomial.

The approximation of the marginal posterior for parameter θ_i can also be calculated using a finite sum as;

$$\tilde{p}(\theta_i|\mathbf{y}) = \sum_k \tilde{p}(\theta_i|\boldsymbol{\phi}_k, \mathbf{y})\tilde{p}(\boldsymbol{\phi}_k|\mathbf{y})\Delta_k \quad (13)$$

In Equation (13), the finite sum is calculated over values of hyperparameters $\boldsymbol{\phi}$, using numerical integration and applying suitable area weights Δ_k .

2.2 Simulation Model

Let y_{it} represent the number of cattle aborted in area i ($i = 1, \dots, I$) at months t ($t = 1, \dots, T$), and let m_i denote the cattle population per area i . Since we are dealing with count data, a Poisson model for y_{it} is assumed.

In this study, a spatio-temporal model which assumed a Poisson distribution has been used. It was formulated in hierarchical manner, by specifying data model $p(y_{it}|\theta)$, the parameter model $p(\theta|\tau)$ as well as the hyperparameter model $p(\tau)$ [15].

Hence the Poisson model is then given by

$$y_{it} \sim \text{Poi}(m_i, \mu_{it}), i = 1, \dots, I, t = 1, \dots, T.$$

where μ_{it} is the expected value of cattle abortion.

In the simulation model, the log-link has been used and the expectation μ_{it} has been broken down into; a fixed intercept β_0 , a temporal term β_t , a spatial term c_i and a spatio-temporal interaction term d_{it} .

The simulated model is then specified as:

$$\log(\mu_{it}) = \eta_{it} = \beta_0 + \beta_t + c_i + d_{it}, \quad i = 1, \dots, I, \quad t = 1, \dots, T. \quad (14)$$

2.2.1 Parameter specification for simulation model

The simulation of model (14) was done by specifying the distributions of some terms in the model.

The temporal term (β_t) is decomposed into a linear and seasonal term given below, a parametric model was assumed for this temporal term.

$$\beta_t = \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{12}\right) + \beta_3 \cos\left(\frac{2\pi t}{12}\right).$$

Spatial component (c_i) is decomposed into two components namely, the structured spatial effect indicating spatial autocorrelation and the unstructured spatial effect indicating residual heterogeneity. This spatial component assumes the Besag York Mollié model (BYM) [16].

$$c_i = \xi_i + \zeta_i.$$

Data for the structured spatial effect ξ_i are given while the unstructured spatial effect ζ_i is assumed to be an independent mean-zero normally distributed variable with unknown variance; τ_ζ^{-1} , i.e $\zeta_i \sim N(0, \tau_\zeta^{-1})$.

The space-time interaction term d_{it} is decomposed into two components:

$$d_{it} = \delta_{1t} + \delta_{2it}.$$

The first term δ_{1t} is a random time process which assumed to follow a first order autoregressive model (AR1), meaning that the current value is based on the immediately preceding value. In simulation model, its normal distribution was specified as follows:

$$\delta_{1t} | \delta_1, s < t \sim N(\rho_\delta \delta_{1,t-1}, \tau_{\delta_1}^{-1}),$$

The second term δ_{2it} is the unstructured space-time interaction which assumed to follow an independently and identically distributed model (IID), In simulation model, its normal distribution was specified as follows $\delta_{2it} \sim N(0, \tau_{\delta_2}^{-1})$.

In this model, a sum-to-zero constraint was postulated to ensure identifiability of the parameters [1].

2.2.2 Simulation set-up

This section briefly describes datasets and variables used in this thesis. We conduct a simulation, where 100 different datasets of aborted cattle are simulated for each of the 306 areas of Flanders, Belgium (Figure 1), over a period of 36 months (3 years).

Flanders is situated in north of Belgium, it is a Dutch-speaking region. Its area is 13,522 km^2 representing 44.3 % of Belgium. Flanders is composed by 5 provinces namely: West Flanders, East Flanders, Antwerp, Limburg and Flemish Brabant, these 5 provinces are split into 306 municipalities.

Map of Flanders, Belgium



Figure 1: *Map showing regions of Flanders Belgium*

We took the response variable to be the number of aborted cattle, and covariate variables as: ID (indicating the structured and unstructured effects of each area), time (representing 36 Months), time2 (indicating that the number of aborted cattle are correlated over time), and ID2 (representing the unstructured space and time interaction).

Table 1 shows the true (hyper)parameters used for simulation.

Table 1: *True (hyper)parameters for the simulation model*

| Description | Parameter | Estimate |
|-------------------|-------------------------------|----------|
| Intercept | | -7.9310 |
| Time effect | Linear | -0.0090 |
| | Sine | -0.0572 |
| | Cosine | -0.0805 |
| Spatial Effect | Structured effect Precision | 1.3828 |
| | Unstructured effect Precision | 27.9474 |
| Space*time Effect | AR1 Correlation | 0.7857 |
| | AR1 Precision | 5.2799 |
| | Unstructured effect Precision | 6.0696 |

2.3 Hyper-prior specification

Choosing prior distributions is very important because the choice of the prior distributions of parameters can affect the posterior significantly [4]. Hence, different prior specifications of the model hyper parameters are specified for the spatial term, autocorrelation term and spatio-temporal interaction term with the assumption of a Poisson distribution.

2.3.1 Default prior (Gamma prior)

A gamma prior on log scale is used as a default prior for INLA [10]. The pdf of a gamma distribution is given by;

$$\pi(\tau) = \frac{b^a}{\Gamma(a)} \tau^{a-1} \exp(-b\tau),$$

where $\tau > 0$, is the parameter of interest, a is the slope parameter and b is the inverse-scale parameter. Both a and b must be positive definite. The mean value of τ is given by $\frac{a}{b}$ and its variance is $\frac{a}{b^2}$. These are characteristics of a $\text{Gamma}(a, b)$ distribution. Furthermore, $\text{Gamma}(a, b)$ is the distribution of variance θ , if $\theta = \log(\tau)$ and τ follows a $\text{Gamma}(a, b)$ distribution.

In model fitting, for both the structured (ξ_i) and the unstructured (ζ_i) spatial effect precisions, default Gamma(1, 0.0005) was used. The same default Gamma(1, 0.0005) was used for both random time process δ_{1t} and unstructured space-time interaction δ_{2it} in fitting the first model [10].

2.3.2 Equivalent gamma prior

For both the structured (ξ) and the unstructured spatial effect precision (ζ_i), Equivalent gamma prior Gamma(0.001, 0.001) on log scale was used.

2.3.3 Halfnormal prior

For both random time process δ_{1t} and unstructured space and time interaction δ_{2it} , halfnormal on log scale for standard deviation instead of the precision was used, it is given by $\log \text{tnormal}(0, 0.001)$.

2.3.4 Prior according to Bernardinelli

The precisions of the structured (τ_ξ) and the unstructured spatial effect (τ_ζ) have been chosen as Gamma(1, 0.023) and Gamma(1, 0.01) respectively. This is prior(B) according to Bernardinelli. Prior(A) was given as Gamma(1, 0.00025) and Gamma(1, 0.0006) for precision of the structured (τ_ξ) and the unstructured spatial effect (τ_ζ) respectively. Likewise, Prior(C) was given as Gamma(1, 0.125) and Gamma(1, 0.286) for precision of the structured (τ_ξ) and the unstructured spatial effect (τ_ζ) respectively [8].

A comparison of these three priors, shows that prior(A) has the largest coefficient of variation, and the greater part of its mass, favouring less extreme geographical variation. Prior(C) is a strong prior, however it is quite concentrated and favours marked geographical variation. Prior(B) is the most appropriate, compared to prior(A) and prior(C). It is considered as better prior in terms of providing marginal and condition posterior distributions of τ_ξ and τ_ζ [8].

2.3.5 Penalised Complexity (PC) Prior

These are informative priors with a single parameter used to control the amount of flexibility allowed in the model. In this case, the user has the freedom to define their own priors.

These priors were developed to be use when dealing with realistically complex Bayesian models, for example, hierarchical models [17].

