UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

The spatio-temporal modeling of prostate cancer in Limburg

Worku Biyadgie Ewnetu

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

SUPERVISOR :

dr. Thomas NEYENS

2017
2018

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

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics

Masterthesis

The spatio-temporal modeling of prostate cancer in Limburg

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

SUPERVISOR :
dr. Thomas NEYENS

Acknowledgments

First and foremost, I would like to thank the almighty God, who helped and blessed me in all my activities, and to the Holy Virgin Mary. I only can say “Glory to the almighty God”.

Secondly, I would like to express my deepest gratitude to my supervisor Dr. Thomas NEYENS, for his motivation and guidance since kickoff date for topic selection. Thank you for giving me the chance to work with you. You helped me by giving advice, unreserved helpful comments and encouraged me to complete the thesis within a scheduled time. I am so very happy and feeling proud for being one of your students.

I would also like to express my gratitude with ample respect for all my Professors and Doctors, who taught me at Hasselt University. Thank you for sharing me your immense knowledge with wonderful experience. I have gained from all of you not only in academia but also the great inspiration that will help me in my endeavor. Next, I would also like to thank the VLIR-UOS for awarding me a full scholarship during my stay at the University of Hasselt, which made my dream true. Thank you for the Flemish government for established this well-organized and fruitful program, which provide a great opportunity for many students like me. The next my great thank goes to all my classmates for that we already spent here with excellent communication and shared distinct experience for the last two years.

Lastly but not the least, my deepest gratitude with enthusiastic respect goes to my beloved father, mother, and sister for your support and advice. I know very well that you sacrificed and spent many challenges to grow me physically and mentally. All my day to day activities is nothing without your good wishes and heartfelt prayers. It is the time to express my heartfelt gratitude to my lovely wife Bizualem Tesfa (multi-hope) for your advice, encouragement, and the time you spent alone in the last two years. This is the result of your prayer.

Worku Biyadgie Ewnetu

June 18, 2018

Pietelbeekstraat 15,

Hasselt, Belgium

Abstract

Background: The global burden of cancer remains an important concern in worldwide, and prostate cancer is the most common cancer for the men population. The age-standardized incidence rates of prostate cancer have increased dramatically because of the increased availability of screening for prostate-specific antigen (PSA) in men without symptoms of the disease. Environmental exposures might be a cause too. In disease mapping, association in nearby areas is common and the assumption of independence to do the analysis using classical likelihood principle may not be efficient. In addition, in the case of a small-area specific study, we usually faced to scarce data, and the classical fixed effects model with maximum likelihood estimation technique often leads to unsatisfactory results in each area. These problems can be overcome by using a Bayesian hierarchical modeling approach which models the spatial and/or temporal effects as random effects, through a prior distribution.

Objectives: The main objective of this study was to assess the evolution of prostate cancer disease risk taking into account both space and time simultaneously and to investigate whether the spatial and temporal effects are separable. It was also aimed at predictions of the true relative risk of the disease across each municipality.

Methods: The data analyzed include yearly incidence counts from prostate cancer, which was subdivided according to 18 age groups in the male population as observed in each of these municipalities during the years 1996-2005 in Limburg. Bayesian hierarchical models to the spatio-temporal disease mapping were implemented in two basic approaches, namely parametric linear time trend and smooth temporal evolution models to account for the evolution of prostate cancer risk in both space and time simultaneously.

Results: From the exploratory analysis, we found an evidence of overdispersion of the incidence counts as well as heterogeneity indicating dependence between the relative risks in neighboring municipalities. The Bayesian hierarchical models were able to remove substantial variability due to small observed as well as expected counts from the raw SIR. The smoothed temporal evolution models showed better fit than a parametric linear trend model to provide a substantial shrinkage of the raw relative risk estimates for each municipality. The results showed inseparable space-time variation in the prostate cancer disease risk in Limburg for the years 1996-2005, hence, it was shown stable patterns for the first four years, and evolving differently for the remaining subsequent years. In conclusion, the results suggested that the time trends for every municipality do not rely on a parametric shape, but flexible to describe the variety of time trends that arise in the data. Moreover, the spatial dependence was very important to describe the behavior of the risk, indeed higher than the temporal one.

Keywords: Overdispersion, Standardization, Relative Risk, Disease Mapping, Bayesian Hierarchical Models.

Contents

List of Tables

List of Figures

List of abbreviations

AIC	Akaike information criterion
ANOVA	Analysis of variance
BGR	Books-Gelman-Rubin
BIC	Bayesian information criterion
CAR	Conditional autoregressive
CH	Correlated heterogeneity
D	Deviance
DIC	Deviance information criterion
GR	Gelman Rubin
LIKAR	Limburg Cancer Registry
MAPE	Mean absolute predictive error
McCLT	Markov chain Central Limit Theorem
McMLLN	Markov chain Law of Large Numbers
McMC	Markov chain Monte Carlo
MSPE	Mean squared predictive error
pD	Effective number of parameters
PG	Poisson-gamma
PL	Poisson-lognormal
PSA	Prostate-specific antigen
RR	Relative risk
SIR	Standardized risk ratio
UH	Unstructured heterogeneity

1 Background of the study

1.1 Introduction

According to the Global Burden of Disease Cancer result explored in 2015, cancer was the second leading cause of death globally, 8 million deaths, whereas cardiovascular diseases being the first [37,38]. Their finding also showed that, between 2005 and 2015, cancer cases increased by 33%, with population aging contributing 16%, population growth 13%, and changes in age-specific rates contributing 4%. For men, the most common cancer globally was prostate cancer (1.6 million cases). Prostate cancer is cancer that begins in tissues of the prostate gland. Located just below the bladder and in front of the rectum, the prostate is the male sex gland responsible for the production of semen. According to [42], the size of the prostate also changes with age. The age-standardized incidence rates of prostate cancer have increased dramatically because of the increased availability of screening for prostate-specific antigen (PSA) in men without symptoms of the disease. Environmental exposures might be a cause too. This motivates public health researchers and great attention is invested to describe the patterns of cancer disease in both geographical and temporal dimensions to provide accurate results for healthcare policymakers.

Disease mapping studies aim to summarize spatial variation in disease risk, in order to assess and quantify the amount of true spatial heterogeneity and the associated patterns, to highlight areas of elevated or lowered risk and to obtain clues as to the disease aetiology [39]. The identification and quantification of patterns in disease occurrence and mortality are recently considered as the primary stages for public health studies to increase the understanding and control of that particular disease [1]. Different sources in literature are using the terms “geographical epidemiology”, “spatial epidemiology”, and “medical geography” interchangeably to describe analytic methods in the study of spatial distribution of disease incidence and mortality. In spatial epidemiology, maps are mostly useful tools for enlightening the potential causes of disease such as monitoring air pollution [33], besides describing the patterns for incidence or prevalence of disease. In addition to disease mapping, the application of spatial models is also increasing in different areas such as crime [32], and ecology [34]. Statistical methods are highly important to analyze public health data starting from the concept of the probability of each person contracting a disease. Spatial statistical methods are used to evaluate differences in observed disease risk rates from different geographic locations, more specifically to separate pattern from noise and identify disease clusters, and to assess the significance of potential exposures for the response of interest [1]. In disease mapping, association in nearby areas is common and the assumption of independence to do the analysis using classical likelihood principle may not be efficient. In addition, in the case of a small-area specific study we usually faced to a scarce data, and the classical fixed effects model with maximum likelihood estimation technique often leads to unsatisfactory results in each area. These problems can be overcome by using a Bayesian hierarchical modeling approach which models the spatial and/or temporal effects as random effects, through a prior distribution based on our prior belief [1, 14, 24]. In this perspective, a random effect is an extra quantity of variation (or variance component) which is estimable within the map and can be ascribed a defined probabilistic structure [28].

This study focuses on modeling prostate cancer incidence data in the years 1996-2005 in Limburg, found in the north-east of Belgium. Previously the data have been analyzed and published by [6] and [36]. In [6], they used Bayesian hierarchical models taken into account spatially structured and unstructured random main effects to show the spatial distribution of the risk. The number of new cases occurring between 1996 and 2005 years was summed over all years and the temporal effect was not considered in the disease mapping process. On the other hand, the work by [36] has been done on the spatio-temporal data structures to investigate whether regional spatial patterns change over time. Here a parametric linear model was used as proposed by Bernadinelli *et al* [16], and temporally independent model proposed by Waller *et al* [17]. Although there is sharing of information in space and time in the popularized parametric linear model, it is restricted to have a predefined linear shape. On the contrary, temporally independent models estimate the risks for each period independently of those from the previous periods, therefore, there is no sharing of information in time [24]. This study was aimed at the spatio-temporal disease mapping by extending the parametric linear trend models to smooth temporal evolution models and hence information can be shared in space and time simultaneously, and to smooth the underlying risk estimates because the data are typically sparse. Several models have been fitted, and comparisons between parametric linear model and smooth temporal evolution models were employed. On one hand, it is important from the epidemiological and public health point of view to provide accurate relative risk patterns mostly characterized by a spatial and/ or temporal structure, which needs to be taken into account simultaneously in the inferential process. On the other hand, more elaborated models are now accessible to estimate the relative risk-free of background noise, and comparing them with respect to goodness-of-fit and predictive ability has its own statistical advantage.

The paper is structured as follows: first, in section 2 we described the methodology part starting from the exploratory data analysis to the statistical models used in the analyses. Bayesian hierarchical models are described in both the spatial only and spatio-temporal data analyses. In this section, detailed about the McMC estimation technique with model diagnostics and information covering sensitivity analysis is also included. In section 3, the data are analyzed and several models are compared with detail descriptions. Section 4 discusses the main results, and finally, we draw conclusions.

1.2 Objectives

The main objective of this study was to assess the evolution of prostate cancer disease risk between 1996 and 2005 years in Limburg taking into account both space and time simultaneously and to investigate whether the spatial and temporal effects are separable. It was also aimed at predictions of the true relative risk of the disease across each municipality.

1.3 Data description

The data in this study was obtained from the Limburg Cancer Registry (LIKAR) [36]. The Limburg Cancer Registry (LIKAR) is designed to register all cancers in the province of Limburg, Belgium. The LIKAR database contains the number of new, histologically or cytologically proven invasive and non-invasive cancers within male or female inhabitants of Limburg between the years 1996 and 2005. The region, Limburg is situated in the north-east of Belgium and divided into 44 contiguous but non-overlapping municipalities or counties. The largest populations in the region found in the middle of the province, in Hasselt. The data analyzed include yearly incidence counts from prostate cancer in the male population as observed in each of these municipalities during the 1996-2005 period. The aggregate of incidence, the population at risk subdivided according to 18 age groups with 5 years gap, and geographic area was registered for each municipality. Incidence, in this case, is the number of new cases contracting prostate cancer in the specified municipality in a given time period.

Table 1.1: Male populations at risk and incidence counts of prostate cancer in Limburg for the years 1996-2005.

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Population	389,655	391,601	393,331	394,797	396,405	398,163	400,193	401,812	403,363	405,465
Incidence	459	504	514	517	604	585	514	677	634	726

In 1996 the population at risk consisted of 775,302 inhabitants, of which 389,655 men and, while in 2005 the population totaled 814,658, of which 405,465 men. Table 1.1 shows in total 5,734 inhabitants of the Belgian province of Limburg were new cases contracting with prostate cancer, of which 5,638 were invasive, and 96 non-invasive tumors during the years 1996-2005. Detailed information about the populations at risk with the incidence of different cancer types for the province of Limburg can be found in [36]. The number of new cases have been increased except for the years 2001 and 2002.

2 Methodology

2.1 Exploratory analysis

The aim of disease mapping in this study was to provide a representation of the spatial distribution of the risk of a prostate cancer disease in male's population of Limburg region. The below construction of standardized relative risk (SIR), the test of overdispersion and spatial autocorrelation were used as preliminary analyses for background measures for disease mapping. This form of mapping is useful for highlighting areas of incidence needing further consideration. For instance, contour plots or surface views of such mapped data can be derived, however, inspection of maps of simple background rates cannot provide the accurate assessment of the statistical significance of disease risk in the given area. The risk in this study reflects the number of people who suffer from the prostate cancer disease (morbidity) in the years 1996-2005 for the population at risk. As a result, considering the incidence data in different groups the importance of age can be explored and potential known confounding factor has been removed before doing any further analysis of the data [11].

2.1.1 Standardization

There is an increasing incidence of cancer with age, with marked increases for ages greater than 40 years, see Figure 2.1. Since incidence proportions reflect estimated average risks for a study areas, municipalities having more people in higher age ranges will have higher values of incidence rates than those of younger populations. As a result, the incidence proportion for two municipalities may appear different, but this difference may be due entirely to the different age distributions within the municipalities, rather than to a difference in the underlying age-specific risk of disease [1, 42]. Rate standardization can be used as a mechanism to adjust summary rates to remove the effect of known risk factors (such as age, gender) and make comparable rates from different populations [1]. The number of persons at risk and the number of incidences counts for prostate cancer across age groups are displayed in Figure 2.1. The number of disease counts observed in a county is clearly a function of the age distribution of the population at risk. As a result, rate standardization has been used to compare observed rates from municipalities with different age distributions. In the perspective of statistics, rate standardization amounts to taking a weighted average of observed age group-specific rates where the weights relate the age distribution in the study population to that in the standard population [1].

Inskip *et al.* [4] reviewed different standardization methods. According to their conclusion, direct and indirect standardization are the most commonly used techniques for summarizing rates and comparing populations. Direct age standardization is used to adjust observed rates to reflect rates and counts that we would observe by applying the observed age-specific rates *directly* to the standard population, while indirect standardization use age-specific rates from the standard population to estimate *indirectly* the numbers of cases expected in each age group in the observed study population. To compare the number of study populations, each standardized against the same standard population, the direct method is advocated because it preserves consistency between the populations. The indirect method requires age-specific rates for the

standard population, the number of people at risk in the study population, and the total number of cases observed in the study population, and so can be used when the direct method cannot be applied [4]. An indirect standardization can be used to seeks to answer the question: what would be the number of cases expected in the study population if people in the study population contracted the disease at the same rate as people in the standard population? [3]. Moreover, in this perspective, age-specific rates for the standard population is unknown, hence internal, which is an aggregate (marginal) standards are used in indirect standardization. Breslow and Day [5] also suggested that an internal standard population (super-population containing the regions of interest) is preferred as compared to an external standard population (entirely separate population), for comparability of the regions. Furthermore, the advantage of using the indirect method is that it has a low standard error [4].

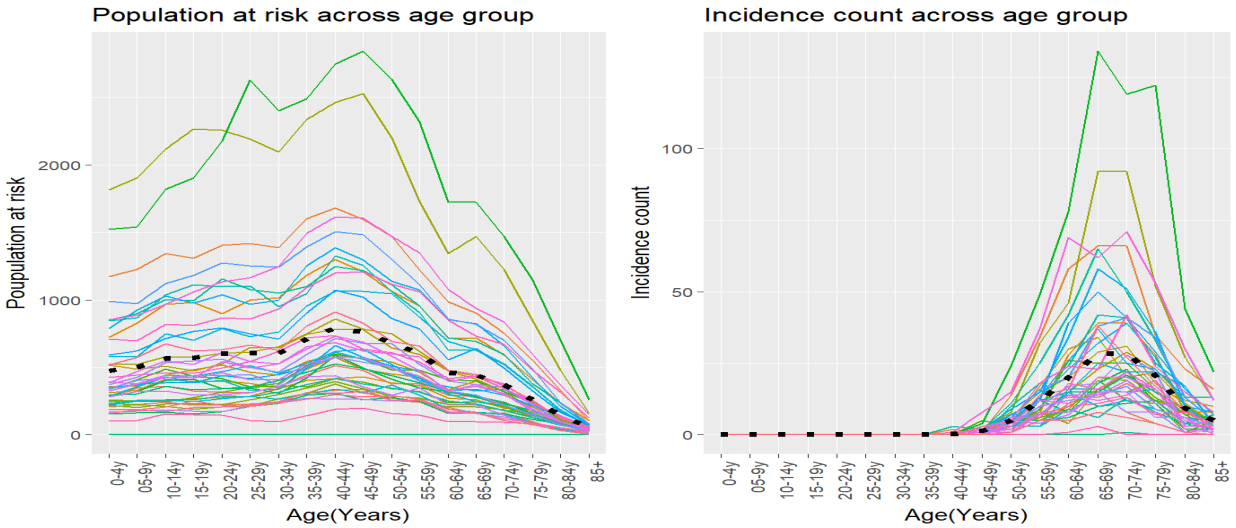


Figure 2.1: Number of incidence counts and populations at risk across age groups.

In this study, we will use the term area throughout to refer to the enumeration districts partitioning the study region, and the term region to refer to the entire study area or a collection of municipalities (Limburg). So that area, municipality, and county would be used interchangeably. For the spatial data, assume that the study region A is divided into N contiguous but non-overlapping areas such that $\boldsymbol{A} = (A_1, A_2, ..., A_N)$, and a response Y_i is observed in each area, thus providing a set of the number of people with prostate cancer disease which is a count data, $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_N)$. Furthermore, let G denote the number of age groups, Y_{ig} is the number of incidences or new cases in age group g for the study population i, n_{ig} denote the number of people at risk in age group g and for study population i. One simple measure of disease risk is the standardized incidence ratio ($\texttt{SIR}_i$) for area i computed with standardized expected disease counts and can be expressed as

$$\texttt{SIR}_i = \frac{Y_i}{E_i} = \frac{\sum_{g=1}^{G} Y_{ig}}{\sum_{g=1}^{G} E_{ig}}, \tag{2.1}$$

where Y_i and E_i denotes the overall observed and expected number of cases of prostate cancer disease for the ith study population, respectively, whereas E_{ig} is the expected number of cases of prostate cancer disease in

age group g for the study population i, and can be obtained by $E_{ig} = \left(\frac{\sum_{i=1}^{N} Y_{ig}}{\sum_{i=1}^{N} n_{ig}}\right) n_{ig}$ involving the principle of internal standardization. Note that the Y_i are assumed to be random variables, while the E_i are as fixed known as a function of n_i, the number of persons at risk for the disease in county i. An incidence ratio value larger than one indicates an area that has a higher observed disease risk than the expected, while a county with an incidence ratio less than one implies a lower observed than the expected disease risk.

The indirect age standardisation for the spatial data given in (2.1) can easily be extended to the spatio-temporal data which adjusts the observed rates by taking into account both space and time dimensions. Let $\texttt{SIR}_{ij}$ denote the standardized incidence rate for the ith area at the jth time ($j = 1, 2, ..., J$), and can be estimated by

$$\texttt{SIR}_{ij} = \frac{Y_{ij}}{E_{ij}} = \frac{\sum_{g=1}^{G} Y_{ijg}}{\sum_{g=1}^{G} E_{ijg}}, \tag{2.2}$$

where Y_{ij} and E_{ij} are now denotes the observed and expected number of cases for the ith area at the jth time, respectively, whereas E_{ijg} is the expected number of cases in age group g for the study population i and time j, and hence it will be simply: $E_{ijg} = \left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{J} Y_{ijg}}{\sum_{i=1}^{N}\sum_{j=1}^{J} n_{ijg}}\right) n_{ijg}$, and then $E_{ij} = \sum_{g=1}^{G} E_{ijg}$. In general, standardization seeks to remove variations in summary measures (rates or proportions) due solely to those known risk factors, so any remaining differences suggest the presence of risk differences other than those adjusted for [1, 6]. Although the SIR gives a useful exploratory information about the spatial distribution of the disease risk but has some major disadvantages that it considers each area in the given region independently and does not account any pattern of spatial structure that will appear in the data. Furthermore, some areas may have a very low or rare expected value mostly for those with a small population such as Herstappe, and the SIR value, in this case, would be susceptible to small random changes in the observed incidence counts [1].