Penalised priors allow for the use of parsimonious models, where a simple model, called the base model is preferred. In cases where there is enough information, a more complex model can be applied but, the prior plays a role of penalizing deviations from the simple model. The prior is used to choose a simple model by measuring the increase in complexity between the base model (g) and a more flexible one (f). The measure of the increase in complexity between densities f and g is given by

$$KLD(f \parallel g) = \int f(x) \log\left(\frac{f(x)}{g(x)}\right) dx,$$

KLD (Kullback-Leibler Divergence) is a measure of the information lost when the base model g is used to approximate a more flexible model f .

To fit a model with PC priors, a type 2- Gumbel distribution was used as PC prior for spatial effect. The density of this distribution is given by:

$$\pi(\tau) = \frac{\lambda}{2} \tau^{-\frac{3}{2}} \exp(-\lambda \tau^{-\frac{1}{2}}), \tau > 0$$

where, τ is the precision parameter and λ is the magnitude of the penalty for deviating from the base model and higher values increase this penalty.

The parameters of this distribution are specified as follows: a scaling parameter $\alpha = 0.01$ was used and the spatial component U is set to be $\frac{0.2}{0.31}$. By plotting this PC priors distribution, it is found that the PC prior puts $\frac{2}{3}$ probability on mixing parameter $\phi < \frac{1}{2}$, this is to set up that a set (U, α) ensures that $\text{Prob}(\phi < U) = \alpha$. This probability determines the degree of penalisation. Those values were also used in model fitting to get the posterior concentrate around 1 [17].

The mixing parameter of the precision of spatial effect controls the marginal precision, the distance above can be rewritten as $d(\phi) = \sqrt{2KLD(\phi)}$.

2.4 Prior selection criteria

Different theoretical methods can be used to determine which model best suits the interest of the study at hand. Some of these methods are;

2.4.1 Unbiased estimator criteria

An estimator is said to be unbiased if the mean of its distribution is equal to the true parameter being estimated. Bias is defined as the difference between the expectation of an estimator and the true parameter. If the difference tends to 0 as $n \rightarrow \infty$, then the estimator is said to be an asymptotically unbiased estimator for the true parameter. Let $T_n = t(X_1, X_2, \dots, X_n)$ represent an estimator. T_n is called an unbiased estimator for θ if $E_\theta(T_n) = \theta$, and the quantity $bias(T_n) = E_\theta(T_n) - \theta$ is the bias of T_n . Among all unbiased estimators, one could choose an unbiased estimator with the smallest mean squared error [18].

2.4.2 Mean squared Error criteria

This is the measure of goodness or closeness of an estimator T_n of θ . To support this property of goodness measure of mean squared error, we could consider mean squared error as a measure of the spread of T_n values about θ . Mean squared error of an estimator T_n for the true parameter is given by

$$MSE_\theta(T_n) = E_\theta[(T_n - \theta)^2].$$

It is estimated by computing:

$$MSE_\theta(T_n) = Var_\theta(T_n) + (bias_\theta(T_n))^2.$$

If $MSE_\theta(T_n) \rightarrow 0$ for all θ , then T_n estimates θ correctly. This means that from the sample, one must be able to identify the true parameter value. While comparing estimators by looking at their respective mean squared error, we prefer one with the smallest mean squared error [18].

2.4.3 Ratio of Variance criteria

The variance of a random variable is a measure of its spread about its mean. The ratio of variances was computed as follows

$$\frac{var_{est}(E_\theta(T_n))}{var_{sim}(E_\theta(T_n))}.$$

The ratio closer to 1 is an indication of the goodness of the estimator T_n of θ .

The variance on numerator is the variance computed based on the 100 simulation results for each parameter.

$$var_{est}(E_{\theta}(T_n)) = \frac{1}{n-1} \sum_{i=1}^{100} ((E_{\theta}(T_{n_i}) - \overline{E_{\theta}(T_n)})^2$$

The variance on denominator is the averaged variances from the model

$$var_{sim}(E_{\theta}(T_n)) = \sum_{i=1}^{100} \frac{var(E_{\theta}(T_{n_i}))}{100}$$

2.4.4 Coverage criteria

Coverage probability can be define as the probability that a process for constructing random regions will produce an interval containing or covering the true parameter [19]. Coverage probabilities have been calculated for each model and each parameter estimate. The calculation was done in three steps below:

- Each model has been fitted for 100 simulated samples
- 2.5% Lower confidence bound and 97.5% upper confidence bound have been obtained from step 1 for each model, each parameter and each sample
- Proportion of samples or simulation for which the true known population parameter is contained in the confidence interval has been computed for each model and each parameter. Hence that proportion is an estimate for coverage probability for the confidence interval [20].

2.4.5 Deviance information criterion

In Bayesian framework, the deviance information criterion is used as a tool to select the best fitted model among all candidate models. It is computed as follows:

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D$$

and,

$$p_D = \overline{D(\theta)} - D(\bar{\theta}).$$

where, p_D is Bayesian measure of complexity, $\overline{D(\theta)}$ is the posterior mean of bayesian deviance, $D(\bar{\theta})$ is the posterior mean of $-2 \log L$.

The rule of thumb for using DIC in model selection is that a difference in DIC of more than 10 takes out the model with the higher DIC, and a difference of less than 5 implies no conclusion among compared models [3].

2.5 Software

All analyses (models fitting and plots) carried out in this thesis are done in R software with emphasis on the INLA package.

3 Results

This chapter contains the main analysis carried out on the simulated data and the results obtained by applying the methodology discussed in Chapter 2 on the 100 simulated datasets.

3.1 Simulated Data

In this section, we present initial graphical representations of the data. This fundamental step is carried out in order to gain a better understanding of the data and its nature. Preliminary graphical representation are used to study the association between the response variable and the set of explanatory variables.

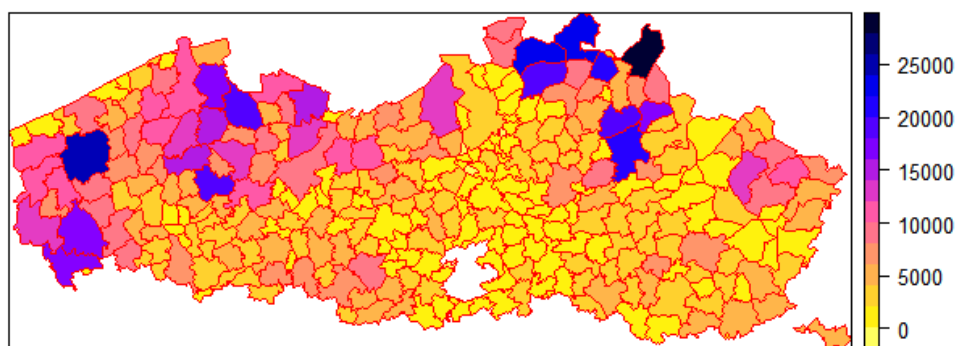


Figure 2: *Map of cattle livestock in Flanders*

Figure 2 represents the distribution of cattle livestock in Flanders. It can be observed that areas in the West and North East of Flanders have a high number of cattle livestock compared to Central and South East regions.

Figures 3 and 4 show the distribution of aborted cattle in all 306 municipalities of Flanders, by considering different simulated datasets and different time points.

Both figures show that there are many areas with very low numbers of cattle abortions. In addition, they show that there is evidence of clustering across municipalities, this means

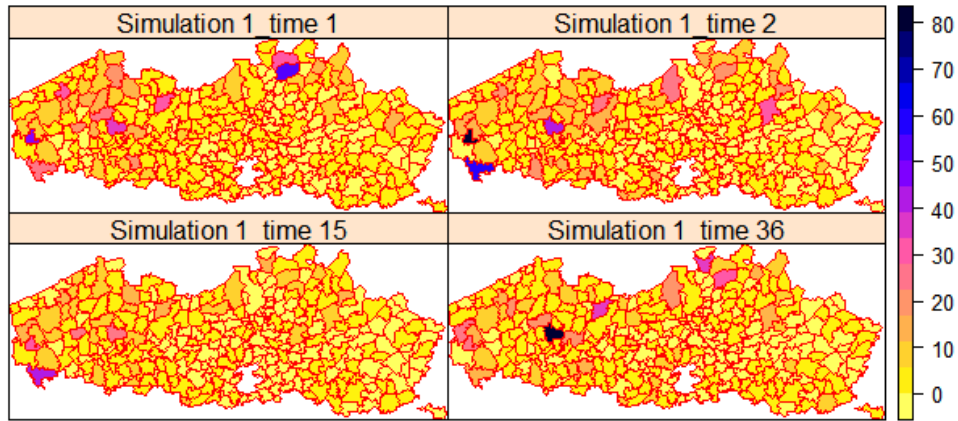


Figure 3: Map showing cattle abortions in Flanders for the first simulation.