2.1.2 Tests for overdispersion

Before conducting any analysis of the presence of spatially structured correlation or unstructured variation, it is more advisable to assess the heterogeneity of the relative risks. We have the observed and standardized expected number of cases, significant differences between these two quantities was carried out using a chi-square test defined by the following test statistic:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left(Y_i - E_i\right)^2}{E_i} \tag{2.3}$$

where Y_i and E_i are the observed and expected number of cases in the ith municipality, and this test statistic, asymptotically follows a chi-square distribution with $N - 1$ degrees of freedom. Another method we carried out to assess the presence of over-dispersion in the count data and in turn heterogeneity in the relative risks is that was proposed by Potthoff and Whittinghill's [27]. In this approach, the alternative hypothesis assumed that the observed incidence counts are distributed from a Negative Binomial distribution, and the

relative risks are drawn from a gamma distribution with mean λ and variance σ^2. This can be stated as

$$H_0 : \theta_1 \ = \ \theta_2 \ = \ ... \ = \ \lambda$$

$$H_1 : \theta_i \sim \texttt{Gamma}(\lambda^2/\sigma^2, \lambda/\sigma^2).$$

Then the test statistic is given by

$$PW = E_+ \sum_{i=1}^{N} \frac{Y_i\big(Y_i - 1\big)}{E_i}. \tag{2.4}$$

The asymptotic distribution of this statistic is Gaussian with mean $Y_+(Y_+ - 1)$ and variance $2N * Y_+(Y_+ - 1)$ [29]. Where Y_+ and E_+ are used to denote the global observed and expected number of cases, respectively.

2.1.3 Spatial autocorrelation

We now move from test of heterogeneity that may be related to unknown factors, which can be geographically structured or unstructured variables, to methods that summarize the extent of observed spatial similarity between nearby areas. A crucial aspect in many applications of the spatial and spatio-temporal data is that the presence of dependence, thus correlation, will exist between spatial units. The main objective for correlation is to measure how strong the tendency is for observations from nearby areas to be more (or less) alike than observations from areas farther apart. Following the Griffith [7] definition, the term spatial autocorrelation implies correlation among the same type of measurement (incidence of prostate cancer disease in this study) taken at different locations, and expresses the amount of spatial dependence. Most global indices of spatial autocorrelation are based on the similarity/dissimilarity of the data values and the spatial proximity between locations such as areas, counties or municipalities. Let S_{ij} denote measure of how similar or dissimilar the data values are at locations i and j, and let $\boldsymbol{W} = \{w_{ij}\}$ denote a spatial proximity matrix (also called spatial connectivity or spatial weight matrix). The (i, j)th element of $\boldsymbol{W}$, denoted w_{ij}, describe the spatial proximity between locations i and j, for i and $j = 1, ..., N$. In other words, w_{ij} can be viewed as weights quantifying the spatial dependence between neighborhood areas i and j. In this setting the measure of similarity S_{ij} depends on random variables defining observations, while the w_{ij} are fixed quantities or weights based on the underlying geography of the regions [1, 8]. Furthermore, the spatial weights matrix is often row-standardized which makes each row sum in the matrix is equal to one, the individual values w_{ij} are proportionally represented. Row-standardization of $\boldsymbol{W}$ is desirable so that each neighbor of an area is given equal weight and the sum of all w_{ij} (over j) is equal to one [1, 8].

The neighborhood structure or the spatial proximity matrix can be defined in different ways, such as based on sharing boundaries, the distance between spatial locations, and nearest neighbors. Since we have irregular areal data, sharing boundaries as proximity measures has been used which is the simplest method and necessarily results in a symmetric spatial proximity matrix. Based on sharing boundaries, the (i, j)th elements of $\boldsymbol{W}$ can be defined as a binary connectivity matrix, that is, $w_{ij} = 1$ if municipalities i and j share some common boundary, and $w_{ij} = 0$, otherwise [1]. In this case the symmetric properties of $\boldsymbol{W}$ are established

because of that $w_{ij} = w_{ji}$ and its diagonal elements equal to zero, being the similarity of the ith region with itself $w_{ii} = 0$.

In general, spatial autocorrelation measures deal with covariation or correlation between neighboring observations of a variable. And thus, it gives us two types of information: similarity of observations and similarity among locations. Two standard statistics that are used mostly to measure strength of spatial association for the case of areal units and interval variables are Moran's I and Geary's C. Moran's I [9] is widely used, and variations of it relate to likelihood ratio tests and best invariant tests for particular models of correlation for normally distributed random variables [7, 8, 10]. The term global shows the fact that all association measures were included in the computation of spatial autocorrelation. Therefore, based on the information described above, Moran's I as a measure of global indexes of spatial autocorrelation with spatial proximity and similarity between areas i and j given as follows:

$$I = \left(\frac{1}{S^2}\right) \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} S_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}}, \tag{2.5}$$

where $\bar{y} = \frac{\sum_{i=1}^{N} y_i}{N}$ is the overall average for regional count, $S^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N}$ and $S_{ij} = (y_i - \bar{y})(y_j - \bar{y})$ measures similarity of incidences between municipalities i and j.

The null hypothesis in the spatial autocorrelation tests states that the near-by regions do not affect one another implies that there is independence and spatial randomness in the data. The alternative hypothesis in contrast to the null implies that there is spatial association or dependence among the areas. To be specific to our study, the research hypothesis states that the near-by municipalities in Limburg region have an association or dependence on the disease risk of prostate cancer. Spatial autocorrelation in the near-by areas is considered to be present when the test statistic such as I computed for a particular pattern takes on a larger value, compared to what would be expected under the null hypothesis of no spatial association [1, 8]. Moreover, spatial autocorrelation testing can be conducted based on randomization or the normal approximation. Randomization distribution can be obtained by reassign data values among the N fixed areas. As a result if I lies in the tails of this distribution, we can reject the assumption of independence. Assuming that the y_i's are observations on random variables Y_i whose distribution is normal, then I has a sampling distribution that is asymptotically normal under the null hypothesis with mean $E(I) = -1/(N-1)$ and variance is given by

$$Var(I) = \frac{N^2 W_1 - N W_2 + 3 W_0^2}{(N-1)(N+1) W_0^2} \; - \; \left(\frac{1}{N-1}\right)^2,$$

with $W_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}, W_1 = 1/2 \sum_{i=1}^{N} \sum_{j=1}^{N} (w_{ij} + w_{ji})^2$, and $W_2 = \sum_{i=1}^{N} (w_{i+} + w_{+j})^2$, with $w_{i+} = \sum_{j=1}^{N}$ and $w_{+j} = \sum_{i=1}^{N} w_{ji}$. Applying the normality assumption under independence with large N, we compare $Z-$ score $= \frac{I - E(I)}{\sqrt{Var(I)}}$ to a standard normal distribution. When the observed value of Moran's I is greater than the expected value $E(I)$ points to a positive spatial autocorrelation (clustered pattern), while a value of Moran's I that is below the expected value indicates a negative spatial autocorrelation (regular

pattern). Note that extremely small numbers in the lower range of cases are observed in a very small town, such as Herstappe and high values partly attributed to highly populated municipalities, such as Hasselt or Genk. This may reveal that the observed spatial similarity in regional deviations from the average regional count may simply be due to variations in the regional at-risk population size. To overcome such problem partly, we replaced counts with incidence proportions, to remove the impact of population heterogeneity in some amounts [1].

2.2 Bayesian spatial modeling

In the conventional likelihood analysis, the individual contribution to the likelihoods is assumed to be independent, and this likelihood can be derived as a product of probabilities [2]. The presence of spatial correlation (or autocorrelation) may have an impact on the structure and form of likelihood models that are assumed for spatial or spatio-temporal data in general. Hence, the independence criterion may not be met in the spatial, and in turn other different approaches would be needed to make an inference.

A traditional Poisson distribution is a popular frequentist approach for modeling count data. However, it is constrained by equidispersion assumption which will not appropriate to model the real data which exhibit overdispersion (underdispersion), and also do not account for the correlation structure. In particular to disease mapping, the maximum likelihood estimator for the first and second moment in the traditional Poisson model is SIR and $SIR \times 1/E_i$, respectively [1,2,12]. Note that this approach can yield large changes in estimate with relatively small changes in expected value since they are inversely proportional. For instance, when a (close to) zero expectation is found, the standardized incidence ratio will be very large for any positive count, and also the zero SIRs do not distinguish variation in the expected count. In conclusion, SIR is a saturated estimate of relative risk, and hence not parsimonious. Thus, simple generalized regression models such as traditional Poisson model and Binomial model often do not capture the extent of variation present in count data. Overdispersion due to unobserved confounders or spatial correlation will usually not be captured by simple covariate models, and often it is appropriate to include some additional term or terms in a model, which can capture such effects [1, 2]. By considering different assumptions for the nature of spatial data, common hierarchical Bayesian spatial modeling has been applied in this paper to produce a map of smoothed standardized incidence rates. Initially, overdispersion or extra-variation can be accommodated by either inclusion of a prior distribution for the relative risk (such as a Poisson-gamma model) or by extension of the linear or non-linear predictor term to include an extra random effect (log-normal model). To consider both structured and unstructured spatial autocorrelation it is often important to include a variety of random effects in a model, this has been discussed using a conditional autoregressive (CAR) convolution model [13].

2.2.1 Poisson-gamma model

Let θ_i denote the true relative risk in the ith area and assumed as random effects, to allow borrowing of strength across neighborhood areas. Assume that the number of cases of prostate cancer disease in each area follow a Poisson distribution (likelihood), i.e., $y_i|\theta_i \sim \texttt{Poisson}(\lambda_i)$ with $\lambda_i = E_i\theta_i$. In the disease

mapping context a commonly assumed prior distribution for the θ_i in a Poisson likelihood model is a gamma distribution and the resulting model is the Poisson-gamma model [2, 12]. The posterior distribution is proportional to the product of the likelihood model and the prior distribution for θ_i:

$$P(\boldsymbol{\theta}|\boldsymbol{y}) \ \propto \ L(\boldsymbol{y}|\theta)g(\theta), \tag{2.6}$$

where $g(\theta)$ is a gamma distribution with parameters α, β, that is $G(\alpha, \beta)$, and $L(\boldsymbol{y}|\theta) = \prod_{i=1}^{N}\{(\lambda_i)^{y_i}\exp(\lambda_i)\}$ is the likelihood assumed dependent on the data. The expectation (μ) and variance (σ^2) for the prior distribution would be respectively α/β and α/β^2, and could allow for extra variation or overdispersion which do not assumed in the traditional Poisson model. In this case assuming that the two parameters in the prior distribution are fixed and known, the relative risk θ_i follows a gamma distribution $G(\alpha+y_i, \ \beta+E_i)$ considered as posterior distribution which emerges in closed form to the conjugacy of the gamma prior with the Poisson likelihood. Thus, a suitable point estimate of the relative risk θ_i might be the posterior mean,

$$E(\theta_i|y_i) = \frac{\alpha + y_i}{\beta + E_i} = \frac{y_i + \frac{\mu^2}{\sigma^2}}{E_i + \frac{\mu}{\sigma^2}} = \frac{E_i(\frac{y_i}{E_i})}{E_i + \frac{\mu}{\sigma^2}} + \frac{(\frac{\mu}{\sigma^2})\mu}{E_i + \frac{\mu}{\sigma^2}} = w_i\texttt{SIR}_i + (1 - w_i)\mu, \tag{2.7}$$

where $w_i = E_i/\{E_i + (\mu/\sigma^2)\}$, which is found between zero and one inclusively. In (2.7), we realized that the Bayesian point estimate is a weighted average of the data based $\texttt{SIR}_i$ for area i, and the prior mean μ. Additionally, the upper and lower values of this w_i in (2.7) will be used in providing the informative/non informative behaviors of likelihood and prior distribution in the estimation of posterior mean. On one hand, the Bayesian point estimate is approximately equal to $\texttt{SIR}_i$ when w_i is close to 1, that is when the expected cases (E_i) is big, so the data are strongly informative, or when the prior variance σ^2 is big, so the prior is weakly informative. On the other hand, it will be approximately equal to μ when w_i is close to 0 (i.e., when E_i is small, so the data are sparse, or when σ^2 is small, so that the prior is highly informative [12]. Note that a Poisson-gamma model accounts only spatially-unstructured heterogeneity (UH), and hence have its own drawbacks due to difficulty to include covariates under the gamma prior. It also fails to capture the spatially structured correlated heterogeneity (CH) that will be found between the near-by areas.

2.2.2 Poisson-lognormal model

Another common popular method in disease mapping to account for the uncorrelated heterogeneity is a Poisson-lognormal model, by assuming a direct linkage between relative risk θ_i and linear or non-linear predictors. A broad discussion can be found in [2, 12]. In this case the number of cases found in each area is still assumed to follow a Poisson distribution with parameter λ_i. Now the parameter of interest θ_i is formulated by the function of random effect ν_i and a vector of fixed covariates ($\boldsymbol{x}_i$) common to all counties or areas. This can be expressed in a form of hierarchical model for relative risk as

$$\theta_i = \exp(\alpha_0 \ + \ \boldsymbol{x_i'\beta} \ + \ \nu_i).$$

This implies that

$$\log(\theta_i) = \alpha_0 \ + \ \boldsymbol{x_i'\beta} \ + \ \nu_i, \tag{2.8}$$

where ν_i is a normal distributed random variable with mean 0 and variance σ_ν^2 representing the heterogeneity random effect (i.e, it has a zero mean Gaussian prior distribution $N(0, \sigma_\nu^2)$), capturing extra-Poisson variability in the log-relative risks. Whereas $\boldsymbol{x_i}$ are explanatory spatial covariates at region-level, having parameter coefficients $\boldsymbol{\beta}$. However, in this specific study we do not have known confounders or possible fixed covariates, and hence the formulation in (2.8) has been applied to incorporate unmeasured confounders via the use of random effects only (i.e., $\log(\theta_i) = \alpha_0 + \nu_i$). Therefore, the random effect ν_i represents the residual or unexplained variability after adjusting the overall mean risk α_0 in the log-relative risks for the ith area, which captures unmeasured areal level covariates. To formulate the full Bayesian hierarchical model, we assigned a vague prior distribution for the intercept α_0 and hyperprior distribution on the precision parameter, $1/\sigma_\nu^2 = \tau_\nu \ \sim \ \texttt{Gamma}(\alpha, \beta)$. Note that the expected count E_i is considered as an offset term in all the fitted models on the natural log scale.

2.2.3 CAR convolution model

In the case of Poisson-lognormal model, the random effect term was incorporated to take into account the unstructured heterogeneity (UH) or overdispertion due to non-geographical variables. There are also other possibilities to include both structured spatial correlation (CH) and UH at the same time. The fully Bayesian convolution model can incorporate both unstructured heterogeneity and spatially structured correlation by including two random effect terms in the model. This was proposed by Besag *et al.* [14] also known as Besag-York-Mollie (BYM) model and widely used in disease mapping. Thus, the logged relative risk θ_i in convolution model becomes:

$$\log(\theta_i) \ = \ \alpha_0 \ + \ \nu_i \ + \ u_i, \tag{2.9}$$

where ν_i and u_i represent UH and CH terms for the ith area, respectively. As we already mentioned in section 2.2.2, the UH terms are assumed to follow normal distribution $\nu_i \sim (0, \sigma_\nu^2)$. On the other hand, the structured correlated heterogeneity terms are assumed to follow an intrinsic conditional autoregressive (CAR) model [13]. The purpose of using an intrinsic CAR prior in the disease mapping is to borrow strength of information between neighbours, yielding a smoothed map for the risk. This is an example of a Markov random field (MRF) prior distribution is assumed, and would be used where the conditional mean of the area effect is based only on its neighborhood areas [15]:

$$u_i | u_{j,i\neq j} \ \sim \ N(\bar{u}_{\delta_i}, \ \sigma_{u,i}^2),$$

$$\bar{u}_{\delta_i} \ = \ \frac{\sum_{j\in\delta_i} u_j}{n_{\delta_i}} = \frac{\sum_{j\neq i}^{N} w_{ij}\delta_j}{\sum_{j\neq i}^{N} w_{ij}},$$

$$\sigma_{u,i}^2 = \frac{\tau_u^{-1}}{n_{\delta_i}},$$

where δ_i is a neighborhood of the ith area, and n_{δ_i} is the number of neighborhood for the ith area, i.e., $\sum_{j\neq i}^{N} w_{ij}$ (see section 2.1.3), $\bar{u}_{\delta_i}$ is the mean of the neighboring u_j values (where $j \in \delta_i$) and τ_u is a precision parameter which controls the degree of smoothing that is the variability of the random effect (for CH) conditional upon the random effects in the neighboring areas [15, 16]. The conditional expectation is the average of the random effects in neighboring areas, while the conditional variance is inversely proportional to the number of neighbors. In this case we can realized that a county with many neighbors has smaller variance. In other words, if the random effects are spatially correlated, then the more neighbors an area has the more information there is from its neighbors about the value of its random effect. Furthermore, adjacency-based weights are used, hence, $w_{ij} = 1$ if area j is adjacent to area i, and 0 otherwise. The full Bayesian hierarchical specification for the convolution model with an intrinsic CAR would be summarized by

$$
\begin{aligned}
y_i|E_i, \theta_i \;&\sim\; \texttt{Poisson}(E_i\theta_i) \qquad (2.10)\\
\log(\theta_i) &= \alpha_0 \;+\; \nu_i \;+\; u_i,\\
\alpha_0 \;&\sim\; N(0,\; \sigma_0^2)\\
\nu_i \;&\sim\; N(0,\; \sigma_\nu^2)\\
u_i|u_{j,i\neq j} \;&\sim\; N(\bar{u}_{\delta_i},\; \sigma_{u,i}^2).
\end{aligned}
$$

The above expression (2.10) reflects that prior distribution would be assigned for all parameters in the model including hyper-priors to the hyper-parameters $\sigma_u^2,\ \sigma_\nu^2,\ \alpha_0$; and posterior sampling of these parameters via Markov chain Monte Carlo (McMC) algorithms [15].

Although, the area-specific random effects in (2.9) can be decomposed into clustering or correlated heterogeneity and uncorrelated heterogeneity components, however, these random effects are not identified, but we are usually interested in the total unobserved confounding, the sum of the two effects is considered as the important component.

2.3 Bayesian spatio-temporal modeling

Up to this point, the focus has been based upon considering the spatial data as being only cross-sections. In this case, a temporal dimension has been added to the spatial dimension. Thus, we are shifting now from spatial disease mapping to spatio-temporal modeling to describe the pattern of the disease risks in space and time simultaneously. Several models are developed based on the concept of hierarchical Bayesian models. Mostly the methods for count data can be extended from the common spatial models by adding linear as well as a nonlinear time dimension. Usually, in Bayesian spatio-temporal disease mapping, three random effects are involved; the spatial random main effect, temporal random main effect, and the interaction between the two main effects or spatio-temporal interaction random effect. Furthermore, the two random main effects can be classified into a structured component and unstructured component to represent the dependence and

heterogeneity variations of disease risks in space and time, respectively [23]. In this study, some basic models starting from simple linear parametric model to more complex or advanced models through the concept of hierarchical Bayesian approaches have been discussed.

2.3.1 Parametric linear trend model

In disease mapping the problems found such as in generalized linear model has been solved by using a Bayesian approach which accounts random effects through a prior distribution. One of the earliest spatio-temporal models for areal data was investigated by Bernardinelli *et al* [16]. This model is appropriate to estimate which areas are exhibiting increasing or decreasing (linear) trends in the response over time. The linear trend model with the area effects (ϕ_i and δ_i) considered as random effects, through a prior distribution based on prior belief on the relative risk can be expressed as

$$\log(\theta_{ij}) \;=\; \alpha_0 \;+\; \phi_i \;+\; (\beta \;+\; \delta_i) * t_j, \tag{2.11}$$

where α_0 is an intercept to denote overall rate, ϕ_i is a spatial random effect, β is a linear trend for time t_j, and δ_i is a spatio-temporal interaction random effect. There is no spatial trend and temporal random effect, but it only has a simple linear time trend and spatial random effects. The area effects here can be used to formulate a spatially unstructured extra-Poisson variation (heterogeneity), or to a spatially structured variation (clustering), the model may even contain both types of variations such as $\phi_i = u_i + \nu_i$ [16,17]. The sign of δ_i has its own meaning that a value of $\delta_i < 0$ implies that the county-specific trend is less steep than the mean trend, whilst a value of $\delta_i > 0$ indicates the county-specific trend is steeper than the mean trend β [16].

We used the same prior distributions, an intrinsic conditional autoregressive (CAR) for both random effects, structured random main effect u_i and the interaction effect δ_i in the clustering model. On one hand, the heterogeneity of the trend, δ_i is assumed to be sampled priorly from a normal distribution with mean 0 and variance σ^2, that is,

$$(\delta_i|\delta_j, j \neq i, \sigma_\delta^2) \sim \; N(0, \sigma_\delta^2). \tag{2.12}$$

The hyper-parameter σ_δ^2 was used to control the amount of variability in the linear time trends [16]. On the other hand, δ_i were assumed to be sampled from normal distribution with mean μ_i and variance σ_i^2, which allows that the mean can depend on the neighboring δ_is. That is,

$$(\delta_i|\delta_j, j \neq i, \sigma_\delta^2) \sim \; N(\mu_i, \sigma_i^2), \tag{2.13}$$

where

$$\mu_i \;=\; \frac{\sum_{j\neq i}^{N} w_{ij}\delta_j}{\sum_{j\neq i}^{N} w_{ij}},$$

$$\sigma_i^2 \;=\; \frac{\sigma_\delta^2}{\sum_{j\neq i}^{N} w_{ij}}$$

where w_{ij} as defined in section (2.1.3) hence equal to 1 if i and j counties are geographically adjacent and 0 otherwise. In this case the amount of variation of the δ_is can be controlled by the variance parameter σ_i^2. Although the parameter σ_δ^2 is included in both models (2.12 and 2.13), it has rather different interpretation. In the unstructured heterogeneity model, it represents the overall or marginal variability of the area random effects, whilst in the clustering model, it controls the variability of the random effect conditional upon the random effects in the neighboring areas. Thus, the variability in the structured model depends upon the number of neighboring areas as well [16]. Finally, a conjugate Inverse-Gamma, Gaussian and a flat hyper-prior distributions have been specified for the random effect variances $(\sigma_u^2, \sigma_\nu^2, \sigma_\delta^2)$, the overall slope parameter (β), and the overall intercept term (α_0), respectively. Although information is shared in both space and time in this parametric form of model, it is too restrictive to be a linear function of time, and seems to be inappropriate for a long period [24].

2.3.2 Smooth temporal evolution models

In the model specified above, a linearity constraint is imposed on the differential temporal trend δ_i. However, it is also possible to reflect it using a dynamic formulation for the linear predictor, such as using models proposed by Knorr-Held [23], and Martiínez *et al* [25], to allow for time trends without restricting to any predefined shape. In the case of smooth temporal evolution models, the evolution of the estimated risk in each area is a smooth function of time. Since temporal evolution is not restricted to any predefined shape, therefore information can be shared in time [24]. In the case of Knorr-Held models, the structured correlated and unstructured heterogeneity spatial and temporal main effects, and the interaction between space and time main effects are accounted. The Knorr-Held model used to express space and time interaction to the log-relative risk for the ith area in the jth time is given by

$$\log(\theta_{ij}) \;=\; \alpha_0 \;+\; u_i \;+\; \nu_i \;+\; \gamma_j \;+\; \phi_j \;+\; \delta_{ij} \tag{2.14}$$

where α_0 is an overall risk level, u_i and ν_i, respectively, represent spatially correlated and independent heterogeneity random main effects. In similar manner, γ_j and ϕ_j denote temporally structured and unstructured main effects, respectively, whereas the parameters δ_{ij} are included to represent the interaction for the ith space and jth time, and capture the variation that cannot be explained by the four main effects. A flat prior for an overall risk level α_0, an intrinsic CAR prior for u_i, random walk with independent Gaussian increment for γ_j, and a Gaussian prior distributions for ν_i and ϕ_j were assumed. Mathematically the prior for the correlated random spatial effects is expressed as,

$$p(u|\tau_u) \propto \exp\Big(-\frac{\tau_u}{2} \sum_{i\sim l} (u_i - u_l)^2 \Big),$$

and for the correlated temporal effect as

$$p(\gamma|\tau_\gamma) \propto \exp\Big(-\frac{\tau_\gamma}{2} \sum_{j=2}^{T} (\gamma_j - \gamma_{j-1})^2 \Big)$$

where τ_u and τ_γ are the precision matrix and assumed to follow gamma hyper-prior distribution to consider full Bayesian formulation.

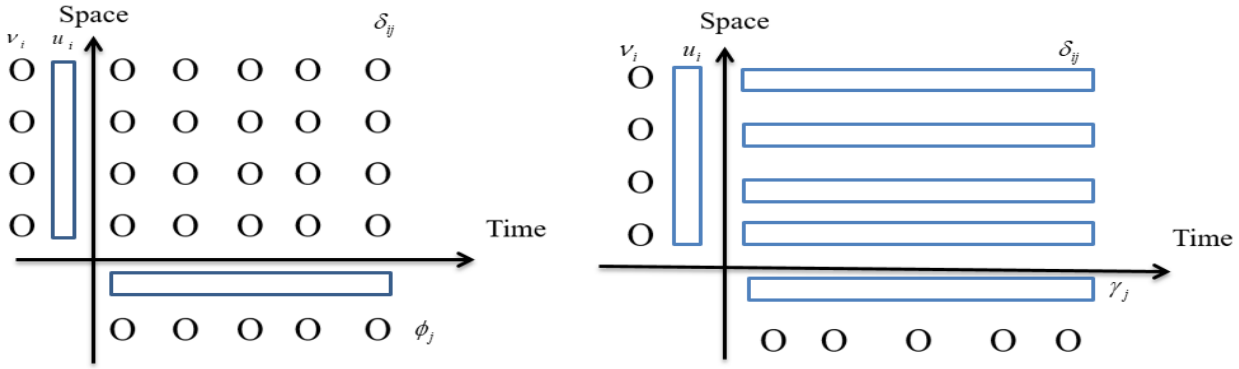


Figure 2.2: Symbolic representation for Type I (left panel) and Type II (right panel) interactions. Circles represent prior independence, rectangles represent prior dependence.

Note that in model (2.14), if $\delta_{ij} = 0$ for all i and j, then separable space and time effects can be reflected, otherwise four types of spatio-temporal interaction can be obtained [23, 26]. Type I interaction is raised when the two unstructured main effects ν and ϕ are assumed to interact, and resulting a priori independent interaction parameters δ_{ij}. Similarly, a Type II interaction was formulated between spatially unstructured heterogeneity and structured temporal main effects; Type III interaction, a combination of structured spatial and unstructured temporal main effects; and lastly Type IV can be obtained by interacting the two structured random main effects (u_i and γ_j). Further detailed explanation can be found from the author [23]. Symbolic representation for Type I and Type II interactions is given in Figure 2.2.

We also employed the method that was proposed by Martínez *et al* [25] illustrated as an autoregressive approach to the spatio-temporal disease mapping. The log-relative risk for the first time period is given by the sum of an intercept and the two random effects, that is

$$\log(\theta_{i1}) \;=\; \alpha_0 \;+\; \gamma_1 \;+\; (1-\rho^2)^{-1/2}.(u_{i1}+\nu_{i1}),\;\; i = 1, 2, ..., N.$$

$$\nu_{i1} \sim N(0, \sigma_\nu^2), \quad \gamma_1 \sim \texttt{CAR.normal}(\sigma_\gamma^2), \tag{2.15}$$

$$\boldsymbol{u_1} \;=\; (u_{11}, ..., u_{N1}) \sim \texttt{CAR.normal}(\sigma_u^2)$$

where α_0 denotes the overall risk level, γ_1 represents the mean deviation of the risks in the first period from the overall mean level, and u_{i1} and ν_{i1} are used to express, respectively, the spatially structured and unstructured random effects in the first period. For the subsequent time periods the log-relative risk at the jth time point was formulated as a function of their estimates in previous periods.

$$\log(\theta_{ij}) \;=\; \alpha_0 \;+\; \gamma_j \;+\; \rho * \big[\log(\theta_{i(j-1)}) - \alpha_0 - \gamma_{j-1}\big] + (u_{ij} \;+\; \nu_{ij})$$

$$\nu_{ij} \sim N(0, \sigma_\nu^2), \quad i = 1, 2, ..., N \quad j = 2, 3, ..., T \tag{2.16}$$

$$\boldsymbol{u_j} \;=\; (u_{1j}, ..., u_{Nj}) \sim \texttt{CAR.normal}(\sigma_u^2)$$

$$\boldsymbol{\gamma} = (\alpha_1, \alpha_2, ..., \gamma_T) \sim \texttt{CAR.normal}(\sigma_\gamma^2)$$

where ρ is used to define the temporal correlation structure, u_{ij} and ν_{ij} are defined as the structured and unstructured spatio-temporal interactions, respectively. A flat prior for the intercept (α_0), uniform prior distribution for ρ, and Inverse-Gamma(a, b) for the variance parameters were used as prior distributions, hence the full Bayesian formulation is kept as well. Finally, the log-relative risk evolution over time and every area in (2.16) has been defined as a linear combination of such a value in the previous period instead of predefined parametric form such as linear or quadratic trend, and hence smoothed temporal evolutions of disease risk can be described.

2.4 Estimation and computation

Once the posterior distribution has been formulated using the product of likelihood and prior distributions, it is expected to evaluate the posterior distributions, which is usually done via posterior sampling. For some simple hierarchical models such as in the Poisson-gamma model, posterior distributions can be simulated directly or by analytical computation from the given formula. For example, an estimate of posterior mean and variance can be directly estimated for the Poisson-gamma posterior distribution (see in section 2.2.1). However, since in most of the Bayesian hierarchical models in disease mapping have two or more levels (hierarchy), and hence the complexity nature of the posterior distributions of the parameters need sampling algorithm [2, 18, 19]. That is when the posterior distribution is not in a closed form, different simulation techniques can be used to approximate them. Let $g_i(\theta_i)$ defined as the prior distributions for the p components of $\boldsymbol{\theta}$ for $i = 1, 2, ..., p$, and $\boldsymbol{L(y|\theta)}$ denotes the likelihood distributions for the data $\boldsymbol{y}$. The posterior distributions of $\boldsymbol{\theta}$ and $\boldsymbol{y}$ can be defined as

$$P(\boldsymbol{\theta}|\boldsymbol{y}) \propto \; \boldsymbol{L(y|\theta)} \prod_{\boldsymbol{i}} g_i(\theta_i). \tag{2.17}$$

Therefore, the objective is to generate a sample from the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{y})$ using Markov chain Monte Carlo (McMC) methods. The default Gibbs sampler algorithm has been used in WinBUGS with blocking mode on. Markov chain provides a sample from posterior distribution and that the summary measures calculated from the chain consistently estimate the required true posterior summary measures [18]. Such summary measures include such as posterior mean, posterior median, and credible intervals. This can be done by applying Markov chain Law of Large numbers (McLLN) and Markov chain Central Limit Theorem (McCLT) [18, 19].

2.5 Assessing convergence of a Markov chain

One possible critical issue in McMC algorithm is convergence for Markov chain which need to check how close to the true posterior distribution. This diagnostic involves convergence to the stationary distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ and verifying the accuracy of the posterior summary measures. This is explained more formally by Lesaffre & Lawson [18] and Robert [19], as convergence in a McMC context is an asymptotic property which implies for a Markov chain that $p_t(\boldsymbol{\theta})$, the distribution of $\boldsymbol{\theta}^t$, grows to the target distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ when t tends to infinity. In verifying the posterior summary measures in regular Monte Carlo setting, for instance, we are concerned with convergence of the empirical average $\hat{\bar{\theta}} = \frac{1}{T}\sum_{t=1}^{T}\theta^t$ to posterior mean ($\bar{\theta}$), where $\hat{\bar{\theta}}$ is used to denote the posterior mean obtained from a McMC chain. Different graphical approaches such as trace plot, autocorrelation plot, and running mean plot, that are readily available software were used for stationarity checking, and the Gelman and Rubin (GR) ANOVA diagnostic as a formal test. When the Markov chain is stationary, θ^{t_1} and θ^{t_2} have the same marginal distribution for arbitrary times or iterations t_1 and t_2. Amongst formal convergence diagnostics in McMC chain, we have used the one proposed by Gelman and Rubin (GR, 1992) which is recommended based on multiple chains with overdispersed starting positions, and used as quantitative and graphical measures as well [18–20]. Suppose that M parallel chains are run for $2T$ iterations, which means that we are taking M dispersed starting points $\theta_m^0 (m = 1, 2, ..., M)$. The first T iterations are discarded and considered as burn-in. Then the M chains (θ_m^t) of length with the remaining T iterations produce means $\hat{\bar{\theta}}_m = \frac{1}{T}\sum_{t=1}^{T}\theta_m^t$ for $m = 1, 2, ..., M$, and the overall McMC mean can be computed by $\hat{\bar{\theta}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\bar{\theta}}_m$. The within- and between- chain variability can be obtained as

$$W = \frac{1}{M}\sum_{m=1}^{M} s_m^2 \qquad B = \frac{T}{M-1}\sum_{m=1}^{M}(\hat{\bar{\theta}}_m - \hat{\bar{\theta}})^2,$$

with $s_m^2 = \frac{1}{T}\sum_{t=1}^{T}(\theta_m^t - \hat{\bar{\theta}}_m)^2$. When there is stationarity in the Markov chain that is when all $\hat{\bar{\theta}}_m$ are unbiased estimates of the true posterior mean, an unbiased estimate of the posterior variance of θ^t in the target distribution can be computed by

$$\widehat{V} \equiv \widehat{Var}(\theta^t|\boldsymbol{y}) = \frac{T-1}{T}W \; + \; \frac{1}{T}B.$$

However, as noted by Gelman, $\widehat{V}$ overestimates the variance of $\hat{\bar{\theta}}_m^t$ because of the large dispersion of the initial distribution, whereas W underestimates this variance, as long as the different sequences $(\hat{\bar{\theta}}_m^t)$ remain concentrated around their starting values [21]. Therefore, GR recommended the ratio $\hat{R} = \frac{\widehat{V}}{W}$ as a convergence diagnostic, which is called the estimated potential scale reduction factor (PSRF). If $\hat{R}$ is substantially greater than 1, further iterations are needed either to increase W_T or to reduce $\widehat{V}$ [18]. In the case of GR-diagnostic, a Gaussian behavior of the sampled parameter values is assumed, and a log-transformation is required for the variance of a normal distribution prior to the application. A nonparametric version of the GR-diagnostic is recommended and computed as the interval-based $\hat{R}_I$:

$$\hat{R}_I = \frac{\texttt{length of total-chain interval}}{\texttt{average length of the within-chain intervals}} \equiv \frac{\widehat{V}_I}{W_I}.$$

The GR-diagnostic was implemented in the R packages CODA and BOA, while a dynamic version of BGR interval diagnostics $\hat{R}_I$ has been done in WinBUGS.

2.6 Model comparison and goodness-of-fit

Until this point, several commonly used models are discussed in both spatial and spatio-temporal analyses. As it has already been mentioned in section 1.2, one of the objectives is making prediction of relative risk of the disease. However, before making any prediction we need to select one best model amongst the fitted Bayesian hierarchical models. Within a likelihood modeling approach, some of the methods that are commonly used as a model choice criterion include Bayesian information criterion (BIC), Akaike information criterion (AIC) and deviance (D). These methods are also widely used in Bayesian hierarchical models. Both AIC and BIC penalize model by incorporating the number of parameters as a penalty for over parametrization. Nevertheless, AIC and BIC have a disadvantage in models with random effects since it is difficult to decide how many parameters are included in the given model. On the other hand, a disadvantage of using the deviance directly is that it does not allow for the degree of parameterization in the model [2,18]. However, the deviance information criterion (DIC) was proposed by Spiegelhalter *et al* [22], and is widely used as an important measure of goodness-of-fit in Bayesian modeling. This is defined as

$$DIC \;=\; 2E_{\boldsymbol{\theta}|\boldsymbol{y}}(D) - D[E_{\boldsymbol{\theta}|\boldsymbol{y}}(\theta)] \;=\; E_{\boldsymbol{\theta}|\boldsymbol{y}}(D) \;+\; pD,$$

where $D(.)$ is the deviance of the model, $pD = E_{\boldsymbol{\theta}|\boldsymbol{y}}(D) - D[E_{\boldsymbol{\theta}|\boldsymbol{y}}(\theta)]$ is the estimated effective number of parameters, and $\boldsymbol{y}$ is the vector of observed data. Thus, the DIC is formulated based on the average deviance and the deviance of the posterior expected parameter estimates [2, 18]. The complexity of each model has been assessed using the effective number of parameters, pD, which is defined as the expected deviance minus the deviance evaluated at the posterior expectations.

Note that the predictive response $p(y_i^{pr}|\boldsymbol{y})$ can be obtained from a converged posterior sample given the current parameters at iteration j ($\boldsymbol{\theta^{(j)}}$).

$$p(y_i^{pr}|\boldsymbol{y}) = \int p(y_i^{pr}|\boldsymbol{\theta^{(j)}})p(\boldsymbol{\theta^{(j)}}|\boldsymbol{y})d\boldsymbol{\theta^{(j)}}. \tag{2.18}$$

The resulting predictive values has marginal distribution $p(y_i^{pr}|\boldsymbol{y})$. Thus, for a Poisson distribution, this can easily generate counts as $y_{ij}^{pr} \leftarrow Pois(E_i\theta_i^{(j)})$ at the jth iteration. As a model selection criterion, we also used the mean squared predictive error (MSPE), which is an average of the item-wise squared error loss given by

$$MSPE_j = \sum_i^N (y_i - y_{ij}^{pr})^2/N \qquad \Rightarrow MSPE = \sum_{j=1}^G MSPE_j/G,$$

where N and G are the number of observations and the sampler sample size, respectively.

2.7 Sensitivity to hyper-prior distributions

One of the key issues in fully Bayesian hierarchical models to disease mapping is the choice of the hyper-prior distributions of the parameters, specifically for the variance components. Most often, the data dominates the prior if the data are large enough and hence it will matter less which values of the hyper-parameters are chosen [28]. However, if we have a scarce data, it is an important consideration to choose a suitable combination of hyper-parameters. Thus, sensitivity analysis has been carried out to different choices of priors or parameters to investigate the influence of the choice of hyper-prior on the estimates of relative risk.

3 Results

3.1 Exploratory data analysis

As we already explained in section 2, the expected incidence counts have been estimated using internal standardization to remove the effect of age on the standardized incidence ratio (SIR). Summary values for observed and expected counts and SIR of prostate cancer in Limburg for spatial only data are shown in Table 3.1. The number of cases of prostate cancer disease in Limburg was 130 on average and ranges from 1 to 596 in the 10 years period 1996-2005. In general, extremely small number of cases have occurred in a very small municipality, such as Herstappe and high values that will be attributed to highly populated cities, such as Hasselt. The average of SIR was 0.997 and ranging from 0.118 to 1.450 for spatial only data. The variance is quite larger than the mean value of observed as well as expected counts which shows the existence of a large amount of overdispersion in the data.

Table 3.1: Summary statistics for the spatial only data.

Variable	Minimum	Mean	Maximum	Variance
Observed	1.000	130.318	596.000	11904.97
Expected	0.823	130.318	527.596	10322.98
SIR	0.118	0.997	1.450	0.056

The distributions of observed and standardized expected counts are shown in Figure 3.1. Although standardization has been done to minimize partly the problem of overdispersion, extra-Poisson variation is still present in the expected counts.

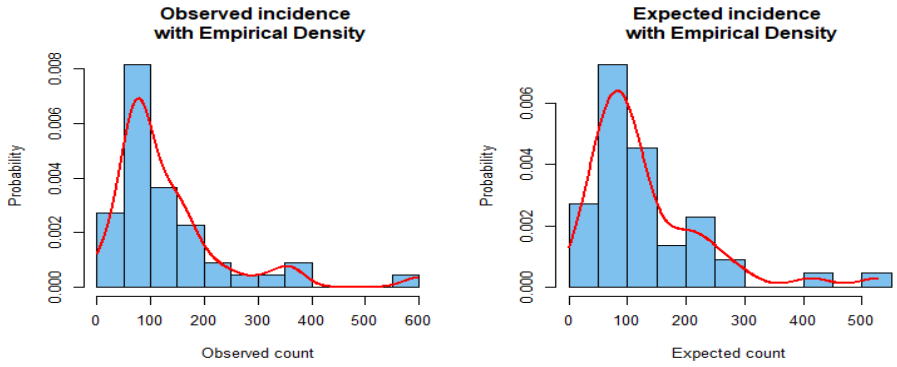


Figure 3.1: Histogram of prostate cancer incidence with empirical density, observed counts(left panel) and expected counts(right panel).