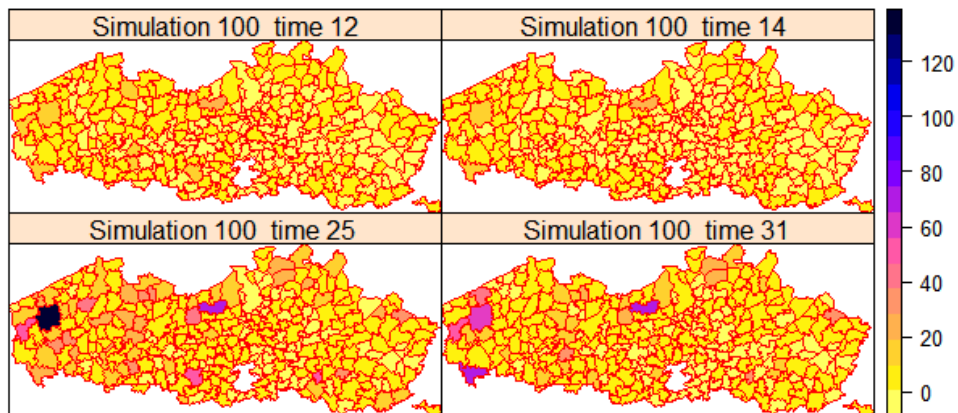


Figure 4: Map showing cattle abortions in Flanders for the last simulations.

that municipalities with high numbers of cattle abortion are surrounded by others with high numbers of cattle abortion. Generally, it can be seen that cattle abortion is spread in the entire Flanders region.

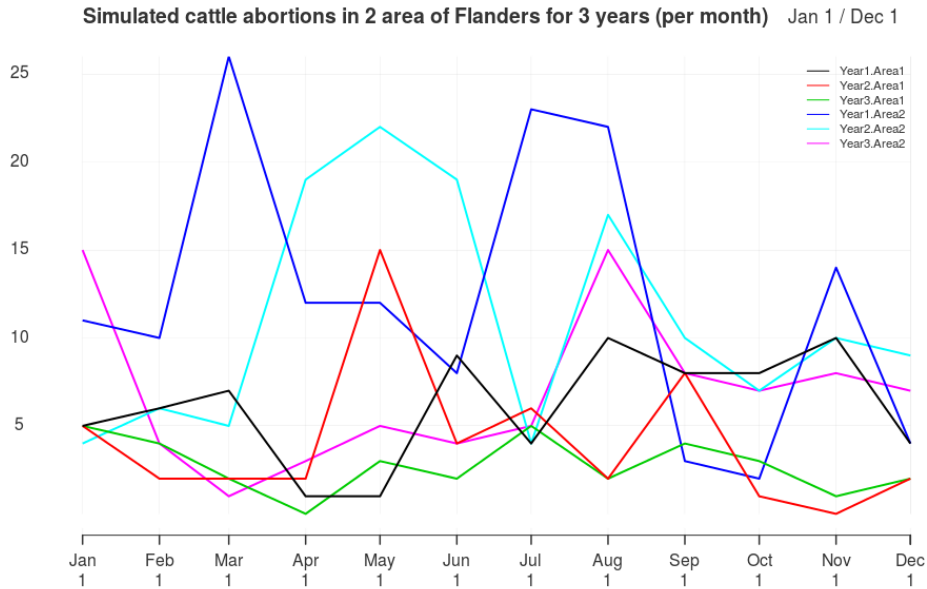


Figure 5: *Evolution of cattle abortion over time for 2 randomly selected areas*

Figure 5 shows individual profiles for two randomly selected areas, Y axis is number of aborted cattle while X axis represents the Months. It represents evolution over time of aborted cattle during three years of study. It can be seen that within the same area, the number of aborted cattle varies depending on years. For instance, in the first area, the evolution of the third year is the lowest compared to the first and second year. Likewise there is a slight change between year 2 and first year. However, the minimum and maximum values are observed in November and May of the second year, respectively.

Between areas, it is clear that the number of aborted cattle is higher in area 2 compared to area 1. Likewise by comparing each year in area 1 with corresponding year in area 2, for all cases area 2 has the highest number of aborted cattle.

3.2 Hyperpriors

To fulfill the objectives of study, different types of priors discussed in section 2.4 have been applied. Default prior, equivalent gamma prior and prior(B) of Bernardinelli follow a gamma distribution, as it is known that in disease mapping context, the most assumed distribution for the parameters in a Poisson model is the gamma distribution [4]. Likewise, a halfnormal prior follows a halfnormal distribution and penalized complexity prior assumes a type-2 Gumbel distribution. These priors have been used to see how they are differently affecting hyperparameter estimates. Some of them are informative priors while others are non informative. Informative priors are halfnormal and penalized complexity prior, while non informative priors are default prior, equivalent prior and prior(B) of Bernardinelli.

3.3 Results from different fitted models

During the analysis, 8 models were fitted on 100 simulated datasets by considering different priors. This means that for each model, one type of prior is used for the structured and unstructured spatial effects. Another type of prior is used for random time process and the unstructured space-time interaction.

Table 2: Comparison of estimated hyperparameters with true hyperparameters by Bias, MSE, coverage and ratio of variance Criterion

| | D+D | E+D | B+D | PC+D | D+HN | E+HN | B+HN | PC+HN |
|------------------------------|-----------|----------|----------|----------|-----------|----------|----------|-------------|
| Bias | | | | | | | | |
| Unstructure effect | 99.6795 | -10.7768 | -16.1114 | -27.5878 | 43.1853 | -13.0710 | -0.0225* | -27.5853 |
| Structure effect | -1.1400 | -1.1385 | -1.1416 | -0.7245* | -1.1387 | -1.1377 | -1.1397 | -0.7309 |
| AR1 Precision | -1.7360 | -1.7367 | -1.7445 | -1.6906* | 1.2E+76 | 1.4E+21 | 4.2E+35 | -3.3779 |
| AR1 Correlation | -0.1144 | -0.1127 | -0.1102 | -0.1138 | 0.0221 | 0.0176* | 0.0200 | 0.0253 |
| Unstr.time space interaction | 0.1743 | 0.1667 | 0.1241 | -0.1138* | 0.1686 | 0.1807 | 0.1774 | 0.1622 |
| MSE | | | | | | | | |
| Unstructure effect | 455894.93 | 12803.05 | 1.6E+16 | 761.09 | 247142.38 | 4.0E+33 | 1379.42 | 760.95* |
| Structure effect | 1.3083 | 1.3032 | 1.3117 | 0.5393* | 1.3047 | 1.3011 | 1.3071 | 0.5493 |
| AR1 Precision | 7.8558 | 8.0147 | 7.9349 | 7.8367* | 6.5E+183 | 3.1E+74 | 9.8E+101 | 102724985.8 |
| AR1 Correlation | 0.0499 | 0.0487 | 0.0473 | 0.0496 | 0.0405 | 0.0395* | 0.0406 | 0.0411 |
| Unstr.time space interaction | 0.2389 | 0.2217* | 0.5673 | 0.2389 | 0.2256 | 0.2471 | 0.2309 | 0.2387 |
| Coverage | | | | | | | | |
| Unstructure effect | 0% | 0% | 4% * | 0% | 0% | 0% | 3% | 0% |
| Structure effect | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| AR1 Precision | 64% | 65% | 60% | 66%* | 36% | 35% | 37% | 33% |
| AR1 Correlation | 84%* | 83% | 82% | 83% | 68% | 70% | 70% | 67% |
| Unstr.time space interaction | 84% | 86% | 86% | 90%* | 82% | 80% | 77% | 85% |
| Ratio of variance | | | | | | | | |
| Unstructure effect | 0.9192* | 1.8922 | 1.1E-13 | 0.3271 | 0.9120 | 3.3E-30 | 2.7768 | 0.3673 |
| Structure effect | 1.0155 | 0.7326 | 1.0065* | 0.3974 | 0.9027 | 0.7121 | 0.9080 | 0.4265 |
| AR1 Precision | 1.5593* | 1.5715 | 1.5695 | 1.5798 | 2.1E-30 | 6.0E-31 | 1.8E-29 | 3.0E-08 |
| AR1 Correlation | 1.5467 | 1.5567 | 1.5336* | 1.6148 | 1.9684 | 2.0172 | 2.0395 | 2.2024 |
| Unstr.time space interaction | 1.2693* | 1.3260 | 7.1E-29 | 1.3382 | 1.5928 | 1.3690 | 1.6494 | 1.2721 |