Since standardization only involves known possible factors, age in this case, hence extra-Poisson variation can occur in structured or unstructured ways. The remaining extra-variation would also be explained by correlation through space, called structured variation, and spatially unstructured overdispersion, due to unobserved factors. In summary, we realized that, although standardization can be used to adjust the effect of age, no account is taken of spatial variations and unknown factors on the estimation of disease risk.

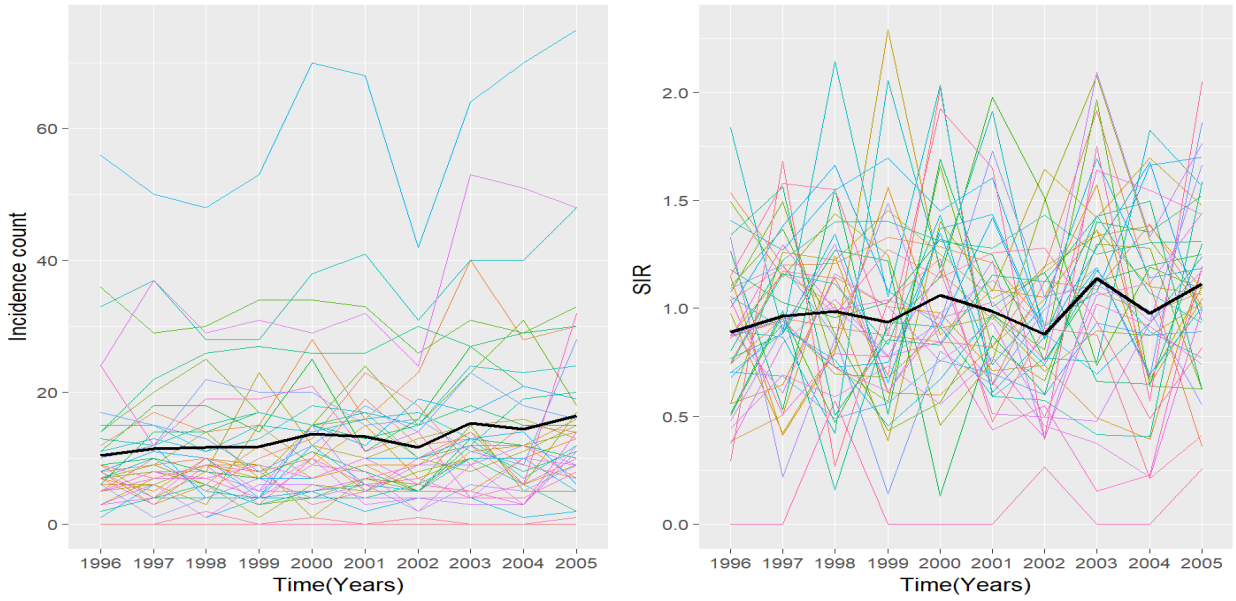


Figure 3.2: Individual trajectories and average evolution of observed incidence counts (left), and standardized incidence ratio (SIR) (right).

It is a good place to look at some simple plots of the incidence count and standardized incidence ratio (SIR) for each county. Figure 3.2 displays the individual municipality trajectories and average evolution of incidence counts and standardized incidence ratio for the period of 1996-2005 years. The individual profile plot clearly shows lots of spatial variation in incidence counts and in turn SIR of prostate cancer in Limburg. These plots also give us some indications about the variability within measurements of the same municipality over time. The average evolution (black line) outlined in the center of each plot depicts slightly an increased rate over time.

The standardized incidence ratio (SIR) estimates of prostate cancer for each municipality at each and combined time points are displayed in the form of maps in Figure 3.3. In the maps, the first row and column show the SIR estimates for only spatial data, called *SIR: only Spatial*. There were extremely low (0) to high (13.34) values of relative risk in the Limburg municipalities. Although the county in question had a low expected count (0.075) and observed count (1), there are several outlying large SMRs with the largest observed SIR equaling 13.34, and it is well known and mentioned that counties with low expected counts tend to be more variable in terms of their SIRs. For instance, an extremely high value of SIR(13.34) is found for Herstappe that has only one case within 10 years measurement. The maps also show a worthwhile spatial as well as temporal heterogeneity. Generally, no immediately obvious patterns emerge. When we compute the standardized expected number of incidence counts for each area to estimate the SIR in section 2.1.1, we have assumed the hypothesis that the risks remain constant in space and time. This shows that the raw estimate of relative risk (SIR) don't have a possibility to reflect the spatial pattern as well as the time trend of relative risks. Thus, SIR is used as an exploratory (noisy) measure of disease risk.

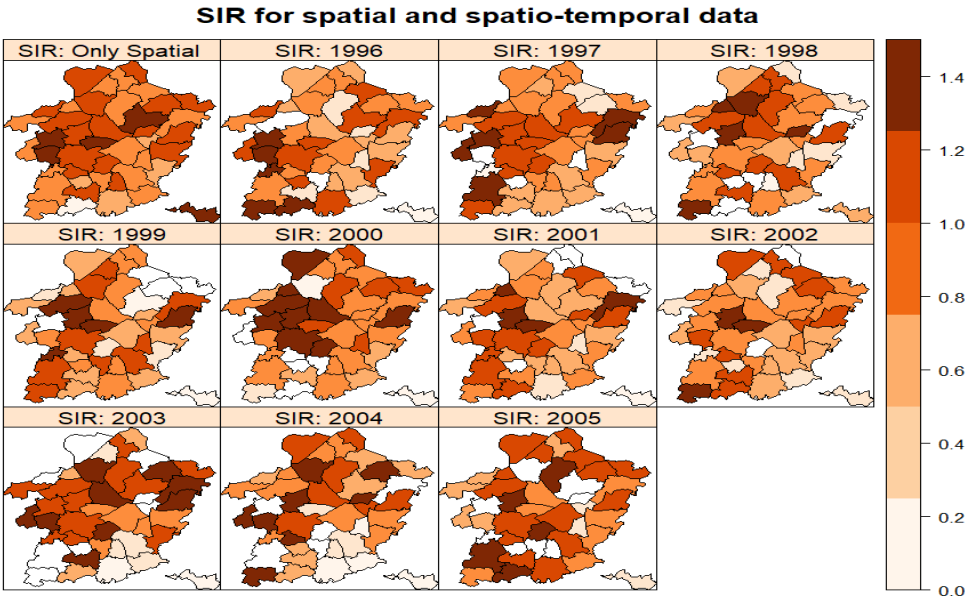


Figure 3.3: Standardized incidence relative risk estimates of prostate cancer in Limburg for the spatial and spatio-temporal data.

Measures of overdispersion and spatial autocorrelation for the spatial data are shown in Table 3.2. Both the chi-square and Potthoff-Whittinghill's test of overdispersion agree with the presence of overdispersion in the incidence counts, and in turn heterogeneity in the relative risk of prostate cancer in Limburg. Moran's I and Geary's C are also shown, used as global indicators of spatial autocorrelation. The estimated Moran's I statistic is 0.2371 with p-value values 0.0027 and 0.0037 under randomization and normality assumptions, respectively. The p-values are found to be less than 0.05 in both assumptions suggesting significant evidence of unexplained spatial autocorrelation in the incidence of prostate cancer after accounting for the effect of age. Similar conclusions can be obtained using Geary's C global indicator showing the presence of spatial association or dependence among the areas.

Table 3.2: Test of overdispersion and global indicators of spatial autocorrelation.

Overdispersion			Spatial autocorrelation			
Test	Statistic	P-value	Indicator	Statistic	P-value	
					randomization	normality
χ^2	232.0479	<0.0001	Moran's I	0.2371	0.0027	0.0037
PW	33957687	<0.0001	Geary's C	0.4469	<0.0001	<0.0001

The histogram in Figure 3.4 shows the distribution of Moran's I values under randomization, where we reassign SIR at random among the 44 municipalities (which makes sense with a spatial association, as noted above). In addition, the observed value of Moran's I =0.2371 is greater than its expected value $E(I) = -0.0232$, showing a positive spatial autocorrelation or clustered pattern, which means that observations from nearby areas tend to be more alike than observations from areas farther apart.

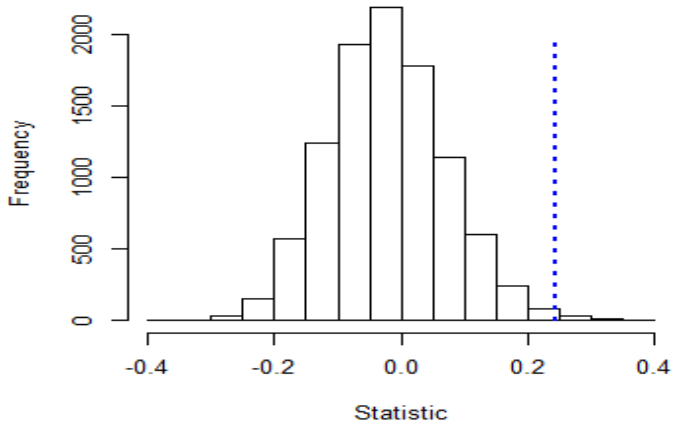


Figure 3.4: Histogram of 9999 simulated values of Moran's I under random permutations of the data, the observed value of Moran's I is marked by a vertical broken line.

In the case of spatio-temporal data, we also checked the spatial autocorrelation of SIR at each time point using Moran's I. These values with corresponding p-values are shown graphically in Figure 3.5. We only found a significant evidence of autocorrelation for the years 1997-1999, 2001 and 2003-2004, that shows the presence of dependence of relative risks in neighborhood areas. In general, testing spatial autocorrelation using Moran's I or Gear's C is usually taken as an exploratory measure, and it is recommended to use the model-based approach as a test of spatial significance.

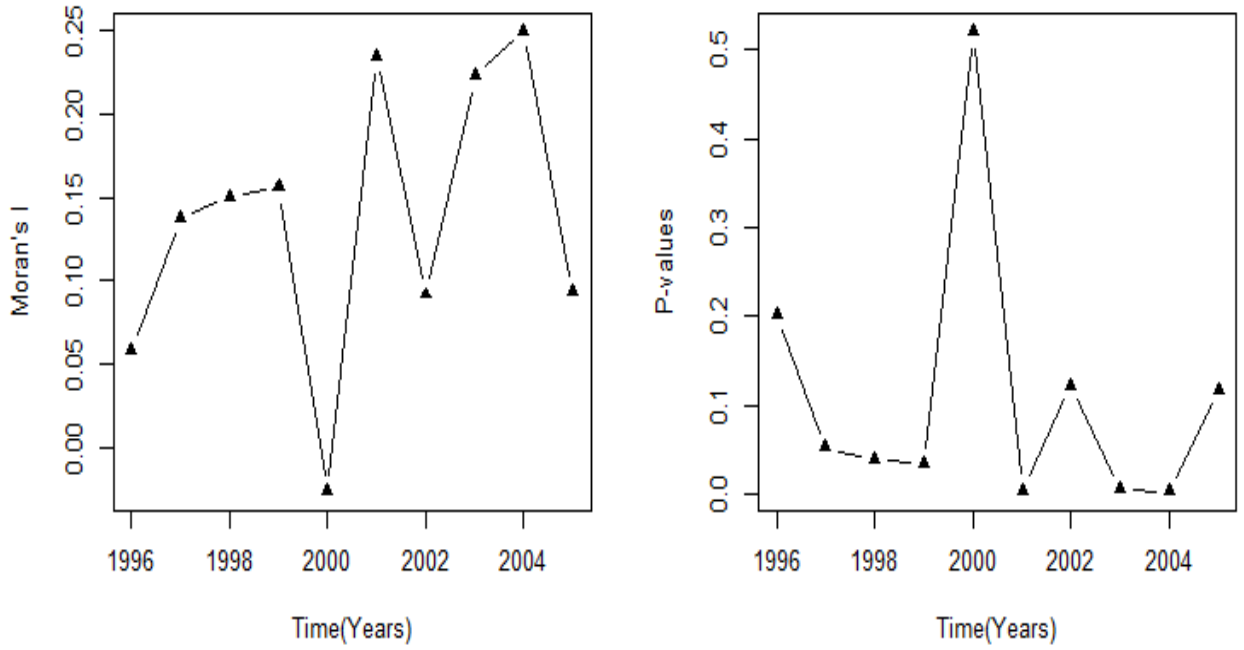


Figure 3.5: Moran's I under random permutations (left), and p-values of Moran's I test (right) for the spatio-temporal data.

3.2 Spatial analysis

As we explained in the methodology part of the study, we applied Poisson-gamma(PG), Poisson-lognormal(PL), CAR (CH), and CAR convolution models for the prostate cancer data summed over all years between 1996 and 2005. We have taken vague prior distributions for all parameters, namely, exp(0.01) for the gamma parameters in PG model, Gamma(0.01, 0.01) to σ_{ν}^{-2} for fitting PL, and Gamma(0.5, 0.005) was taken for σ_{u}^{-2} and σ_{ν}^{-2} in CAR (CH+UH) and CAR(CH) modeling. The MCMC convergence for each model was checked in both graphical plots and formal diagnostic tests. To increase the reliability of the estimates we used a large number of iterations (90,000) with burn-in of 45,000, and only one of 10 them was saved to lower the autocorrelation. A multivariate trace and autocorrelation plots to show the mixing rate, running-mean and density plots to show the stability of the posterior distribution also investigated for the deviance of each model and displayed in Figures B.10 - B.13 in the appendices. The estimated value to the potential scale reduction factor $\hat{R}_I$ for the fitted spatial models are presented in Table 3.4. It has been seen that the values of $\hat{R}_I$ for all parameters were approximately equal to 1 and revealed well converged, and reliable results can be obtained from the fitted models.

Once we have assessed the convergence of a Markov chain for the fitted models, the deviance information criteria (DIC) and the estimated effective number of parameters (pD), the mean absolute predictive error (MAPE), and mean-squared predictive error (MSPE) were used to select the best outperform model. The posterior estimates of these relative measures of goodness of fit for fitting each of the spatial and aspatial models are shown in Table 3.3.

Table 3.3: Posterior estimates of goodness-of-fit statistics for the fitted spatial models.

Type	Model	$\bar{D}$	$\hat{D}$	pD	DIC	MAPE	MSPE
aspatial	PG	335.013	300.792	34.220	369.233	12.099	260.489
	PL	336.627	301.214	35.412	372.039	12.142	262.925
Spatial	CH	325.672	293.892	31.780	357.452	11.827	252.259
	CH+UH	324.693	292.479	32.214	356.906	11.829	252.454

One can see that there is no strong clear preference among the fitted models in terms of MAPE values. Looking at models that only accounts for the unstructured variation the Poisson-gamma model showed less decisive preference than the Poisson-lognormal model. On the other hand, the spatial pattern models the CAR convolution model seems favored to CAR (CH) in terms of DIC and pD value. Generally to investigate the structured and/or unstructured heterogeneity random effect the goodness-of-fit results revealed that the CAR models outperform the other aspatial fitted models since spatially structured and unstructured heterogeneity can be accounted in the estimation of relative risk.

The posterior estimated value for the intercept, the variance of spatially-unstructured random effect (σ_{ν}^2) and the variance of spatially-structured random effect (σ_u^2) with corresponding credible intervals are displayed in Table 3.4. When we compute the ratio of the two random effects for the CAR convolution model results in Table 3.4, $\sigma_u^2/\sigma_{\nu}^2 = 0.0106/0.0000418 = 1345.63$, indicates that a spatially structured variation highly

dominates the unstructured over-dispersion. Which implies that the residual disease risk due to spatially structured variation is larger than unstructured overdispersion. Moreover, the variance of structured random effect in CAR convolution is approximately equal to that of CAR (CH). This can be expected because of that the random effects were assumed to have a lognormal prior distribution in both models.

Table 3.4: Posterior estimates, standard deviation, credible interval, and $\hat{R}_I$ for the spatial models.

Model	Parameter	estimates	sd	2.50%	97.50%	$\hat{R}_I$
PG	a	26.2943	8.1763	14.0597	46.4505	1.0023
	b	26.4386	8.2476	14.0500	46.7300	1.0023
	σ_ν^2	0.0413	0.0130	0.0212	0.0718	1.0022
PL	Intercept	0.9759	0.0363	0.9061	1.0500	1.0009
	σ_ν^2	0.0025	0.0018	0.0006	0.0072	1.0018
CH	Intercept	0.9618	0.0169	0.9288	0.9953	1.0009
	σ_u^2	0.0117	0.0077	0.0031	0.0312	1.0009
CH+UH	Intercept	0.9628	0.0187	0.9264	1.0010	1.0010
	σ_u^2	0.0106	0.0080	0.0021	0.0317	1.0012
	σ_ν^2	4.18E-05	0.0001	1.59E-06	0.0002	1.0011

The posterior estimated value for the relative risk for each municipality obtained from the fitted models are visualized in Figure 3.6. We can see from the map that the disease risk levels are fairly constant across the fitted models with the exception of some municipalities such as Voeren and Wellen. Although the CAR models are considered as the best to fit the data in terms of DIC as well as MSPE values, the spatial patterns of the relative risks were very similar in all the fitted models. Nevertheless, as it has shown in Figure 3.7, the credibility intervals containing the relative risk for the aspatial models are a little bit wider than the spatial models (CAR models) in most of the estimates.

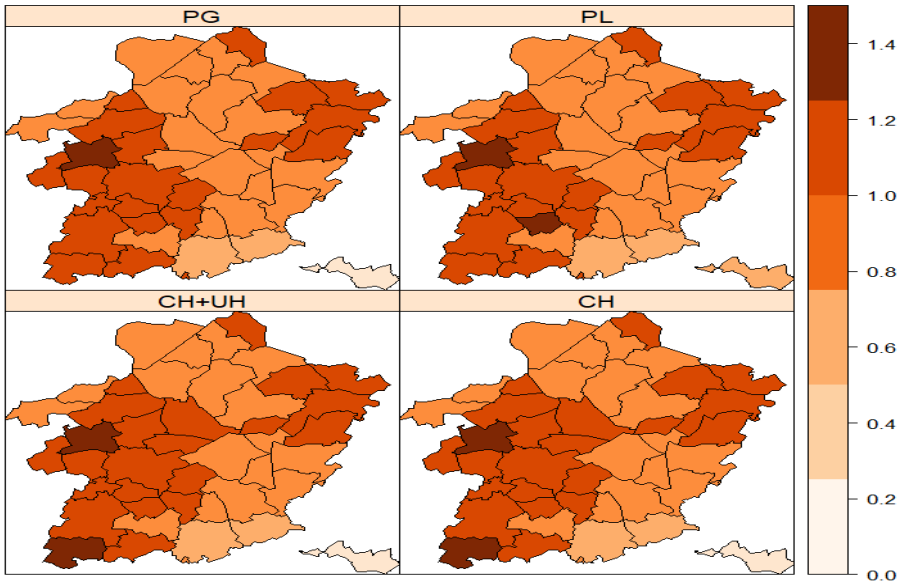


Figure 3.6: Relative risk estimates obtained from the fitted spatial models

The credible intervals of the relative risk obtained from the Poisson-gamma model coincide with Poisson-lognormal model, while the two spatial models also showed similar credible intervals. In the north-east,

south-west and some center of the region, such as Lummen, Heusden-Zolder, Sint-Truiden, Hasselt, Gingelom, Beringen, Wellen, and Maaseik shows elevated prostate cancer rate relative to the other municipalities. In general, the range of the relative risk value in the fitted models was within 0.321 and 1.370, which implies comparing with the SIR, we have obtained a smoother map with fewer extremes in the relative risk estimates in all the fitted models.

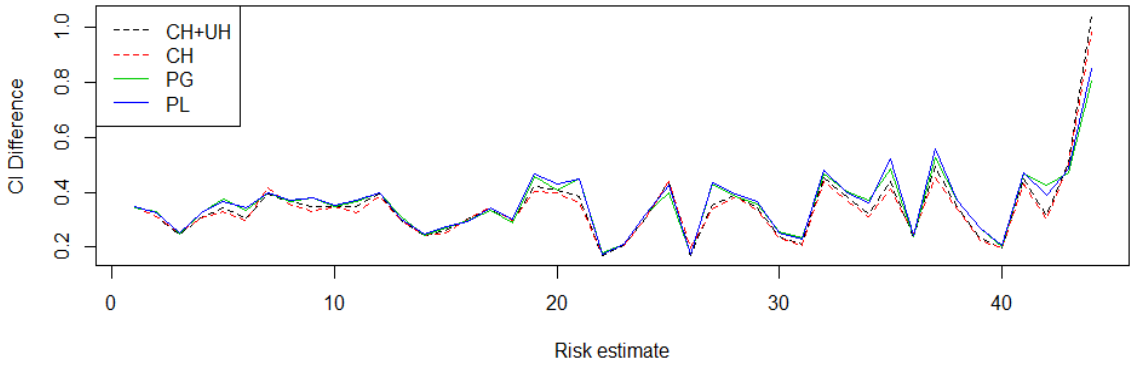


Figure 3.7: Credible interval difference for the posterior estimates of relative risk obtained from the fitted spatial models.

3.3 Spatio-temporal analysis

Since the main objective of this study was to assess the evolution of prostate cancer risk in space and time at the same time, hence more emphasis was given to the spatio-temporal disease mapping. As it has illustrated in section 2.3, we were fitted several models to the spatio-temporal data which can be classified as parametric linear trend type model, and smooth temporal evolution models that reflects inseparable space-time interaction. We considered an intrinsic CAR normal distribution for the spatially correlated random effect, u_i, while first-order random walk for the structured temporal effect, and normal distribution with mean 0 and variance σ_ν^2 and σ_ϕ^2, respectively for the unstructured spatial and temporal random effects. We have also used an intrinsic CAR prior distribution for the Type II interaction, while it comes from a normal distribution in the case of Type I interaction. Furthermore, we considered vague hyper-priors for all precision hyper-parameters $(\sigma_u^{-2}, \sigma_\nu^{-2}, \sigma_\gamma^{-2}, \sigma_\phi^{-2}, \sigma_\delta^{-2})$, and a flat prior for the overall risk level (α_0).

We also used similar techniques as spatial analysis to assess the convergence of McMC algorithm. For many of the spatio-temporal models described in Table 3.5, we initiated two chains with 100,000 total iterations and 50,000 burn-in for each chain, and 10 thinning value to accelerate the convergence time. However, it has been taken a total sample of 160,000 with burn-in of 80,000 iterations to get convergence results for the Knorr-Held models. Convergence was checked by visual examination of time series plots of samples for each chain, running mean plot and BGR plots. Following the BGR formal diagnostic procedures the estimated values of potential scale reduction factor $\hat{R}_I$ for all parameters in the fitted models are displayed in Table A.2 in the appendices. The estimated value of $\hat{R}_I$ revolves around 1, which indicates an acceptable degree of convergence of the Markov chain. The fitted models with corresponding relative goodness-of-fit statistics

such as DIC and MSPE are also shown in Table 3.5. The first model B_0 was fitted with the only overall temporal trend and spatially unstructured heterogeneity random effect and has large DIC score (DIC= 270.550) showing that it is far from the best model compared to all the others.

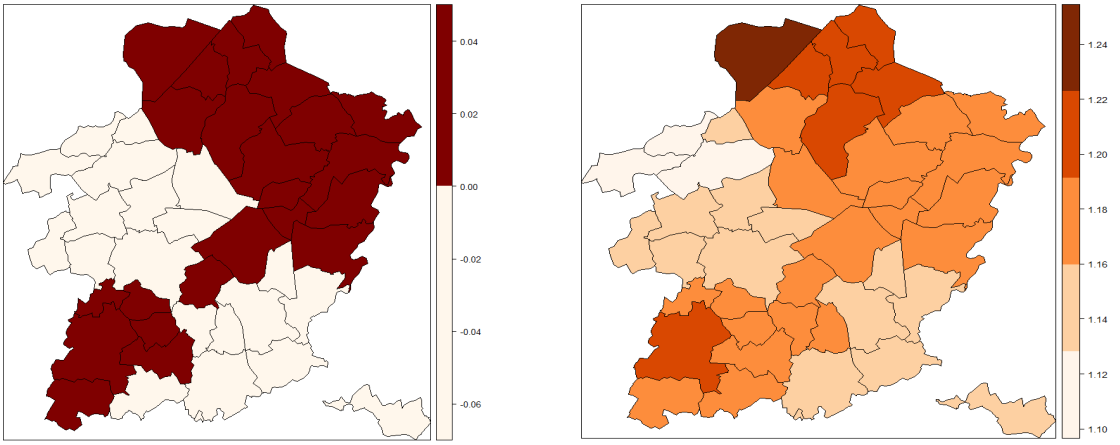
Table 3.5: Posterior estimates of goodness-of-fit statistics for the fitted spatio-temporal models.

Model type	Model	pD	DIC	MAPE	MSPE
	Linear trend models				
Linear	B_0: $\alpha_0 + \nu_i + \beta t_j$	36.300	2370.550	4.001	29.906
	B_1: $\alpha_0 + u_i + \nu_i + (\beta + \delta_i)t_j$	35.558	2355.200	3.988	29.760
	B_2: $\alpha_0 + u_i + (\beta + \delta_i)t_j$	35.876	2355.980	3.985	29.710
	B_3: $\alpha_0 + \nu_i + (\beta + \delta_i)t_j$	38.597	2370.300	3.992	29.621
	B_4: B_3, but $\delta_i \sim N(0, \sigma_\delta^2)$	40.162	2369.660	4.001	29.823
	Smooth temporal evolution models				
Main effect	K_0: $\alpha_0 + u_i + \nu_i + \gamma_j + \phi_j$	39.278	2344.380	3.938	28.643
	K_1: $\alpha_0 + u_i + \nu_i + \gamma_j + \phi_j + \delta_{ij}$	80.268	2336.200	3.783	26.130
Type I	K_2: $\alpha_0 + u_i + \nu_i + \phi_j + \delta_{ij}$	90.395	2337.050	3.820	26.800
Interaction	K_3: $\alpha_0 + \nu_i + \gamma_j + \phi_j + \delta_{ij}$	85.180	2351.360	3.833	26.830
	K_4: $\alpha_0 + u_i + \nu_i + \gamma_j + \phi_j + \delta_{ij}$	57.373	2337.760	3.879	27.630
Type II	K_5: $\alpha_0 + u_i + \nu_i + \gamma_j + \delta_{ij}$	50.695	2342.440	3.845	26.991
Interaction	K_6: $\alpha_0 + \nu_i + \gamma_j + \delta_{ij}$	65.827	2355.720	3.888	27.680
	AR(1): Model (2.16)	128.508	2343.520	3.725	25.396
Autoregressive	AR(2): Extension of AR(1)	120.269	2344.580	3.747	25.687
	AR(3): AR(1), but without ν_{ij}	78.845	2336.110	3.817	26.701

The parametric models B_1, B_2, B_3 and B_4 have been fitted to account both space and time with restricting to linear trend, which was proposed by Bernardinelli *et al* [16]. Specifically, model B_1 found in the first row and second column of Table 3.5 was fitted by accounting both the spatially structured, u_i and unstructured random main effects, ν_i along with the interaction, δ_i between space and linear temporal effect. The other three models B_2 and B_3 were fitted to make a more parsimonious model by removing the two random main effects one at a time. Similarly, model B_4 is model B_3 but the spatio-temporal interaction term δ_i comes from Gaussian distribution with mean 0 and variance σ_δ^2. We made the analysis with centering time to reduce the correlation between area specific intercept and trend. We have taken a flat prior for the overall risk level, α_0 and Gamma(0.5, 0.0005) for the precision hyper-parameters, $\sigma_u^{-2}, \sigma_\nu^{-2}$ and σ_δ^{-2}. Models B_1 and B_2 are virtually indistinguishable in terms of pD, MAPE and MSPE values. However, model B_2 was fitted without ν_i, which shows that adding the unstructured heterogeneity effect doesn't give an improvement on the model fitting.

On the other hand, models B_3 and B_4, respectively, with CAR and Gaussian prior to δ_i were fitted neglecting u_i, and the DIC value increases in both cases, hence the spatially structured random main effect has a tangible significant effect on the estimation of relative risk. In general, models B_1 and B_2 showed better fit than the others, however, model without spatially unstructured random main effect (B_2) is more parsimonious than the full model. The posterior estimated values with corresponding 95% credibility intervals for the variance components and trend are shown in Table A.2. For instance for model B_2, we found a significant time trend with an increased prostate cancer incidence over time between 1996 and 2005 ($\beta = 0.1506$;

CI=[0.0569, 0.2432]). As $\exp(\beta) = \exp(0.1506) = 1.1625$ is the rate ratio between two consecutive years, the risk was multiplied by approximately 1.1625 every year. This finding then shows that there is an increasing trend in the incidence rate from prostate cancer in Limburg.



(a) (b)

Figure 3.8: Posterior expected values for differential trend(δ_i) (a), and temporal trend ($\exp(\beta+\delta_i)$) obtained by linear trend model B_2 (b).

We also mapped the differential trend δ_i to investigate the geographical variation of the area-specific trend. An interesting point displayed in Figure 3.8a is that, a value of $\delta_i < 0$ (white) implies that the area-specific trend is less steep than the mean trend, whilst a value of $\delta_i > 0$ (highlighted areas) implies that the area-specific trend is steeper than the mean trend. However, there is not an extreme value of differential trend that would show the area-specific trend grossly deviates from the mean trend. Figure 3.8b also shows the temporal trend estimates in the counties of Limburg for the period 1996-2005. This is computed as $\exp(\beta+\delta_i)$ from the temporal model B_2 terms. In general, the incidence relative risk increases in the north, north-east and some part of south-west of the region.

The first four smooth temporal evolution models shown in Table 3.5 are Knorr-Held models and fitted to account spatially and temporally structured and unstructured main effects (K_0) at the same time, and inseparable space-time interactions (K_1, K_2, K_3). As we can see from the mathematical expression, the main effect model K_0 represents separable spatial and temporal variation at the same time in the estimation of relative risks. Whereas the other three models, K_1, K_2 and K_3 were formulated to represent inseparable Type I space-time variation. More specifically, model K_1 was fitted including all random main effects and the interaction between unstructured spatial and temporal effects ($\nu_i \times \phi_j = \delta_{ij}$), hence reflecting Type I inseparable variation proposed by Knorr-Held. In a similar manner, the remaining two models (K_2 and K_3) are simply the reduced form of K_1 and fitted by removing the structured main effects in order to get a more parsimonious model description. The hyper-prior distributions for all precision parameters were Gamma(0.5, 0.0005) in all Knorr-Held models [41]. Comparing the main effects model, K_0 (DIC= 2344.38) with the other Type I interaction models, K_1(DIC= 2336.20) and K_2(DIC= 2337.050), and Type II interaction models,

K_4(DIC= 2337.760) and K_5 (DIC= 2342.440), we can observe that the interaction models can ruled out the main effect model. This shows that there is a considerable evidence to include the interaction effect. In addition, the DIC value for K_1 is the smallest in Type I interaction models, with slight evidence to be ruled out the other nested models. In terms of model complexity, not surprisingly, the main effects model has the smallest effective number of parameters (pD=39.278).

In Table 3.5, models listed under Type II interaction were fitted to investigate the interaction between correlated temporal effect and spatially unstructured random effect ($\nu_i \times \gamma_j = \delta_{ij}$). Namely, model K_4 was fitted including all random main effects with an interaction effect. Similarly, model K_5 and model K_6 are model K_4 but by removing temporally heterogeneous and spatially structured main effects, respectively. As we can see in Table 3.5, there is a considerable difference in the posterior summary values of DIC, suggesting that model K_4 shows best fit; adding the spatial correlated random main effect and heterogeneous temporal main effects show substantial improvement on the fitted model.

Lastly, the three models AR(1), AR(2), and AR(3) found in the last three rows of Table 3.5 have been fitted to describe an autoregressive approach to the spatio-temporal disease mapping. Model AR(1) was fitted that already defined in (2.16), and model AR(3) is model AR(1) but it only has the structured term u_{ij} for every time period. Thus, we have three precision parameters in model AR(1), namely, σ_γ^{-2}, σ_u^{-2} and σ_ν^{-2}, while only σ_γ^{-2} and σ_u^{-2} in model AR(3). On the other hand, the mathematical formulation for model AR(2) is the same as AR(1), however, we extended that the precision parameters for the first period assumed to be different from those described in the following periods. That is, we have two precision parameters, $\sigma_{u_1}^{-2}$ and $\sigma_{\nu_1}^{-2}$ in the first period, and the other two different precision parameters, σ_u^{-2} and σ_ν^{-2} for the subsequent periods in addition to the common σ_γ^{-2} that captures an intercept random temporal effect. We have used Gamma$(0.5, 0.0005)$ as a hyper-prior distribution for all precision parameters with U$(-1, 1)$ prior for the correlation parameter ρ. The DIC value for AR(1), AR(2) and AR(3) are, respectively, 2343.52, 2344.58, and 2336.11, showing that model AR(3) has better performance than the other two. Model AR(1) fits also slightly better than model AR(2) in terms of DIC value but with a cost of complexity. In this approach, there is no any improvement that can be gained by including the unstructured spatio-temporal interaction term ν_{ij}.

Due to the large number of models fitted here we considered one model to make sensitivity analysis, to include the interpretations of estimates and to do predictions. Once we made several nested model comparisons, the next step is to select one model from each nested model shown in Table 3.5. The predictions of relative risk, and further interpretation about the posterior estimated values were illustrated based on the best fitted model. From the posterior summary values of DIC, 2355.200(B_2), 2336.20(K_1), 2337.760(K_4), and 2336.11(AR(3)) are the lowest DIC values from the linear, Type I interaction, Type II interaction, and autoregressive type models, respectively. Among these model, the difference in DIC is less marked between K_1, K_4 and AR(3). Based on this results, it appears clear that models K_1, K_4 and AR(3) represent the three most promising approaches for our data. In this case, we need to make a comparison considering model

complexity, which can be measured by the effective number of parameters found in each model. The model containing Type II spatio-temporal interaction effect (K_4) has around 57 effective parameters, whereas the model with Type I interaction effects (K_1) and AR(3), respectively, has about 23 and 21 additional effective parameters. Models K_1 and AR(3) has the best DIC score, but it requires substantial undertaking in terms of model complexity and therefore may not be an efficient choice in the statistical world, hence model K_4 was selected for further analyses. In other words, a model with Type II spatio-temporal interaction and spatial and temporal CH and UH random main effect, which is inseparable, appears to be the best choice overall. On the other hand, the least favorable models that have been fitted in our dataset in the spatio-temporal analysis are the linear parametric trend models. Therefore, it seems to be advisable to include the interaction between the spatially unstructured and temporal correlated random effects in the process of relative risk estimation. This can be thought of as unobserved spatial covariates for each pixel (i, j), that do not have any structure in space but in time. We now describe detailed analyses with Type II interaction, model K_4 as this model being less complex in terms of pD.

As a diagnostics analysis, we have also used the posterior distribution of the deviance residual for each space and time point, which can be estimated as

$$D_{ij} = \texttt{sign}(y_{ij} - E_{ij}\theta_{ij}) * \sqrt{2 * \left\{ y_{ij} \log\left(\frac{y_{ij}}{E_{ij}\theta_{ij}}\right) - (y_{ij} - E_{ij}\theta_{ij}) \right\}}$$

where θ_{ij} and E_{ij}, respectively, defined as the true relative risk and the expected incidence count for the ith municipality at the jth time. Figure 3.9 gives the deviance residual versus predicted diagnostics plot and the posterior distribution of the deviance for the final fitted model. The posterior mean of the deviance residual is plotted against the posterior mean of the predicted incidence counts, which doesn't show strong outliers in terms of deviance residuals.

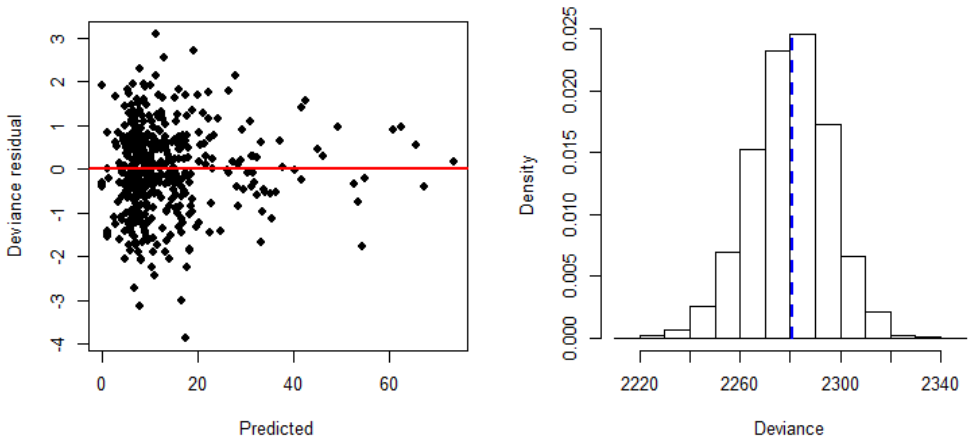


Figure 3.9: The deviance residual versus predicted diagnostics plot (left panel) and the posterior distribution of the deviance (right panel) for the final fitted model.

The histogram in the right panel of Figure 3.9 represents the empirical distribution of residual deviance based on 4,000 samples. Once more again, some of the convergence graphical plots for the final fitted model can

be found in the appendices, such as the running mean plot in Figure B.6, the autocorrelation plot in Figure B.7, the Geweke and the BGR diagnostic plots in Figure B.8 for each variance parameter and deviance. All these plots showed acceptable convergence in the Gibbs sampler.

3.3.1 Sensitivity analysis to priors

At the third stage of the fitted models in hierarchical Bayesian analysis, we required specifications of prior distributions for the second stage parameters (precision components) σ_l^{-2}s. As it has been described in section 3.3, the hyper-prior distributions of the variance components are generally set to be vaguely specified as a Gamma(a, b) distribution with scale and shape parameters a and b, respectively, which gives mean a/b, and variance a/b^2. In order to investigate the influence of hyper-prior distribution specifications on the results, a sensitivity analysis has been done for the spatial and temporal precision parameters, $\sigma_u^{-2}, \sigma_\nu^{-2}, \sigma_\gamma^{-2}, \sigma_\phi^{-2}$, and σ_δ^{-2}. Three hyper-prior distributions, Gamma$(0.01, 0.001)$, Gamma$(0.001, 0.001)$, and Gamma$(0.5, 0.0005)$ were used for all the precision parameters, to avoid problems with improper hyper-priors [16, 23, 41], and one uniform distribution, U(0,100) for the standard deviations, σ_ls [2]. The hyper-priors are arranged in increasing order of their dispersion, that is Gamma$(0.001, 0.001)$ are more dispersed than Gamma$(0.01, 0.001)$, and the latter have high dispersion than Gamma$(0.5, 0.0005)$.

Table 3.6: Posterior means for the variances, DIC and pD estimated from model K_4 under each prior.

	Hyper-priors			
Parameter	Gamma(0.001,0.001)	U(0,100)	Gamma(0.01,0.001)	Gamma(0.5,0.0005)
σ_u^2	0.0118	0.10679	0.01145	0.01136
σ_ν^2	3.83E-05	0.00262	4.22E-05	1.26E-05
σ_γ^2	6.65E-05	0.00323	5.37E-05	1.91E-05
σ_ϕ^2	5.66E-05	0.00467	6.07E-05	2.38E-05
σ_δ^2	6.64E-06	0.00172	7.79E-06	4.23E-06
DIC	2338.47	2339.76	2339.64	2337.76
pD	56.66	56.294	52.393	57.373

The posterior summary measures of variance components and model-fit statistics obtained with those hyper-prior distributions used as sensitivity analyses are shown in Table 3.6. The Gamma(0.5, 0.0005) prior to the precision components gives the smallest posterior estimates of DIC value (2337.76) relative to the others with a similar estimate for variances. The spatial distributions of the posterior mean of relative risks for the years 1996-2005 obtained from model K_4 under each prior distribution are displayed in Figures B.1 -B.4 in the appendices. All these maps showed that stable estimates for the relative risks can be obtained from each hyper-prior distributions, which indicates that results are robust against the specification of the prior distribution doesn't have an influence on the results in this specific study. In all cases, the spatially structured random main effect accounts larger posterior estimates of variance relative to the other effects. For instance, with Gamma(0.5, 0.0005) prior, the variability of the relative risk is attributed more to the spatially structured effects ($\sigma_u^2 = 0.01136$) than to the uncorrelated heterogeneity ($\sigma_\nu^2 = 0.0000191$). An interesting point to mention here is that the computational time of McMC under Uniform prior distribution was twice of the Gamma prior distributions.