* means the smallest value through the row

In table 2, the meaning of abbreviated priors are the following:

- D+D: Default prior is used for the structured and unstructured spatial effect, in combination with Default prior for the random time process and the unstructured space-time interaction (Model1).
- E+D: Equivalent prior is used for the structured and unstructured spatial effect, in combination with Default prior for the random time process and the unstructured space-time interaction (Model2).
- B+D: Prior(B) of Bernardinelli is used for the structured and unstructured spatial effect, in combination with Default prior for the random time process and the unstructured space-time interaction (Model3).
- PC+D: Penalised Complexity prior is used for the structured and unstructured spatial effect, in combination with Default prior for the random time process and the unstructured space-time interaction (Model4).

- D+HN: Default prior is used for the structured and unstructured spatial effect, in combination with halfnormal prior for the random time process and the unstructured space-time interaction (Model5).
- E+HN: Equivalent prior is used for the structured and unstructured spatial effect, in combination with halfnormal prior for the random time process and the unstructured space-time interaction (Model6).
- B+HN: Prior(B) of Bernardinelli is used for the structured and unstructured spatial effect, in combination with halfnormal prior for the random time process and the unstructured space-time interaction (Model7).
- PC+HN: Penalised Complexity prior is used for the structured and unstructured spatial effect, in combination with halfnormal prior for the random time process and the unstructured space-time interaction (Model8).

First panel of table 2 shows the mean bias values calculated across all 8 fitted models. For each model and each hyperparameter estimate, the mean bias is calculated by averaging all biases across 100 simulated datasets. The values were rounded off to the nearest four decimals. By checking all mean biases for all hyperparameters across 8 models, it can be seen that model4 (PC+D), has smaller values of mean bias for three hyperparameters among 5 hyperparameters. These three hyperparameters are the structured spatial effect, AR1 precision and the unstructured space time interaction. This is an indication that priors used in fitting model4 is a good proposal prior for spatio temporal model in term of providing estimates with minimum bias.

Second panel of table 2 shows the mean squared error values calculated across all 8 fitted models. For each model and each hyperparameter estimate, the mean of mean squared error is calculated by averaging all mean squared errors across 100 simulated datasets. Among 8 models, still model4 provides estimates with minimum mean squared error for 2 hyperparameters(structured spatial effect and AR1 precision). Hence priors used in fitting model4 proposed in first panel, for the best unbiased estimator, still being proposed in Second panel, by considering unbiasedness estimator with minimum squared Error.

Third panel of table 2 represents the coverage probability for 95% confidence intervals for each hyperparameter estimate across 8 models. It can be seen that model4 has higher

coverage compared to other 7 models, as it has highest coverage for both AR1 precision and unstructured time space interaction. Still priors used to fit model4 are preferable in terms of coverage probability.

Fourth panel of table 2 represents the ratio of variances between variance of hyperparameters estimates and average variance from the model fitting. In this panel, model1 can be chosen as a good prior, since it provides three hyperparameter estimates among 5 with ratio closure to 1.

After obtaining PC prior as a good proposal prior for spatio-temporal model, three additional models have been fitted to see the effect of this PC prior, when they are used for other components of the model. However these results are equally compared with previous ones.

- PC+PC+PC: Penalised Complexity prior is used for the structured and unstructured spatial effect, random time process and the unstructured space-time interaction (Model9).
- PC+PC+D: Penalised Complexity prior is used for the structured and unstructured spatial effect, random time process, in combination with Default prior for the unstructured space-time interaction (Model10).
- PC+D+PC: Penalised Complexity prior is used for the structured and unstructured spatial effect, for the unstructured space-time interaction in combination with Default prior for the random time process (Model11).

In table 3 below, comparison of results from model4, model9, model10 and model11, shows that still model4 performs better in terms of mean bias, however for the remaining criterion, values are much more closure for model4 and model10.

Table 3: Comparison of estimated hyperparameters with true hyperparameters by Bias, MSE, coverage and ratio of variance Criterion

| | PC+D | PC+PC+PC | PC+PC+D | PC+D+PC |
|------------------------------|----------|----------|-----------|----------|
| Bias | | | | |
| Unstructure effect | -27.5878 | -27.5875 | -27.5869* | -27.5900 |
| Structure effect | -0.7245* | -0.7263 | -0.7264 | -0.7268 |
| AR1 Precision | -1.6906 | -1.4361 | -1.3950* | -1.7275 |
| AR1 Correlation | -0.1138* | -0.1398 | -0.1422 | -0.1144 |
| Unstr.time space interaction | -0.1138* | 0.1662 | 0.1760 | 0.1158 |
| MSE | | | | |
| Unstructure effect | 761.0879 | 761.0707 | 761.0397* | 761.2145 |
| Structure effect | 0.5393* | 0.5417 | 0.5431 | 0.5429 |
| AR1 Precision | 7.8367 | 6.2636 | 6.2386* | 8.0034 |
| AR1 Correlation | 0.0496* | 0.0519 | 0.0535 | 0.0510 |
| Unstr.time space interaction | 0.2389 | 0.2342 | 0.2420 | 1.2083* |
| Coverage | | | | |
| Unstructure effect | 0% | 0% | 0% | 0% |
| Structure effect | 0% | 0% | 0% | 0% |
| AR1 Precision | 66% | 70% | 71%* | 66% |
| AR1 Correlation | 83% | 77% | 88%* | 81% |
| Unstr.time space interaction | 90%* | 90%* | 87% | 86% |
| Ratio of variance | | | | |
| Unstructure effect | 0.3271* | 0.3034 | 0.3106 | 0.2630 |
| Structure effect | 0.3974* | 0.3891 | 0.3778 | 0.3970 |
| AR1 Precision | 1.5798 | 1.2783* | 1.3060 | 0.3898 |
| AR1 Correlation | 1.6148 | 1.3720* | 1.4068 | 1.6712 |
| Unstr.time space interaction | 1.3382 | 1.4017 | 1.2419* | 0.4832 |

* shows the smallest value through the row

By comparing the mean bias, mean squared error, coverage probability and ratio of variance for each hyperparameter for 4 respective models, it can be seen that the variabilities are very low. This is different from what we can observe from table 2, where the variabilities in mean bias, mean squared error, coverage probability and ratio of variance are highly depending on priors used. From this, we can suggest that hyperparameters estimated using PC prior are more stable as compared to those estimated using other given priors.

Further comparisons have been done in terms of DIC. Table below contains the averaged DIC for all fitted models.