3.3.2 Risk prediction

Once we have selected one best-fitted model which can give fairly stable estimates with different priors, the next step goes to the prediction of the risk of prostate cancer in Limburg. The predicted relative risk of prostate cancer using model K_4 was displayed in the form of a map in Figure B.4. There are clear and fairly stable spatial patterns in the first four years, and it highlights the geographical distribution of risks; the western, and north-eastern parts of Limburg appear at higher risk. As it can be seen in the map, the spatial pattern shows some changes for some counties such as Nieuwerkerken, Hamont Achel and Lummen in the years 1996-1999, but does not change much more within 2000-2001. The maps also show substantial increasing disease risk starting from 1999 with an exception for the year 2002. As indicated by a gradual darkening of the municipalities, prostate cancer incidence rates in Limburg are increasing differently over time with the exception of the year 2002. Note that the cut-off points in the maps were chosen for convenience, based on the distribution of the estimated posterior mean of relative risks, and that the lowest category includes values ranging from 0.26 to 0.99, and the highest category includes values ranging from 1.001 to 1.54 with the overall average of 0.99. Some lower and upper extreme SIR estimates have disappeared and much of the map has been smoothed.

Up to this point, we have been discussing mapping the posterior mean of the relative risk, but it does not make full use of the output of the analysis that provides, for each municipality, samples from the whole posterior distribution of the relative risk [2, 40]. As a further illustration, we need to deal with the maps for the posterior probability of estimated relative risk that exceed one for each municipality. We call it exceedance probability that can be used in the detection of municipalities that have an unusual risk or to assess elevated(reduced) areas. These probabilities were estimated from $\hat{P}(\hat{\theta}_{ij} > 1) = \sum_{g=1}^{G} I(\hat{\theta}_{ij}^{g} > 1)/G$, where $\hat{\theta}_{ij}^{g}$ is the sampled value of $\hat{\theta}_{ij}$ from a posterior sample of size G, and $I(a) = 1$ if a true and 0 otherwise. The threshold for the probability that used to detect the elevated areas was 0.975, and 0.025 for the reduced risks [2]. Which means that any municipality with exceedance probability >= 0.975, has been clustered as elevated area, while exceedance probability <= 0.025 detected as an area with reduced risk.

The posterior probability that the relative risk (RR) of each area higher than one is displayed in the form of a map in Figure 3.10. The municipalities of the region that are linked to elevated and reduced risks were highlighted by green and orange colors, respectively. It was shown that only the two municipalities were consistently detected as elevated areas (Lummen and Heusden Zolder) between 1997 and 1999, while a clear pattern with reduced risk found in the south-eastern parts of the region. For the remaining subsequent years, the number of elevated areas has been increased differently, and the number of detected areas with reduced risk slightly decreased with an exception for the year 2002. This could be explained due to the fact that, the model includes the spatio-temporal interaction effect to estimate the relative risk. In other words, this suggests that the spatial distribution of the risk of prostate cancer in Limburg shows a flexible pattern at each time point.

Similar results were obtained from the maps of a 95% credibility intervals shown in Figure B.5 in the appendices. Mapping the credibility intervals for the relative risk has an interesting importance to reflect significantly elevated (reduced) areas. In this aspect, a 95% credibility interval that includes 1 corresponds to insignificant relative risk, otherwise, it shows significant risk.

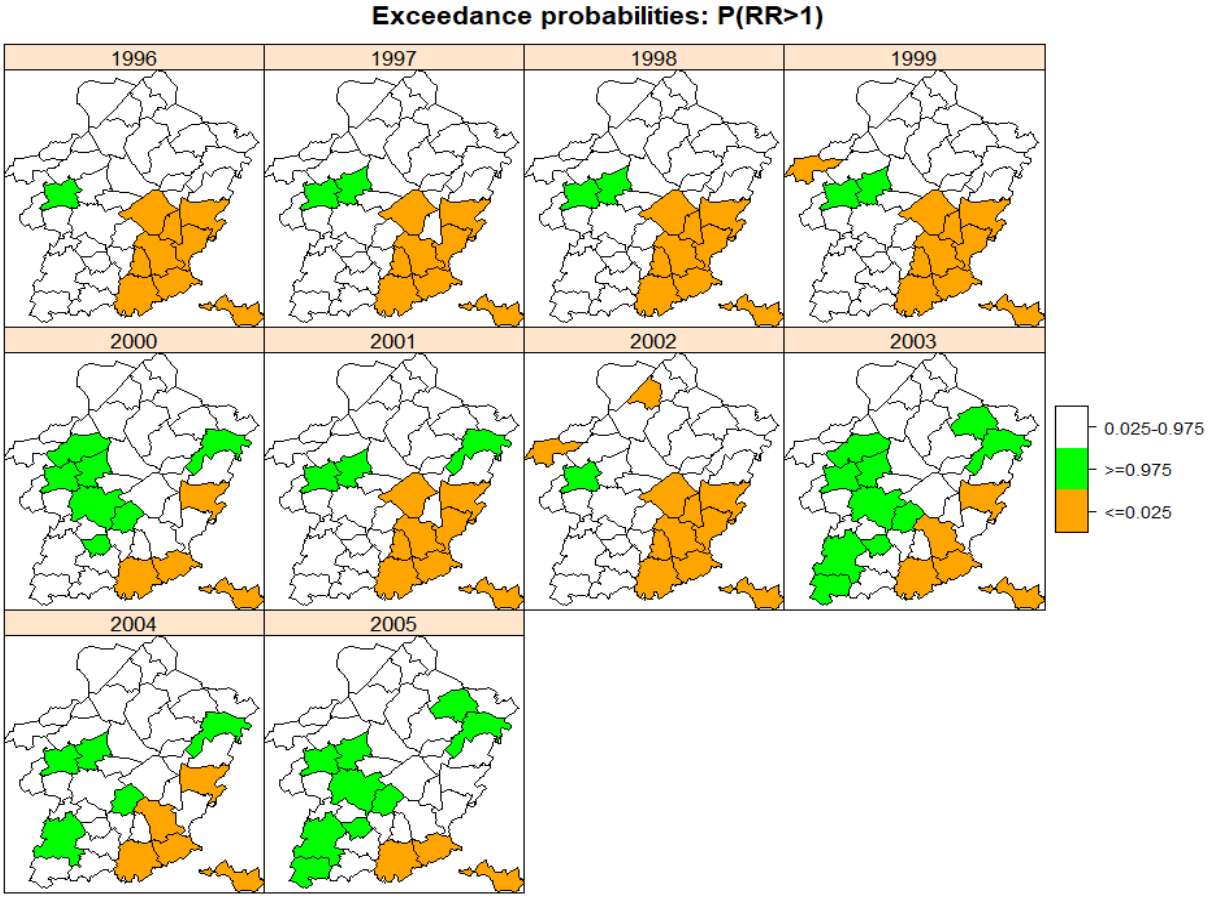


Figure 3.10: Exceedance probabilities for the relative risk obtained from the final fitted model (K_4).

To investigate how the Bayesian hierarchical models provide adequate smoothing of the background rates or SIR, we presented the evolution of the background SIR and posterior mean of relative risk for some randomly selected municipalities in Figure 3.11, and average evolution in Figure 3.12. By convenience, the legend included in the average evolution plot (Figure 3.12) can also use for the individual trajectories. In Figure 3.11, neighbouring areas are plotted in the same row, for instance, Hamont Achel, Neerpelt and Bocholt are neighbours. It has been seen that neighbouring municipalities have similar with a flexible shape to describe the variety of time trends that arise in the data, while those areas which are not neighbours evolve in a different manner. Moreover, the prediction plots showed neighbours municipalities which have relative risks that are consistently above or below even coincides with one. This reflects that risks are smoothed, accounting for sharing information in both spatial and temporal neighbours.

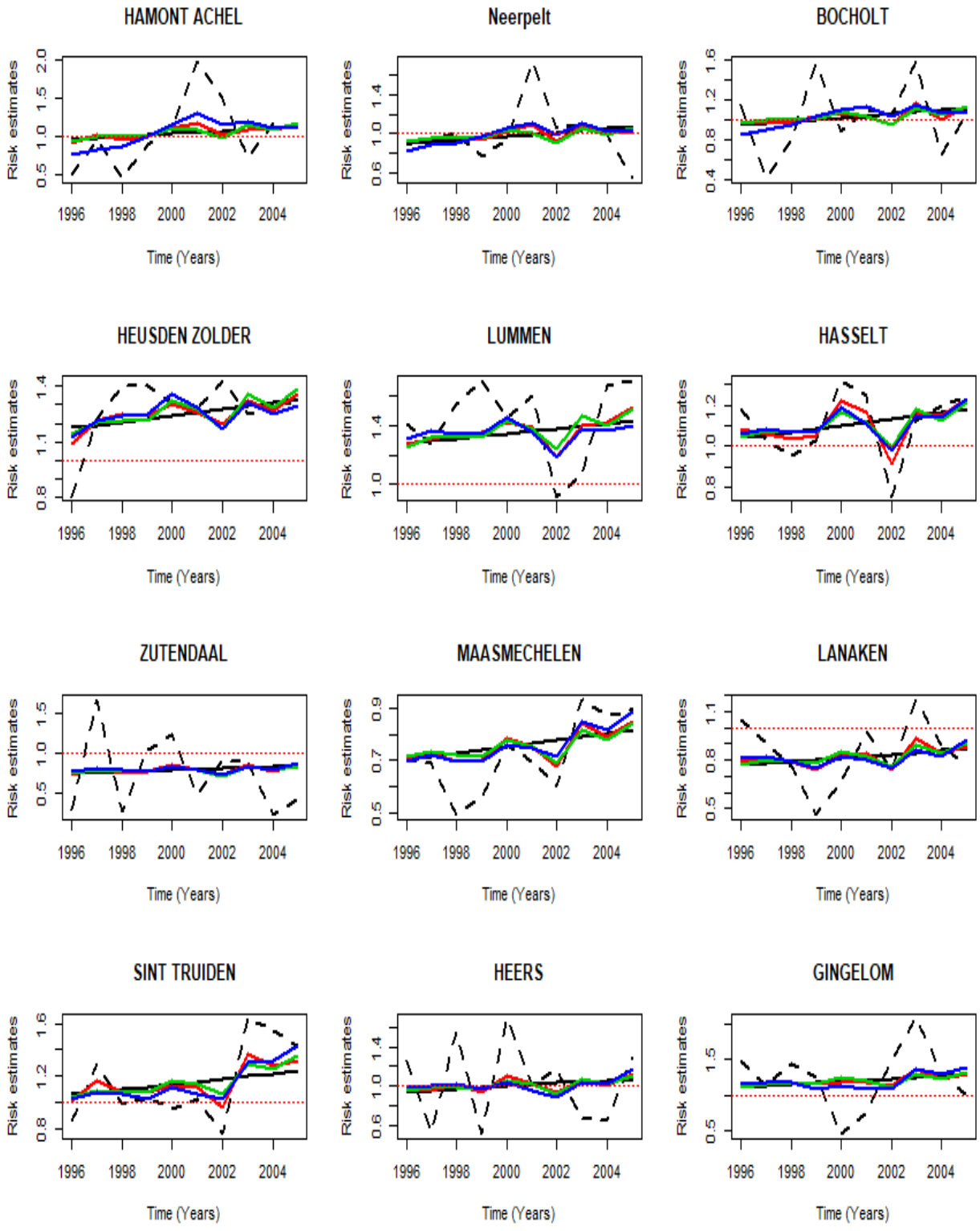


Figure 3.11: Standardized incidence and predicted risk estimates profiles for randomly selected municipalities. B2= model B_2, K1=model K_1, K2=model K_2, AR3=model AR(3).

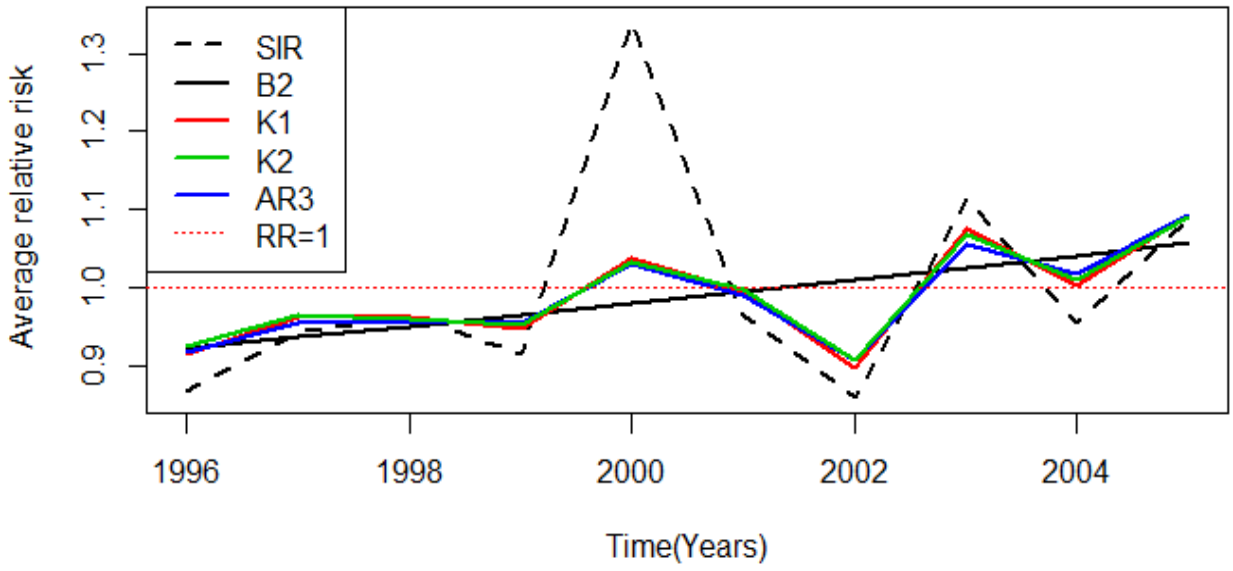


Figure 3.12: Average evolution of SIR and posterior mean of relative risk estimates.

The average evolution of raw SMRs showed unstable variability over time, which could be strongly influenced by the size of the population at risk, leads to a noisy of the true unobserved risks, whereas the Bayesian hierarchical models give posterior mean relative risk that evolving around 1. Also, these average evolution and individual trajectories showed additional implication, which illustrated that the smooth temporal evolution models (models K_1, K_2 and AR(3)) revealed better-smoothed estimates than the parametric linear trend model (B_2). Almost similar estimates for the relative risk has been obtained from the three smooth temporal evolution models. In general, the results suggested that the time trends for every municipality do not rely on a parametric shape, but flexible to describe the variety of time trends that arise in the data.

Note that in general, the lowest estimated relative risk was 0.26 (95% CI: 0.17–0.43) and the highest 1.54 (95% CI: 1.22–1.89) obtained from the final spatio-temporal model. Thus, we removed from the data random variability due to the small observed as well as expected counts. We are now dealing with a smoother map with less extremes in the relative risk estimates. This is an important illustration to understand that how the statistical model can able to smooth the risk estimates to eliminate random noise while at the same time avoiding over smoothing that might flatten any true variations in risk [40].

4 Discussion and conclusion

The main aim of this study was to assess the evolution of prostate cancer disease risk taking into account both space and time simultaneously and to investigate whether the spatial and temporal effects are separable, and finally to predict the relative risk for each area of the region. To address these main objectives, the data were analysed using several Bayesian hierarchical models, which accounts for the spatial and temporal effects as random effects, through prior distributions. In general, we examined four models for the spatial only data, and thirteen inseparable space-time interactions with two separable models, for the spatio-temporal dataset. In all cases for the precision parameters, we used Gamma priors which are conditionally conjugate for the given parameters, and hence computationally convenient to perform a Gibbs sampler draw of σ_l^{-2} in the posterior distribution [23]. We have also used Gibbs sampling techniques in McMC estimation, and inference for all models have been carried out using **R2WinBUGS** package for running **WinBUGS** from R [31].

Before conducting any analysis, we carried out internal rate standardization to remove the effect of age that makes comparable rates from different populations. The overall standardized rate ratio for the counties of Limburg in the years 1996-2005 was 1.001, while 0.997 for the spatial only data. We also assessed the presence of overdispersion, and found a significant indication for an extra variation in the incidence counts in turn heterogeneity in relative risks, which describes the presence of unadjusted geographically structured or unstructured unknown confounding variables. In addition, the exploratory test of spatial autocorrelation provides a substantial dependence among standardized incidence ratio taken at different locations.

First, for the spatial only data overdispersion or extra-variation was accommodated by some common Bayesian hierarchical models, such as Poisson-gamma and Poisson-lognormal models used to account spatially unstructured heterogeneity, a CAR model to include spatially structured variation, and finally CAR convolution model to employ both spatially structured and unstructured heterogeneity effects in the estimation of relative risk. Based on the relative measure of goodness-of-test, an intrinsic CAR and CAR convolution models provide the best fit for our dataset. This would be an indication that geographically structured variables dominate amongst the unobserved confounding variables that will affect the risk of prostate cancer in Limburg. This was in line with what we have observed strong spatial dependence in the spatial autocorrelation test under the exploratory data analysis. Although the spatial distribution for the posterior mean of relative risk showed almost similar patterns in all spatial models and a smoother map with fewer extremes in the relative risk estimates has been obtained, the aspatial models provide slight wider credible intervals for the relative risk in most parts of the region.

Bayesian hierarchical models to the spatio-temporal disease mapping were implemented in two basic approaches, namely parametric linear time trend and smooth temporal evolution models to account for the evolution of prostate cancer risk in both space and time simultaneously. The parametric linear trend models were conducted to describe which areas are showing increasing or decreasing trends in the prostate cancer risk over time. To extend this predefined linear restriction, a dynamic formulation for the spatial and tem-

poral random effects were assessed through inseparable space-time variation and autoregressive approach to spatio-temporal disease mapping. To ensure that contiguous areas are likely to be similar, we used spatially structured random effects through an intrinsic conditional autoregressive normal prior, while for contiguous periods temporally structured random effect with a first-order random walk prior was assumed, which allows flexible shapes in the evolution curves [24].

Based on our results, it appears that Knorr-Held models having Type I and Type II interactions, and an autoregressive approach investigated by Martínez with structured spatio-temporal interaction effect represent the three most promising approaches for our data. In terms of model complexity, model with Type II inseparable space-time interaction was found to be the best to describe the distribution of prostate cancer risk in Limburg. This result agreed with that was fitted by Knorr-Held to the respiratory cancer dataset [23], and the Type II interaction model was favored to Type I in his findings and also offered a lower deviance than Type III and IV interactions. Comparing to smoothed temporal evolution models, it was clearly shown that linear trend models, while parsimonious, are far from the best model. A variety of space-time models were fitted to very low birth weight count data for the counties of Georgia by Lawson [2], and he saw that the Knorr-Held models favored to the parametric linear trend model. On the other hand, a comparison of spatio-temporal disease mapping models conducted by [30] to Ischaemic heart disease count data showed that model AR(3) as the best fits in terms of DIC score, besides substantial undertaking in terms of computational time. In their finding, the Knorr-Held model with Type I interaction also showed better fit relative to the parametric linear trend model.