Table 4: *Deviance Information Criterion values for all fitted model*

| Model | DIC |
|-------------------|------------|
| Model1 (D+D) | 30760.7338 |
| Model2 (E+D) | 30769.8777 |
| Model3 (B+D) | 30770.0757 |
| Model4(PC+D) | 30701.0643 |
| Model5 (D+HN) | 30765.9278 |
| Model6 (E+HN) | 30769.1848 |
| Model7 (B+HN) | 30715.6009 |
| Model8 (PC+HN) | 30771.2679 |
| Mod9 (PC+PC+PC) | 30771.8755 |
| Model10 (PC+PC+D) | 30722.7086 |
| Model11 (PC+D+PC) | 30771.6479 |

In table 4, comparison of the DIC values indicates that seven deviance information criterion(model2, model3, model5, model6, model8, model9 and model11) are nearly equal, the difference in DIC for each 2 models is less than 5, meaning no clear conclusion can be obtained from them.

On the other hand, by comparing DIC of all models, the smallest DIC is observed on model4, and the difference in DIC between each model and model4 is greater than 10, this indicated that model4 is selected as it has the smallest DIC. This is an other indication that priors used in fitting mode4 (PC+D) are good prior specification in spatio-temporal model.

4 Discussion and Conclusion

The main objective of this study was to investigate the effect of different prior specifications in a spatio-temporal model. Data simulation was done in R software by using true known parameters and hyperparameters given in table 1, then the parameters and hyperparameters were estimated by fitting simulation model (spatio-temporal model) to 100 simulated data in R. By specifying different priors, 11 models have been fitted by using INLA package which is alternative to MCMC.

Different priors specified are the following: Default prior, equivalent prior, prior(B) of Bernardinelli, halfnormal prior and Penalized Complexity prior. Default prior on log scale was chosen as INLA default prior given by $\text{Gamma}(1, 0.0005)$, it provided minimum information, hence it was considered as minimally informative prior [10]. Equivalent prior on log scale was equally considered, where the mean and precision are given equal value $\text{Gamma}(0.001, 0.001)$. Prior(B) of Bernardinelli which is a vague (noninformative) prior based on a chi-square distribution was used. It was chosen as $\text{Gamma}(1, 0.023)$ for the structured spatial effect and $\text{Gamma}(1, 0.01)$ for the unstructured spatial effect [8]. Halfnormal on log scale for standard deviation instead of the precision was used, it is given by $\log \text{tnormal}(0, 0.001)$, this type of prior is informative prior [3]. Lastly, Penalized Complexity prior was used, this prior is given by a type-2 Gumbel distribution with scaling parameter given by 0.01 and a location (spatial component) parameter given by $\frac{0.2}{0.31}$. This PC prior can be vague, weakly informative or strongly informative prior, depending on the way user adjusts a scaling parameter. In this study, it was used as informative priors [17]. The difference between PC prior and other used priors is that the PC prior are defined on individual components, while for non informative priors, priors depend on the structure of global model. In addition, PC prior assumed a heavy dependence between the structured and unstructured spatial effect which is different from other priors assumption [17].

Model1 fitted using default INLA prior for all components of the model does not provide good fit to the data, its DIC value is among the largest values, however it provides good fit in terms of ratio of variance. This result is in contrast with the following study, where a spatio-temporal model was fitted on counts of low birth weight using INLA default priors. Counts of low birth weight was defined if weight is less than 2500g, data are for 159 countries in the US state of Georgia during 2000-2010. In term of DIC a model containing

temporal effect, both the structured and unstructured spatial effect, the random time process and the unstructured space time interaction was found to provide more suitable fit to the data regardless of the added complexity [10].

In table 2, PC prior provides good fit in terms of minimum mean bias, minimum mean squared error as well as coverage probability. This is similar with study conducted on Larynx data in Germany with the interest to find spatial effect and the effect of an ecological covariate by applying PC priors. It was found that only spatial component has a contribution to the marginal variance and the effect of the ecological covariate appears to decrease towards the straight line(base model) [17].

Similarly, it was found that hyperparameters estimated using PC prior are more stable as compared to those estimated using other given priors as the variabilities in hyperparameters estimates are very low.

In terms of Deviance information criterion, the obtained DIC values for 11 models were compared, the DIC of model4 was found to be the smallest, indicating that model4 provides good fit to the data. In other words, as we fitted the same model by specifying different priors, this smallest DIC indicated that priors used in model4 are good choice priors for spatio-temporal model. This confirm the interpretation of other models using prior(B) of Bernardinelli, equivalent prior and halfnormal prior which did not provide good fit in terms of all criterion. That interpretation about model4 makes sense even if default prior alone did not provide a good fit in this analysis, we have seen an example where it performed well, [10], it is not surprising to obtained a combination of PC prior with default prior providing good fit than PC alone used to all model components or any other combination of priors.

From these, it can be concluded that PC in combination with default prior can be used as priors specification for spatio-temporal model. PC prior used is a type-2 Gumbel distribution with scaling parameter given by 0.01 and a location (spatial component) parameter given by $\frac{0.2}{0.31}$ and default prior used is Gamma(0.001, 0.001), it was used for both the random time process and the unstructured space time interaction. This conclusion is similar to the conclusion obtained from the analysis of Larynx data in Germany [17], it was recommended to construct PC prior not as the best prior, rather avoiding " cut and paste" prior choices from other similar research articles, and as alternative it was proposed to

use obtained tools to specify weakly informative or informative priors with a well defined shrinkage to get PC Prior which is believed to provide good fit as tool for real world applied statistics. Likewise, the researchers may know a better prior and applied it for their studies, however knowing a better default proposal for how to construct priors is a considerable advantage.

To strengthen the current evolution of Bayesian analysis, depending on few available prior specifications related researches conducted, further study should focus on other type of Bayesian hierarchical models for instance Spatial Survival Analysis and Spatial Longitudinal Analysis to scrutinize the deep awareness of their multi-application.

References

- [1] Schrödle, B., and Held, L. (2011). Spatiotemporal disease mapping using INLA. *Environmetrics*, 22(6), 725-734.
- [2] Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92(438), 607-617.
- [3] Lesaffre, E., Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.
- [4] Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.
- [5] Syversveen, A. R. (1998). Noninformative bayesian priors. interpretation and problems with construction and applications. *Preprint statistics*, 3, 1-11.
- [6] Fabrizi, E., Greco, F., and Trivisano, C. (2016). On the specification of prior distributions for variance components in disease mapping models. *Statistica*, 76(1), 93.
- [7] Wu, Y. (2006). *Bayesian tree models* (Vol. 68, No. 05).
- [8] Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). Bayesian estimates of disease maps: how important are priors?. *Statistics in medicine*, 14(2122), 2411-2431.
- [9] Haugsand, E. B. (2017). *Kingman's subadditive ergodic theorem and its application* (Master's thesis, NTNU).
- [10] Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013). Spatial and spatio-temporal models with R-INLA. *Spatial and spatio-temporal epidemiology*, 7, 39-55.
- [11] Blangiardo, M., and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- [12] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97-109.
- [13] Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2), 319-392.

- [14] Tierney, L., and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393), 82-86.
- [15] Ensoy, C. (2014). Forecasting and prediction models for demography and animal infectious diseases (Doctoral dissertation).
- [16] Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1), 1-20.
- [17] Simpson, D. P., Rue, H., Martins, T. G., Riebler, A., and Sørbye, S. H. (2014). Penalising model component complexity: A principled, practical approach to constructing priors. arXiv preprint arXiv:1403.4630.
- [18] Mood, A. M. (1950). *Introduction to the Theory of Statistics*.
- [19] Oxford University Statistics: Constructing Confidence Intervals. Retrieved in 2002 at <https://www.stats.ox.ac.uk/pub/bdr/IAUL/Course1Notes5.pdf>
- [20] Coverage probability of confidence intervals: A simulation approach. Retrieved on 26 July 2016 at <https://blogs.sas.com/content/iml/.../coverage-probability-confidence-intervals.html>