Regarding model performance in this specific study, Table 3.5 shows that those models which are fitted without including the spatial correlated random effects have the least performance. Indeed, the DIC score for Type I interaction model K_3 increases from 2336.20 to 2351.360, while that DIC value increases 0.85 units in model K_2 (model without temporally correlated random main effect). This has also been seen for Type II interaction model K_6 increases from 2337.760 to 2355.720, while the DIC value increases by 4.68 units in model K_5. This revealed that sharing information among municipalities has been shown to improve the model more than sharing information among periods. Overall, the posterior estimated value for the intercepts term along with its corresponding credible interval showed that the overall risk is approximately equal to one for all the fitted models. This could be due to the fact that we have used the marginalized total population as a standard population for each municipality to estimate the standardized expected counts. Note that in general, the maps in spatio-temporal models agreed with findings from spatial models, which adds to our confidence that this cluster is real, rather than simply a product of random variation.

The results of the Bayesian hierarchical models have typically been presented in the form of maps displaying the mean of the posterior distribution of the relative risk for each municipality. The smoothed temporal evolution models showed better fit than a parametric linear trend model to provide a substantial shrinkage of the raw relative risk estimates for each municipality. Based on the Bayesian hierarchical models,

a substantial random variability due to the small observed as well as expected counts has been removed from the data. Although the parametric linear trend models do not provide a good fit to our data, the global overall trend suggested an increasing risk of prostate cancer in Limburg. In this particular study, the results suggested that the time trends for every municipality do not rely on a parametric shape, but flexible to describe the variety of time trends that arise in the data. In conclusion, we have seen that sharing information among municipalities has been shown to improve the model more than sharing information among periods. This suggested in general that the spatial dependence is very important to describe the behavior of the risk in this specific data, indeed higher than the temporal one.

Some municipalities of the south-western parts of the region were detected as elevated risk in terms of the posterior probability that exceeds a threshold of one, whilst a clear pattern with reduced risk found in the south-eastern parts of the region. Therefore, we recommended to the epidemiologist to conduct further investigation on these specific areas. Finally, the area-specific random effects and the random time effects were assumed to be independent in this study. Thus, we recommended to extend this assumption for further studies by allowing a prior correlation between them.

References

[1] Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). John Wiley & Sons.

[2] Lawson, A. B. (2009). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

[3] Mausner, J. S. & S. Kramer (1985). *Epidemiology: An Introductory Text*. Philadelphia: W. B. Saunders.

[4] Inskip, H., Beral, V., Fraser, P., & Haskey, J. (1983). Methods for age-adjustment of rates. *Statistics in Medicine*, 2(4), 455-466.

[5] Breslow, N. E., & Day, N. E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases*, 28(5-6), 289-303.

[6] Neyens, T., Faes, C., & Molenberghs, G. (2012). A generalized Poisson-gamma model for spatially overdispersed data. *Spatial and spatio-temporal epidemiology*, 3(3), 185-194.

[7] Griffith, D. A. (1992). What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espace géographique*, 265-280.

[8] Fischer, M. M., & Wang, J. (2011). *Spatial data analysis: models, methods and techniques*. Springer Science & Business Media.

[9] Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17-23.

[10] Tiefelsdorf, M. (2006). *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I* (Vol. 87). Springer.

[11] Elliot, P., Wakefield, J. C., Best, N. G., & Briggs, D. J. (2000). *Spatial epidemiology: methods and applications*. Oxford University Press.

[12] Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press.

[13] Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4), 733-746.

[14] Beasag, J., York, J., & Moilie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of Mathematical Statistics*, 43, 1-59.

[15] Lawson, A. B. (2014). Hierarchical modeling in spatial epidemiology. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6), 405-417.

[16] Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, 14(21-22), 2433-2443.

[17] Waller, L. A., Carlin, B. P., Xia, H., & Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92(438), 607-617.

[18] Lesaffre, E., & Lawson, A. B. (2012). *Bayesian biostatistics*. John Wiley & Sons.

[19] Robert, C. P. (2004). *Monte carlo methods*. John Wiley & Sons, Ltd.

[20] Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

[21] Gelman, A. (1996). Inference and Monitoring Convergence in Markov Chain Monte Carlo in Practice. WR Gilks, S. Richardson, and DJ Spiegelhalter, eds.

[22] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.

[23] Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18), 2555-2567.

[24] López-Quilez, A., & Munoz, F. (2009). Review of spatio-temporal models for disease mapping. *Final Report for the EUROHEIS*, 2.

[25] Martínez-Beneito, M. A., López-Quilez, A., & Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in medicine*, 27(15), 2874-2889.

[26] Knorr-Held, L., & Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in medicine*, 17(18), 2045-2060.

[27] Potthoff, R. F., & Whittinghill, M. (1966). Testing for homogeneity: II. The Poisson distribution. *Biometrika*, 53(1/2), 183-190.

[28] Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). *Disease mapping with WinBUGS and MLwiN* (Vol. 11). John Wiley & Sons.

[29] Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717). New York: Springer.

[30] Anderson, C., & Ryan, L. M. (2017). A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International journal of environmental research and public health*, 14(2), 146.

[31] Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from RJ Stat. Soft. 12 (3): 1-16.

[32] Law, J., Quick, M., & Chan, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. *Journal of quantitative criminology*, 30(1), 57-78.

[33] Lee, D., & Shaddick, G. (2010). Spatial Modeling of Air Pollution in Studies of Its Short-Term Health Effects. *Biometrics*, 66(4), 1238-1246.

[34] Ver Hoef, J. M., Peterson, E., & Theobald, D. (2006). Spatial statistical models that use flow and stream distance. *Environmental and Ecological statistics*, 13(4), 449-464.

[35] World Health Organization (2012). International agency for research on cancer. https://www.wcrf.org/int/cancer-facts-figures/ data-specific-cancers/ prostate-cancer-statistics

[36] Lousbergh, D., Cloes, E., Op De Beeck, L., Rummens, J. L., Vanden Brande, J., Faes, C., ... & Hensen, K. (2007). Ten years of cancer in the Belgian province of Limburg: 1996-2005. LIKAR, Hasselt.

[37] Fitzmaurice, C., Allen, C., Barber, R. M., Barregard, L., Bhutta, Z. A., Brenner, H., ... & Fleming, T. (2017). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA oncology*, 3(4), 524-548.

[38] Melaku, Y. A., Appleton, S. L., Gill, T. K., Ogbo, F. A., Buckley, E., Shi, Z., ... & Fitzmaurice, C. (2018). Incidence, prevalence, mortality, disability-adjusted life years and risk factors of cancer in Australia and comparison with OECD countries, 1990–2015: findings from the Global Burden of Disease Study 2015. *Cancer epidemiology*, 52, 43-54.

[39] Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical methods in medical research*, 14(1), 35-59.

[40] Richardson, S., Thomson, A., Best, N., & Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, 112(9), 1016.

[41] Johnson, G. D. (2004). Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. *International Journal of Health Geographics, 3(1), 29.*

[42] American Cancer Society. Global Cancer Facts & Figures (2015). 3rd Edition. *Atlanta: American Cancer Society.*

Appendices

Appendix A: Tables

Table A.1: Summary statistics for the spatio-temporal data

	Observed		Expected		SIR	
Year	mean	variance	mean	variance	mean	variance
1996	10.432	107.274	11.405	85.314	0.868	0.150
1997	11.455	103.835	11.747	89.831	0.942	0.153
1998	11.682	90.455	12.118	94.952	0.965	0.180
1999	11.750	109.401	12.478	100.082	0.916	0.225
2000	13.727	152.389	12.840	105.039	1.339	3.615
2001	13.295	150.771	13.199	110.054	0.965	0.184
2002	11.682	78.780	13.564	115.078	0.859	0.122
2003	15.386	175.731	13.930	120.280	1.112	0.260
2004	14.409	188.759	14.312	126.464	0.956	0.214
2005	16.500	199.093	14.724	132.854	1.089	0.189

Table A.2: Posterior estimated values with credibility intervals for the variance components and the potential scale reduction factor $\hat{R}_I$ for the spatio-temporal models.

parameter	mean	sd	2.50%	97.50%	Rhat	mean	sd	2.50%	97.50%	$\hat{R}_I$
	B_1					B_2				
Intercept	0.9578	0.0176	0.9226	0.9927	1.0010	0.9573	0.0169	0.9242	0.9908	1.0011
β	0.1503	0.0468	0.0598	0.2405	1.0010	0.1506	0.0472	0.0569	0.2432	1.0009
σ^2_ν	0.00001	3.94E-05	2.98E-08	0.0001	1.0009	/	/	/	/	/
σ^2_u	0.0139	0.0096	0.0032	0.0396	1.0011	0.0145	0.0095	0.0039	0.0392	1.0010
σ^2_δ	0.0013	0.006	4.21E-08	0.0123	1.0031	0.0015	0.0073	4.33E-08	0.0139	1.0012
	B_3					B_4				
Intercept	0.9750	0.0358	0.9043	1.0460	1.0010	0.9750	0.0358	0.9043	1.046	1.0010
β	0.1511	0.0499	0.0529	0.2514	1.0010	0.1511	0.0499	0.0529	0.2514	1.0010
σ^2_ν	0.0026	0.0019	0.0006	0.0073	1.0009	0.0026	0.0019	0.0006	0.0073	1.0009
σ^2_δ	0.0003	0.001	4.88E-08	0.0027	1.0015	0.0003	0.0010	4.88E-08	0.0027	1.0015
	K_0					K_1				
Intercept	0.9951	0.0103	0.9801	1.0200	1.0011	0.9900	0.0010	0.9980	1.0020	1.0010
σ^2_u	0.0145	0.0105	0.0032	0.0421	1.0010	0.0141	0.0106	0.0029	0.0415	1.0009
σ^2_ν	3.63E-05	9.44E-05	2.11E-07	0.0003	1.0013	3.86E-05	9.59E-05	2.09E-07	0.0003	1.0009
σ^2_γ	5.95E-05	0.00033	1.86E-07	0.0004	1.0011	5.79E-05	0.0006	1.85E-07	0.0004	1.0010
σ^2_ϕ	5.53E-05	0.00019	4.11E-07	0.0003	1.0026	5.23E-05	0.0002	3.25E-07	0.0003	1.0009
σ^2_δ	/	/	/	/	/	0.0002	0.0002	2.43E-06	0.0007	1.0010
	K_2					K_3				
Intercept	0.9904	0.0106	0.9700	1.0110	1.0009	0.9850	0.0010	0.9980	1.002	1.0009
σ^2_u	0.0142	0.0109	0.0029	0.0426	1.0012	/	/	/	/	/
σ^2_ν	6.04E-05	0.00012	3.54E-07	0.0004	1.0013	0.0024	0.0018	0.0006	0.0071	1.0012
σ^2_γ	/	/	/	/	/	1.43E-05	0.0001	3.40E-08	0.0001	1.001
σ^2_ϕ	6.47E-05	0.0002	1.45E-06	0.0004	1.0014	2.18E-05	7.39E-05	6.45E-08	0.0001	1.0011
σ^2_δ	0.0002	0.0002	2.89E-06	0.0007	1.0034	0.0002	0.0002	1.91E-07	0.0007	1.0041
	K_4					K_5				
Intercept	0.9900	0.00101	0.9981	1.0020	1.0009	0.99	9.98E-04	0.998	1.002	1.0011
σ^2_u	0.0114	0.00971	0.0016	0.0364	1.0010	0.0115	0.0092	0.0020	0.0359	1.0009
σ^2_ν	4.22E-05	0.00013	2.06E-07	0.0003	1.0009	1.15E-05	5.08E-05	2.78E-08	9.37E-05	1.0011
σ^2_γ	5.37E-05	0.00026	2.03E-07	0.0004	1.0011	7.54E-05	2.78E-04	2.10E-07	5.06E-04	1.0010
σ^2_ϕ	6.07E-05	0.00030	4.03E-07	0.0004	1.0012	/	/	/	/	/
σ^2_δ	7.79E-06	1.10E-05	2.79E-07	3.77E-05	1.0011	3.72E-06	6.04E-06	5.32E-08	2.12E-05	1.0049
	K_6					AR(3)				
Intercept	0.9999	0.0003	0.9994	1.0010	1.0010	0.9512	0.0271	0.8989	1.0050	1.0009
σ^2_ν	0.0012	0.0014	4.65E-07	0.0047	1.0303	/	/	/	/	/
σ^2_u	/	/	/	/	/	0.0009	0.0011	4.08E-05	0.0041	1.0084
σ^2_γ	6.69E-05	0.0002	1.47E-07	0.0005	1.0012	0.0002	0.0004	4.52E-06	0.0009	1.0010
σ^2_δ	2.78E-05	3.84E-05	2.68E-07	0.0001	1.0225	/	/	/	/	/
	AR(1)					AR(2)				
Intercept	0.9561	0.0382	0.8819	1.0320	1.0009	0.9566	0.0388	0.8817975	1.0340	1.0012
σ^2_γ	0.0001	0.0004	3.37E-06	0.0008	1.0010	0.0001	0.0004	3.32E-06	0.0008	1.0014
$\sigma^2_{\nu_1}$	0.0002	0.0002	3.85E-05	0.0008	1.0061	0.0005	0.0010	7.52E-06	0.0027	1.0012
$\sigma^2_{u_1}$	0.0017	0.0016	0.0001	0.0058	1.0109	0.0090	0.0162	1.56E-05	0.0496	1.0026
$\sigma^2_{\nu_2}$	/	/	/	/	/	0.0002	0.0002	2.59E-05	0.0007	1.0015
$\sigma^2_{u_2}$	/	/	/	/	/	0.0012	0.0015	2.11E-05	0.0052	1.0061

Appendix B: Figures

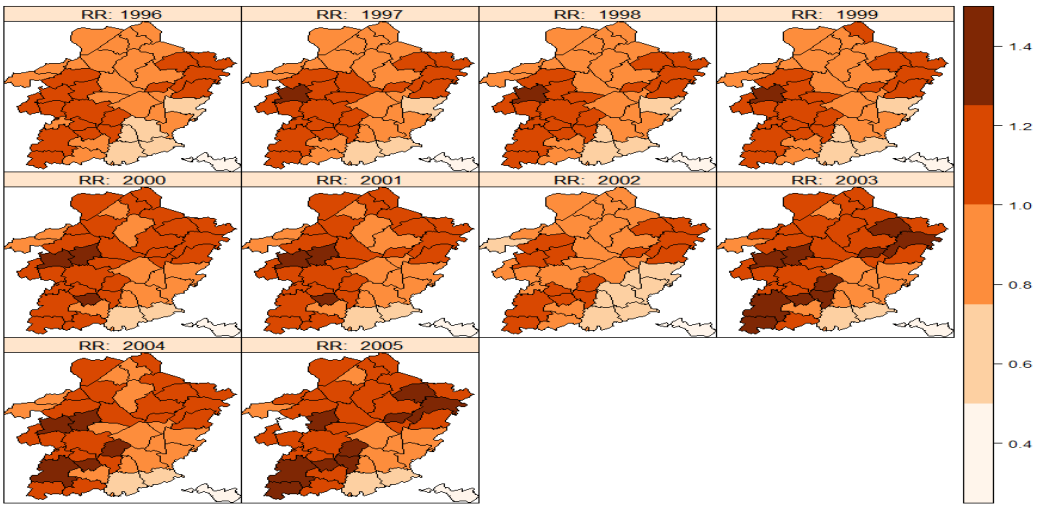


Figure B.1: Posterior estimated values of relative risks obtained from model K_4 under Gamma(0.001, 0.001) prior.

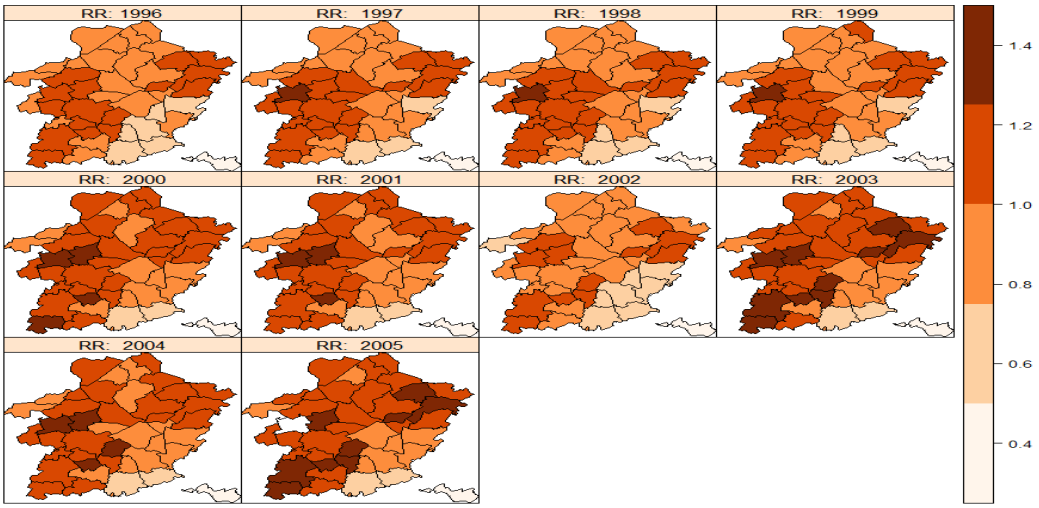


Figure B.2: Posterior estimated values of relative risks obtained from model K_4 under U(0, 100) prior.

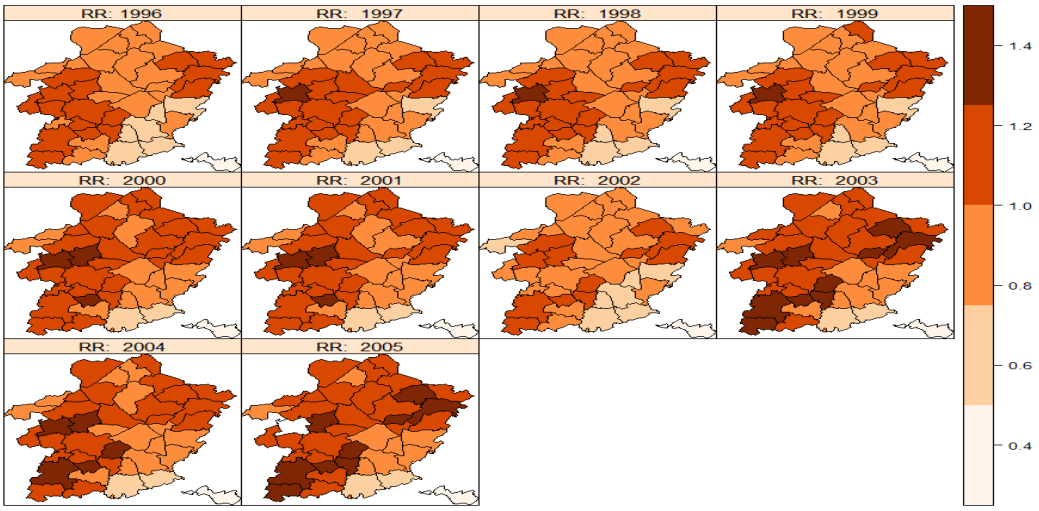


Figure B.3: Posterior estimated values of relative risks obtained from model K_4 under Gamma(0.01, 0.001) prior.

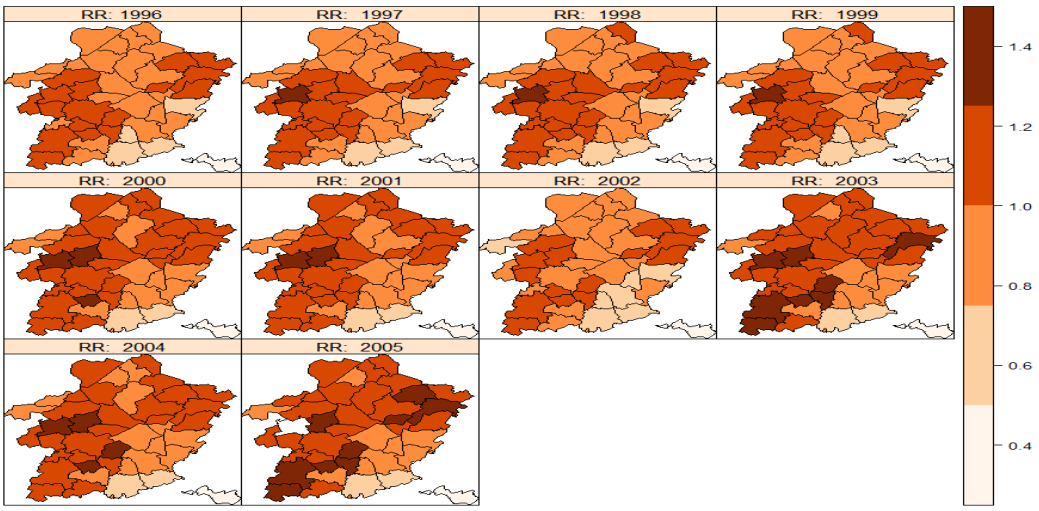


Figure B.4: Posterior estimated values of relative risks obtained from model K_4 under Gamma(0.5, 0.0005) prior.