Appendix

Tables

Table 5: Comparison of estimated parameters with true parameters by Bias, MSE, coverage and ratio of variance Criterion

| | D+D | E+D | B+D | PC+D | D+HN | E+HN | B+HN | PC+HN |
|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| Bias | | | | | | | | |
| Intercept | 7.7335 | 7.7325 | 7.7335 | 7.7383 | 7.7581 | 7.7509 | 7.7623 | 7.7611 |
| time | -0.0030 | -0.0030 | -0.0030 | -0.0030 | -0.0039 | -0.0038 | -0.0044 | -0.0039 |
| sin1 | 0.0207 | 0.0208 | 0.0206 | 0.0209 | 0.0229 | 0.0189 | 0.0149 | 0.0184 |
| cos1 | -0.0224 | -0.0225 | -0.0225 | -0.0225 | -0.0227 | -0.0226 | -0.0241 | -0.0225 |
| MSE | | | | | | | | |
| Intercept | 60.2777 | 60.2307 | 60.2442 | 60.3223 | 63.4231 | 63.2861 | 63.1777 | 62.8701 |
| time | 0.0010 | 0.0010 | 0.0010 | 0.0005 | 0.0017 | 0.0017 | 0.0017 | 0.0016 |
| sin1 | 0.0749 | 0.0744 | 0.0746 | 0.0742 | 0.0793 | 0.0807 | 0.0805 | 0.0806 |
| cos1 | 0.0639 | 0.0636 | 0.0637 | 0.0636 | 0.0699 | 0.0700 | 0.0700 | 0.0698 |
| Coverage | | | | | | | | |
| Intercept | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| time | 90% | 92% | 93% | 90% | 98% | 97% | 98% | 98% |
| sin1 | 88% | 88% | 89% | 90% | 91% | 90% | 91% | 92% |
| cos1 | 88% | 88% | 89% | 91% | 93% | 91% | 91% | 93% |
| Ratio of variance | | | | | | | | |
| Intercept | 0.6026 | 0.6704 | 0.6719 | 0.6508 | 0.0419 | 0.0433 | 0.0476 | 0.0539 |
| time | 1.0809 | 1.0674 | 1.0667 | 1.0505 | 0.3610 | 0.3716 | 0.3592 | 0.3948 |
| sin1 | 1.3655 | 1.3866 | 1.3830 | 1.3831 | 1.1047 | 1.1405 | 1.1526 | 1.1515 |
| cos1 | 1.2635 | 1.2705 | 1.2634 | 1.2716 | 1.0360 | 1.0310 | 1.0405 | 1.0338 |

Table 6: Comparison of estimated parameters with true parameters by Bias, MSE, coverage and ratio of variance Criterion

| | PC+D | PC+PC+PC | PC+PC+D | PC+D+PC |
|-------------------|---------|----------|---------|---------|
| Bias | | | | |
| Intercept | 7.7383 | 7.7340 | 7.7343 | 7.7382 |
| time | 0.0051 | 0.0053 | 0.0052 | 0.0051 |
| sin1 | 0.0209 | 0.0212 | 0.0210 | 0.0208 |
| cos1 | -0.0225 | -0.0225 | -0.0201 | -0.0225 |
| MSE | | | | |
| Intercept | 60.3223 | 60.1799 | 60.1838 | 60.3222 |
| time | 0.0010 | 0.0009 | 0.0009 | 0.0010 |
| sin1 | 0.0742 | 0.0732 | 0.0730 | 0.0744 |
| cos1 | 0.0636 | 0.0624 | 0.0620 | 0.0641 |
| Coverage | | | | |
| Intercept | 0% | 0% | 0% | 0% |
| time | 90% | 90% | 87% | 91% |
| sin1 | 90% | 90% | 88% | 89% |
| cos1 | 91% | 89% | 87% | 88% |
| Ratio of variance | | | | |
| Intercept | 0.6508 | 1.0260 | 1.0290 | 0.6462 |
| time | 1.0505 | 1.4348 | 1.4469 | 1.0430 |
| sin1 | 1.3831 | 1.4466 | 1.4518 | 1.3809 |
| cos1 | 1.2716 | 1.3298 | 1.3301 | 1.2498 |

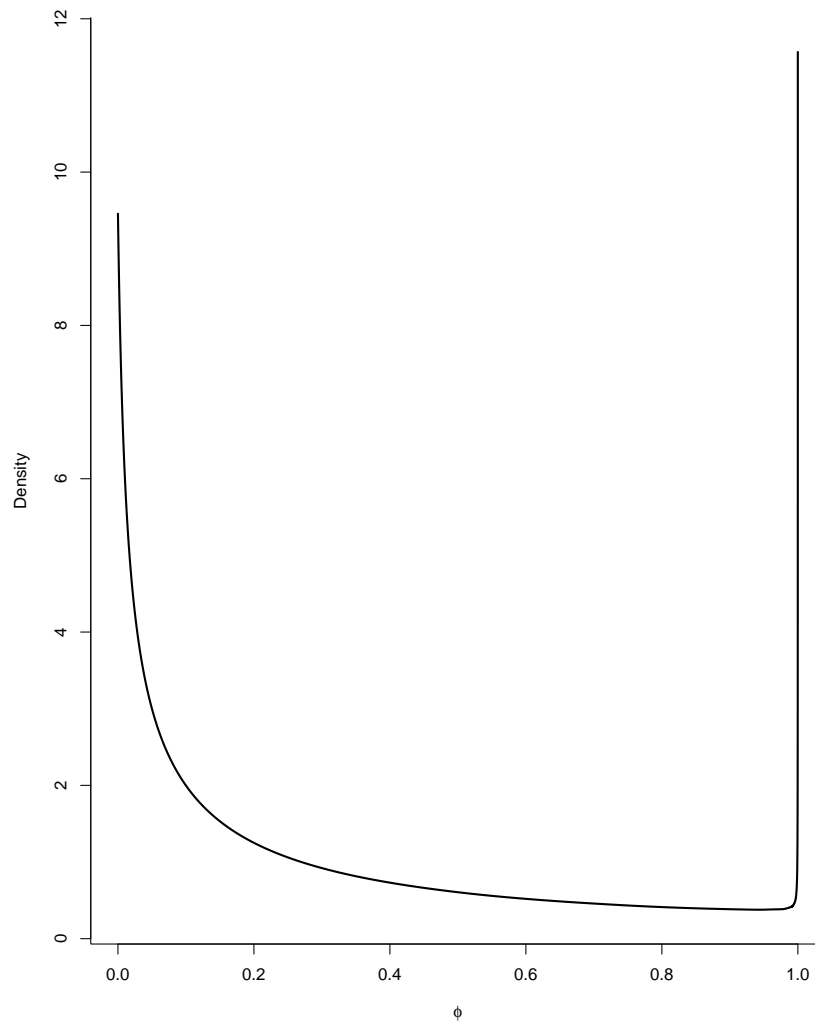


Figure 6: *PC prior density for the mixing parameter ϕ*

R codes

```
setwd("E:/TEST")
library(Matrix)
library(maptools)
library(sp)
library(spdep)
library(INLA)
library(Matrix)
loc.map = "E:\\TEST\\FlandersMap\\"
flanders = readShapePoly(paste(loc.map, 'vlaanderen.shp', sep=''))
plot(flanders)
####SIMULATION###
sIM<- function(i){
  set.seed(i)
  nTime = 36 # Number of time points
  nArea = 306 #Number of areas/locations
  N = nArea*nTime
  #fixed effect
  beta0=-7.9310 # Intercept
  ##Temporal effect
  beta1=-0.0090 # Linear
  beta2=-0.0572 # Sine
  beta3=-0.0805 # Cosine
  timeT = c(1:nTime)
  time = rep(timeT,nArea)
  bt=beta1*time+beta2*sin(2*pi*time/12)+beta3*cos(2*pi*time/12)
  length(bt)
  ###Spatial effect
  load("StructuredSpatialEffect.RData")
  xi_i
  xi=1.3828 # Structured effect Precision
  zeta=27.947 # Unstructured effect Precision
  ci= xi_i+rnorm(nArea,0,zeta^(-0.05))
}
```

```

#xi_i is known , no need of its distribution
ci=rep(ci,each=nTime)#306*36
####Space-time
rho_d=0.7857 # AR1 Correlation
tau_d=5.2799 # AR1 Precision
delta=6.0696 # Unstructured effect Precision
d1t =c(1:nTime)*0 #set the first values to 0
for(t in 2:nTime){
d1t[t]=rnorm(1,rho_d*d1t[t-1],tau_d^(-0.5)) #rho_d is the correlation,
#tau_d is the precision
}
dit=rep(d1t,nArea)+rnorm(N,0,delta^(-0.5))
###MEAN FUNCTION###
log_mu_it = beta0 +bt +ci+dit
lambda_simnew = exp(log_mu_it)*rep(flanders$CATTLE,each=nTime)
length(lambda_simnew)
###SIMULATED DATA###
y_simnew = rpois(n=N, lambda=lambda_simnew)
data_simnew = data.frame("ID"=rep(c(1:306),each=nTime),
"time"=rep(timeT,nArea),"y_simnew"=y_simnew)
##NEW VARIABLES ##
ID2=1:nrow(data_simnew)
time2=time
sin1=sin(2*pi*time/12)
cos1=cos(2*pi*time/12)
data_new=cbind(data_simnew, ID2,time2,sin1,cos1)
return(data_new)
}
Nsim<-100
seedi<-1:Nsim
datasets<-sapply(seedi, function(x){sIM(x)}, simplify = F)
#Creating datasets containing 100 simulated data
SIM_i=rep(NA,Nsim)
for (i in 1: length(SIM_i)) {

```

```

SIM_i[i]=paste("SIM",i,sep = '')
names(datasets)=SIM_i

# Saving datasets
data1<-datasets[[i]]
save(data1,file=paste("E:\\TEST\\SIMULATION\\datasets\\data",i,
'.RData',sep=""))
}
#####
##NEIGHBOUR = SHARING BOUNDARIES##
neighmat_sb1 <- nb2mat(neighbours,style="B",zero.policy=TRUE)
neighb.graph <- mat2listw(neighmat_sb1)
neighb.adj <- "E:/TEST/adjacencyFlanders.adj" #location
nb2INLA(neighb.adj, neighb.graph$neighbours)

##PLOT CATTLE OF POPULATION ON MAP

plotclr <-c(rev(bpy.colors(n = 99, cutoff.tails = 0.1, alpha = 1.0)))
spplot(flanders, "CATTLE", col.regions=plotclr,
names.attr = "CATTLE",
colorkey=list(space="right",cex=0.4,height=1,width=0.8),
scales = list(draw = F), col="red",as.table = TRUE)

##PLOT 2 SIMULATED DATA ON MAP

plotclr <-c(rev(bpy.colors(n = 99, cutoff.tails = 0.1, alpha = 1.0)))
spplot(flanders, c("SIM1_time1", "SIM1_time2","SIM1_time15","SIM1_time36" ),
col.regions=plotclr,
names.attr = c("Simulation 1_time 1","Simulation 1_time 2",
"Simulation 1_time 15","Simulation 1_time 36"),
colorkey=list(space="right",cex=0.4,height=1,width=0.8),
scales = list(draw = F),col="red",as.table = TRUE)
##
plotclr <-c(rev(bpy.colors(n = 99, cutoff.tails = 0.1, alpha = 1.0)))

```

```

spplot(flanders,c("SIM100_time12", "SIM100_time14","SIM100_time25",
"SIM100_time31"), col.regions=plotclr,
names.attr = c("Simulation 100_time 12","Simulation 100_time 14 ",
"Simulation 100_time 25 ","Simulation 100_time 31"),
colorkey=list(space="right",cex=0.4,height=1,width=0.8),
scales = list(draw = F),col="red",as.table = TRUE)

##Linear trend plot for 2 selected simulate data##

library(timeDate)
library(timeSeries)
library(graphics)
library(zoo)
library(xts)
library(PerformanceAnalytics)

pcp <- read.csv("regions.csv", header = T) #import the new created dataset
#containing total cattle abortion in 2 areas for 3 years per month.
pcp
PCP <- ts(pcp, frequency = 12) #change it to a timeseries format
PCP
PCP1 <- ts(PCP[ , -1], frequency = 12) #remove the first column
PCP1
plot(as.xts(PCP1), xlab = "Months", ylab = "Cattle abortions",
major.format = "%Y-%m", gpars=list(lty=c(1:6)),
main = 'Simulated cattle abortions in 2 Areas of
Flanders for 3 years (per month)',cex.main = 0.6, ylim = c(0, 26),
cex=1.1, yaxis.right = F, grid.ticks.lwd = 0.1, xaxt='n')

addLegend("topright", on=1, legend.names = c("Year1.Area1", "Year2.Area1",
"Year3.Area1", "Year1.Area2", "Year2.Area2", "Year3.Area2"),
lty=c(1, 1,1,1,1,1), lwd=c(1,1,1,1,1, 1), cex = 0.65)

#####

```

```
### model 1 : Default prior+Default prior ###
```

```
formula1=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,  
hyper=list(prec.unstruct= list(prior=  
"loggamma", param=c(1,0.0005)),prec.spatial=list(prior=  
"loggamma", param=c(1,0.0005))))+f(time2,model="ar1",  
hyper=list(prec= list(prior= "loggamma", param=c(1,0.0005))))+  
f(ID2,model="iid",hyper=list(prec= list(prior=  
"loggamma", param=c(1,0.0005))))  
#for (i in 1: length(Nsim)){  
for (i in 1:100 ){  
mod1<- inla(formula1,family="poisson",data=datasets[[i]],  
control.compute = list(dic=T))  
# saving output  
save(mod1,file=paste("E:\\TEST\\SIMULATION\\default\\default",i,  
' .RData',sep=""))  
}
```

```
### model 2 : Equivalent prior+Default prior###
```

```
formula2=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,  
hyper=list(prec.unstruct= list(prior=  
"loggamma", param=c(0.001,0.001)),prec.spatial=list(prior=  
"loggamma", param=c(0.001,0.001))))+f(time2,model="ar1",  
hyper=list(prec= list(prior=  
"loggamma", param=c(1,0.0005))))+f(ID2,model="iid",  
hyper=list(prec= list(prior="loggamma", param=c(1,0.0005))))  
for (i in 1: 100){  
mod2<- inla(formula2,family="poisson",data=datasets[[i]],  
control.compute = list(dic=T),control.predictor=list(compute=TRUE))  
save(mod2,file=paste("E:\\TEST\\SIMULATION\\equivalent\\equivalent",i,  
' .RData', sep=""))  
}
```

```

### model 3 : Prior according to Bernardinelli+default prior##

formula3=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,
hyper=list(prec.unstruct= list(prior=
"loggamma", param=c(1,0.01)),prec.spatial=list(prior=
"loggamma", param=c(1,0.023))))+f(time2,model="ar1",
hyper=list(prec= list(prior="loggamma", param=c(1,0.0005))))+
f(ID2,model="iid",hyper=list(prec= list(prior=
"loggamma", param=c(1,0.0005))))
for (i in 1: 100){
mod3<- inla(formula3,family="poisson",data=datasets[[i]],
control.compute = list(dic=T),control.predictor=list(compute=TRUE))
save(mod3,file=paste("E:\\TEST\\SIMULATION\\bernar\\bernar",i,
'.RData',sep=""))
}

###model4: pc prior+ default prior##

g = neighb.adj
Q = INLA:::inla.pc.bym.Q(g)
Q = INLA:::inla.scale.model(Q, constr=list(A=matrix(1, 1, n), e=0))
u = 0.2/0.31
alpha = 0.01
phi.u = 0.5
phi.alpha = 2/3 ## prob(phi < phi.u) = phi.alpha

formula4 = y_simnew ~ 1 +time+sin1+cos1+f(ID,
model = "bym2",
graph=g,
scale.model = TRUE,
constr = TRUE,
hyper=list(
phi = list(

```



```

prior = "pc",
param = c(phi.u, phi.alpha),
initial = -3),
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))+
f(time2,model="ar1",hyper=list(prec= list(prior=
"loggamma", param=c(1,0.0005))))+f(ID2,model="iid",
hyper=list(prec= list(prior=
"loggamma", param=c(1,0.0005))))
for (i in 1: 100){
mod4 = inla(formula4, data =datasets[[i]], family = "poisson",
control.compute = list(dic=T),
control.predictor = list(compute=TRUE))
save(mod4,file=paste("E:\\TEST\\SIMULATION\\pct\\pc",i,'.RData',sep=""))
}

```

#####model5: Default prior+halfnormal prior##

```

formula5=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,
hyper=list(prec.unstruct= list(prior=
"loggamma", param=c(1,0.0005)),prec.spatial=list(prior=
"loggamma", param=c(1,0.0005))))+f(time2,model="ar1",
hyper=list(prec= list(prior="logtnormal", param=c(0,0.001))))+
f(ID2,model="iid",hyper=list(prec= list(prior=
"logtnormal", param=c(0,0.001))))
for (i in 1: 100){
mod5<- inla(formula5,family="poisson",data=datasets[[i]],
control.compute = list(dic=T),
control.predictor=list(compute=TRUE))
save(mod5,file=paste("E:\\TEST\\SIMULATION\\halfnormal\\default\\default",i,
'.RData',sep=""))
}

```

```

##model6: Equivalent prior+half normal prior##

formula6=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,
hyper=list(prec.unstruct= list(prior=
"loggamma", param=c(0.001,0.001)),prec.spatial=list(prior=
"loggamma", param=c(0.001,0.001))))+f(time2,model="ar1",
hyper=list(prec= list(prior= "logtnormal", param=c(0,0.001))))+
f(ID2,model="iid",hyper=list(prec= list(prior=

"logtnormal", param=c(0,0.001))))
for (i in 1: 100){
#for (i in 1: length(Nsim)){
mod6<- inla(formula6,family="poisson",data=datasets[[i]],
control.compute = list(dic=T),control.predictor=list(compute=TRUE))
save(mod6,file=paste("E:\\TEST\\SIMULATION\\halfnormal\\equivalent
\\equivalent",i,'.RData',sep=""))
}

####model7: Prior according to Bernardinelli+half normal prior##
formula7=y_simnew~1+time+sin1+cos1+f(ID,model="bym",graph=neighb.adj,
hyper=list(prec.unstruct= list(prior=
"loggamma", param=c(1,0.01)),prec.spatial=list(prior=
"loggamma", param=c(1,0.023))))+f(time2,model="ar1",
hyper=list(prec= list(prior= "logtnormal", param=c(0,0.001))))+
f(ID2,model="iid",hyper=list(prec= list(prior=

"logtnormal", param=c(0,0.001))))
for (i in 1: 100){

mod7<- inla(formula7,family="poisson",data=datasets[[i]],
control.compute = list(dic=T),control.predictor=list(compute=TRUE))

```

```

save(mod7,file=paste("E:\\TEST\\SIMULATION\\halfnormal\\bernar\\bernar",i,
'.RData',sep=""))
}

###model8: pc prior+ half normal prior##
g = neighb.adj
Q = INLA:::inla.pc.bym.Q(g)
Q = INLA:::inla.scale.model(Q,  constr=list(A=matrix(1, 1, n), e=0))
u = 0.2/0.31
alpha = 0.01
phi.u = 0.5
phi.alpha = 2/3 ## prob(phi < phi.u) = phi.alpha
formula8 = y_simnew ~ 1 +time+sin1+cos1+f(ID,
model = "bym2",
graph=g,
scale.model = TRUE,
constr = TRUE,
hyper=list(
phi = list(
prior = "pc",
param = c(phi.u, phi.alpha),
initial = -3),
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))
+f(time2,model="ar1",hyper=list(prec= list(prior=
"logtnormal", param=c(0,0.001))))
+f(ID2,model="iid",
hyper=list(prec= list(prior= "logtnormal", param=c(0,0.001))))
for (i in 1: 100){
mod8 = inla(formula8, family = "poisson",data =datasets[[i]],
control.compute = list(dic=T),control.predictor = list(compute=TRUE))
save(mod8,file=paste("E:\\TEST\\SIMULATION\\halfnormal\\pc\\pc",i,
'.RData',sep=""))
}

```

```

###model9: pc for everything##

formula9 = y_simnew ~ 1 +time+sin1+cos1+f(ID,
model = "bym2",
graph=g,
scale.model = TRUE,
constr = TRUE,
hyper=list(
phi = list(
prior = "pc",
param = c(phi.u, phi.alpha),
initial = -3),
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))f(time2,model="ar1",
hyper=list(
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))f(ID2,model="iid",
hyper=list(
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))
for (i in 1:100){
mod9 = inla(formula9, family = "poisson",data =datasets[[i]],
control.compute = list(dic=T),
control.predictor = list(compute=TRUE))
save(mod9,file=paste("E:\\TEST\\SIMULATION\\B\\pcpcpc\\pc",i,
'.RData',sep=""))
}

```

```

###model10: pc+pc+d##

formula10 = y_simnew ~ 1 +time+sin1+cos1+f(ID,
model = "bym2",graph=g,scale.model = TRUE,
constr = TRUE, hyper=list(phi = list(prior = "pc",
param = c(phi.u, phi.alpha),
initial = -3),
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))f(time2,model="ar1",
hyper=list(
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))f(ID2,model="iid",hyper=list(prec= list(prior=
"loggamma", param=c(1,0.0005))))
for (i in 1:100){
mod10 = inla(formula10, family = "poisson",data =datasets[[i]],
control.compute = list(dic=T),
control.predictor = list(compute=TRUE))
save(mod10,file=paste("E:\\TEST\\SIMULATION\\B\\pcpcd\\pc",i,
'.RData',sep=""))
}

###model11: pc+d+pc##
formula11 = y_simnew ~ 1 +time+sin1+cos1+f(ID,
model = "bym2",
graph=g,
scale.model = TRUE,
constr = TRUE,
hyper=list(
phi = list(

```

```

prior = "pc",
param = c(phi.u, phi.alpha),
initial = -3),
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))+f(time2,model="ar1",hyper=list(prec= list(prior=
"loggamma", param=c(1,0.0005))))+f(ID2,model="iid",
hyper=list(
prec = list(
prior = "pc.prec",
param = c(u, alpha),
initial = 5)))

for (i in 1:100){
mod11 = inla(formula11, family = "poisson",data =datasets[[i]],
control.compute = list(dic=T),
control.predictor = list(compute=TRUE))
save(mod11,file=paste("E:\\TEST\\SIMULATION\\B\\pcdpc\\pc",i,
'.RData',sep=""))
}

###Plotting PC prior###
neighmat_sb1 <- nb2mat(neighbours,style="B",zero.policy=TRUE)
neighb.graph <- mat2listw(neighmat_sb1)

neighb.adj <- "E:/TEST/adjacencyFlanders.adj" #location
nb2INLA(neighb.adj, neighb.graph$neighbours)

write.new.figures = TRUE

inla.dev.new.hook = function()
{
cex.lab = 1.4

```

```

cex.axis = 1.4
par(cex.lab=cex.lab, cex.axis=cex.axis)
}
inla.dev.new()

g = neighb.adj
Q = INLA:::inla.pc.bym.Q(g)
Q = INLA:::inla.scale.model(Q, constr=list(A=matrix(1, 1, n), e=0))
u = 0.2/0.31
alpha = 0.01
phi.u = 0.5
phi.alpha = 2/3 ## prob(phi < phi.u) = phi.alpha
if (FALSE) {
## make figures of the priors
for (g in "neighb.adj") {
log.prior = INLA:::inla.pc.bym.phi(Q=Q, rankdef=1, u=phi.u,
alpha = phi.alpha)
phis = 1/(1+exp(-seq(-10, 10, len=10000)))
inla.dev.new()
plot(phis, exp(log.prior(phis)), type="l",#plotting a prior knowledge
lwd = 2, bty = "l",
xlab = expression(phi),
ylab = "Density")
if (length(grep("^sar", g)) > 0) {
fnm = "datasets[[1]]-prior.ps"
} else {
fnm = "neighb-prior.ps"
}
if (write.new.figures)
dev.print(postscript, file=fnm)
}
if (write.new.figures)
system("which psfix && psfix *-prior.ps")
}

```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Prior specification of spatio-temporal models.

Richting: **Master of Statistics-Biostatistics**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Uwineza, Marie Aimee

Datum: **23/01/2018**