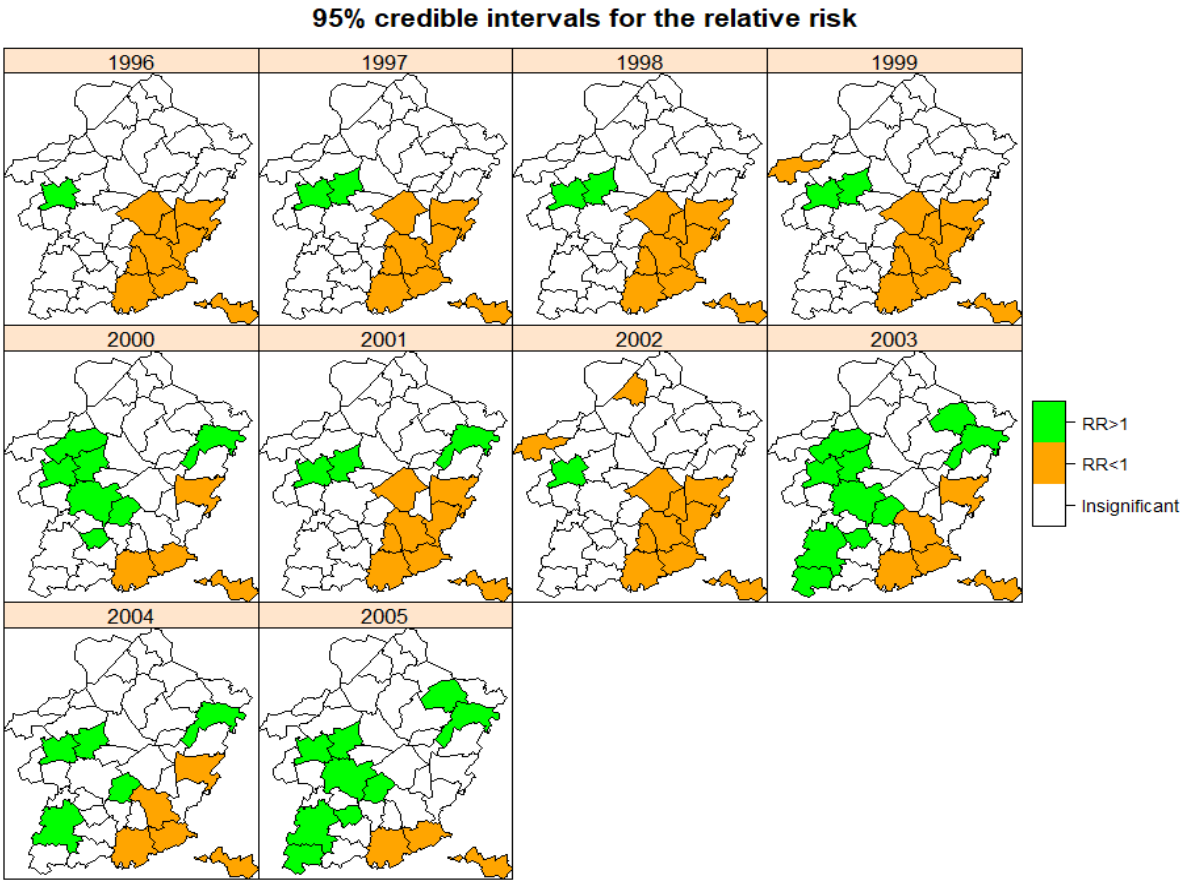


Figure B.5: Credible intervals for the relative risk of prostate cancer obtained from model K_4 for the years 1999-2005 in Limburg.

Convergence Diagnostics for a spatio-temporal model

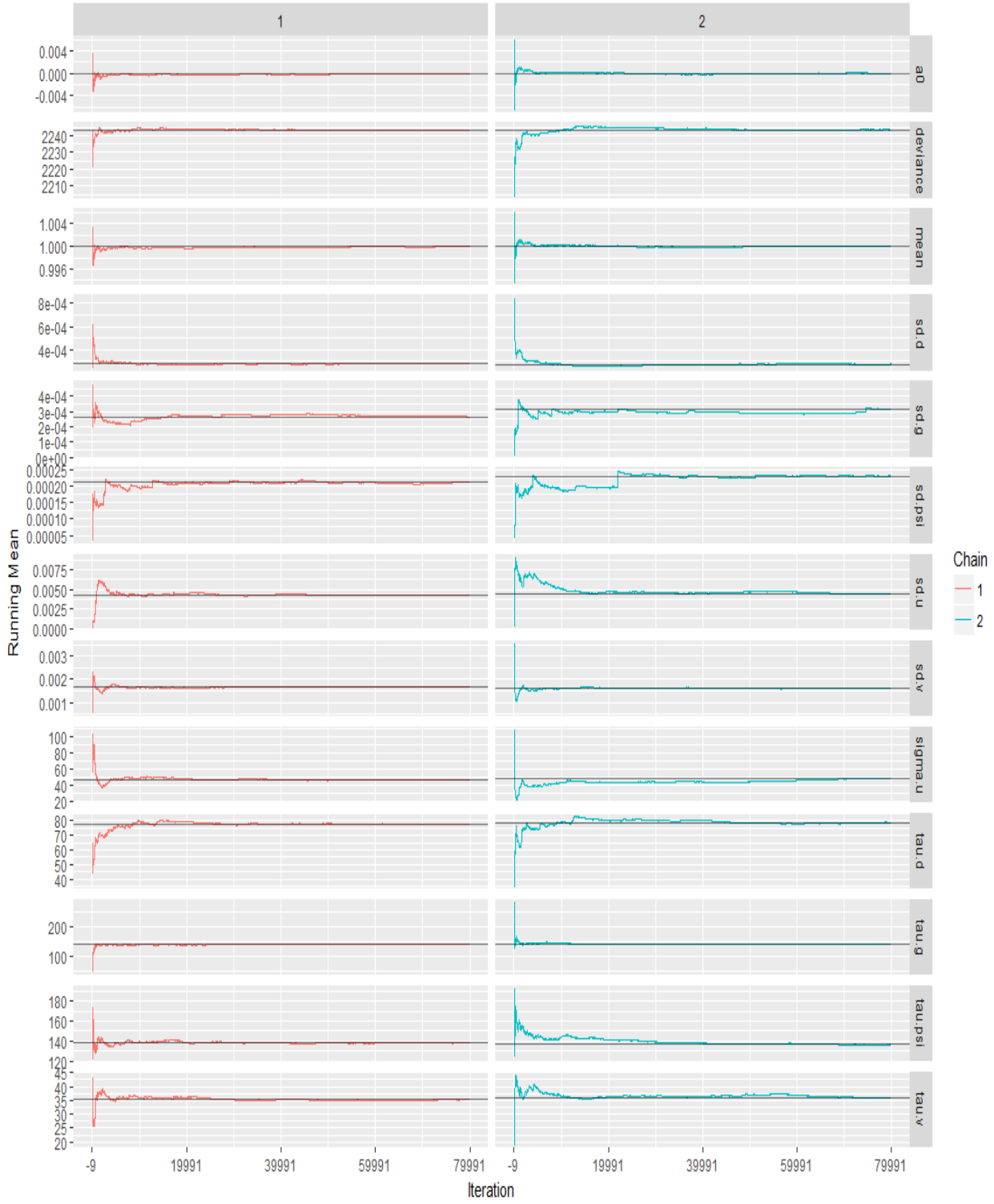


Figure B.6: Running mean plot for the final spatio-temporal fitted model

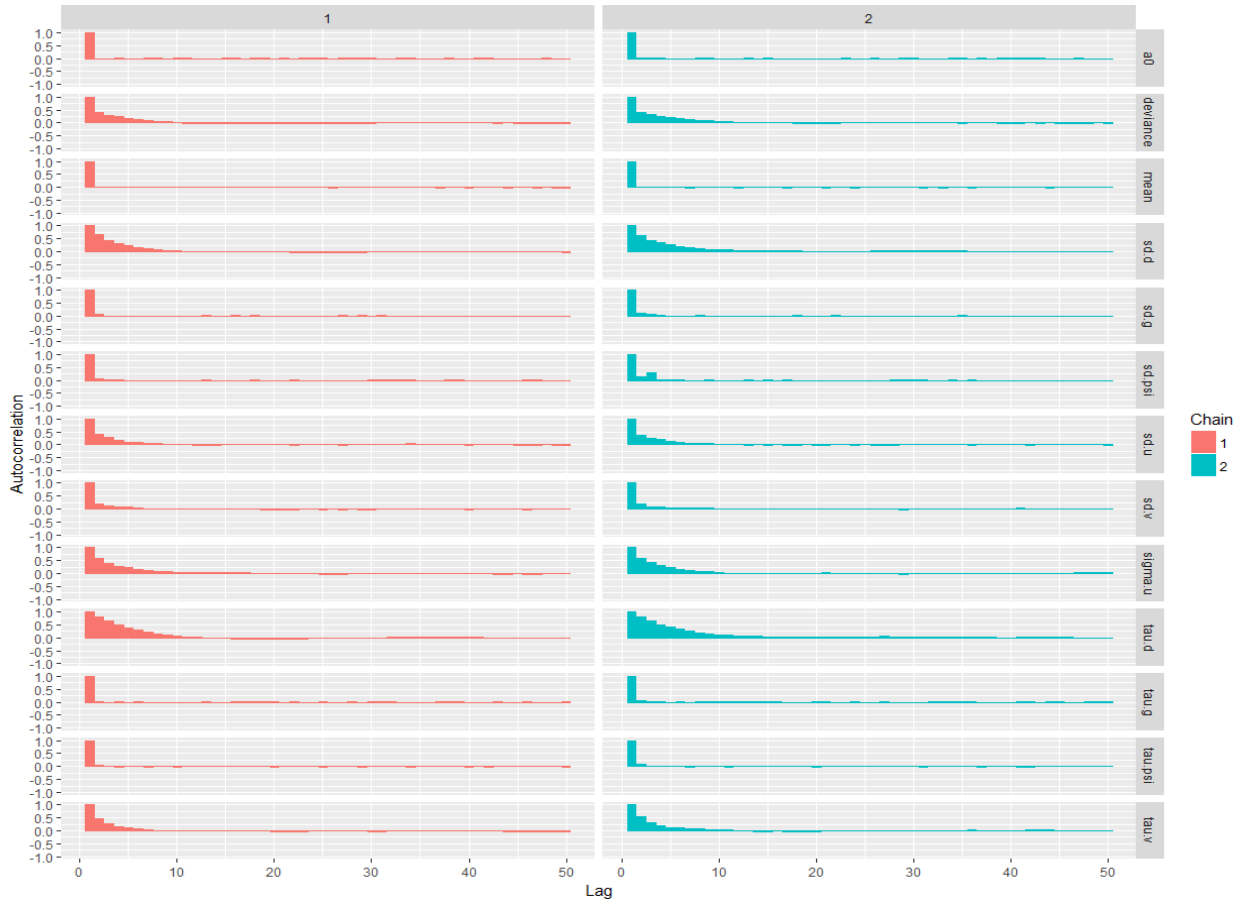


Figure B.7: Autocorrelation plot for the spatio-temporal fitted model k_4.

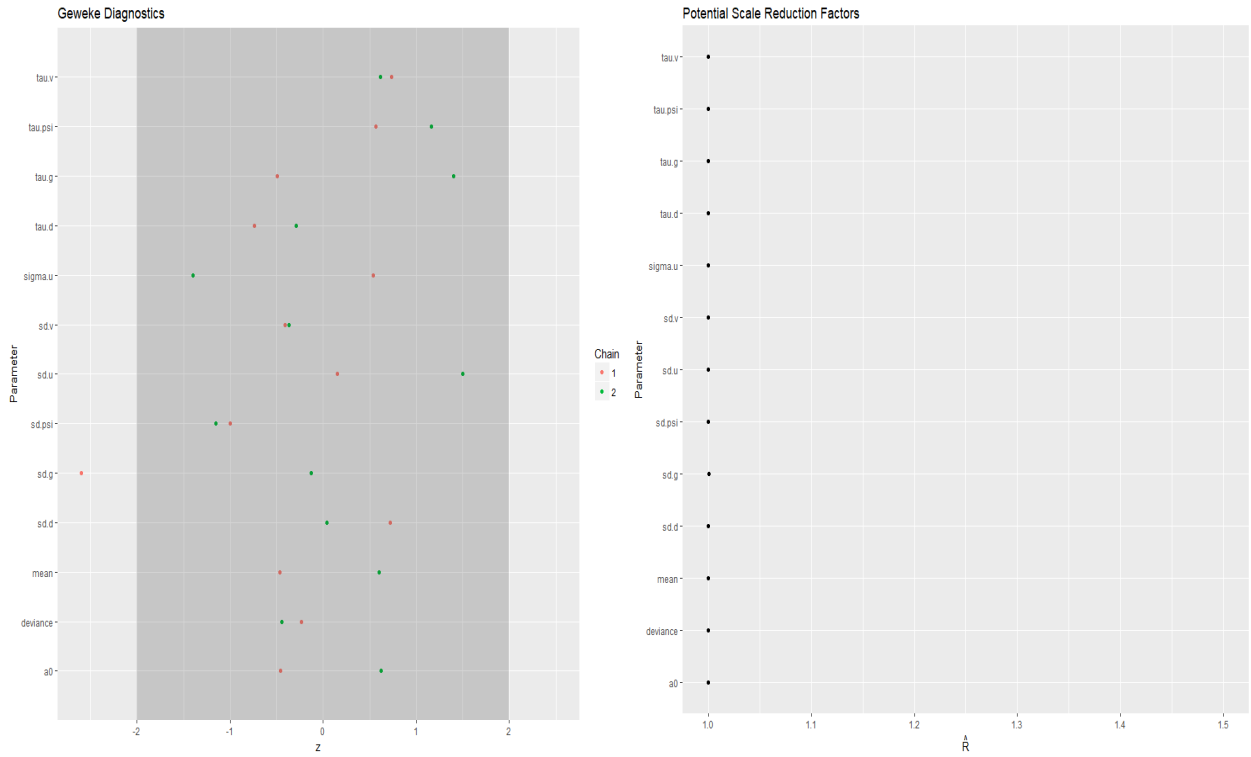


Figure B.8: Geweke diagnostics and potential reduction factors for model K_4.

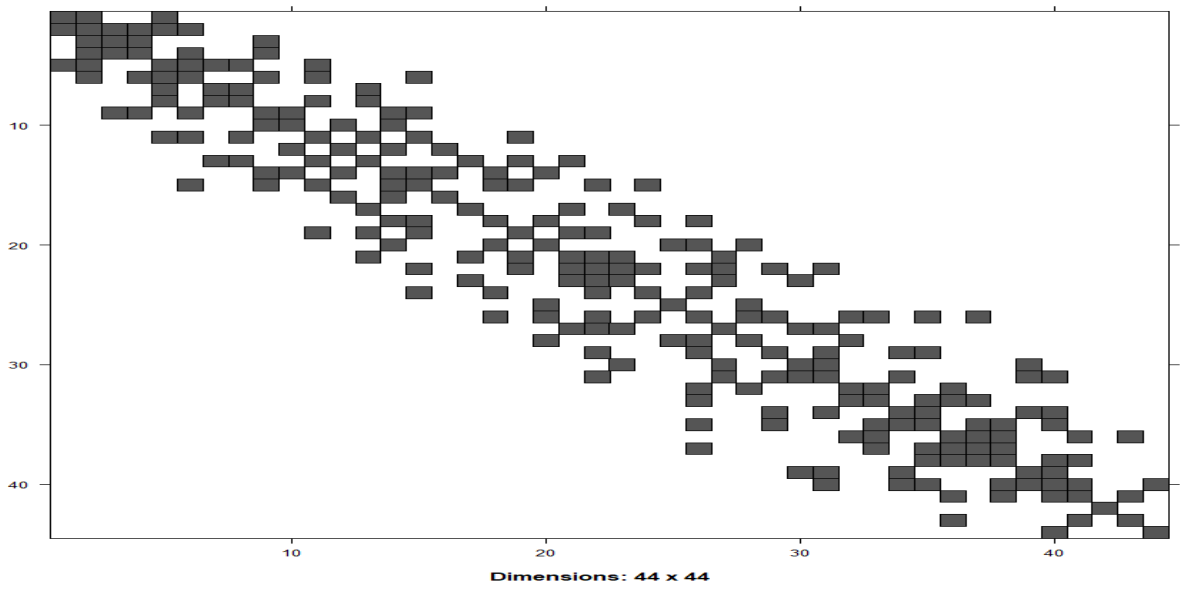


Figure B.9: Adjacency matrix for the Limburg: rows and columns identify areas; squares identify neighbors.

Diagnostics plots for the spatial models

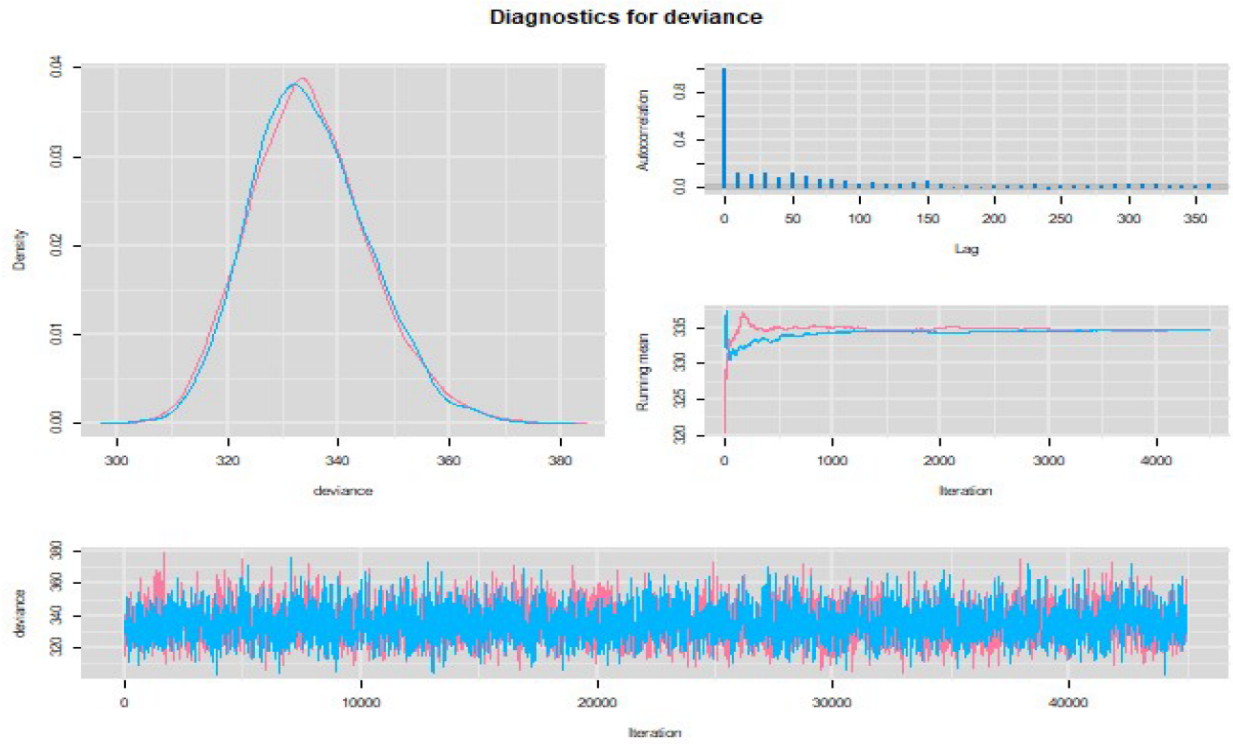


Figure B.10: Diagnostics plots for the Poisson-gamma model

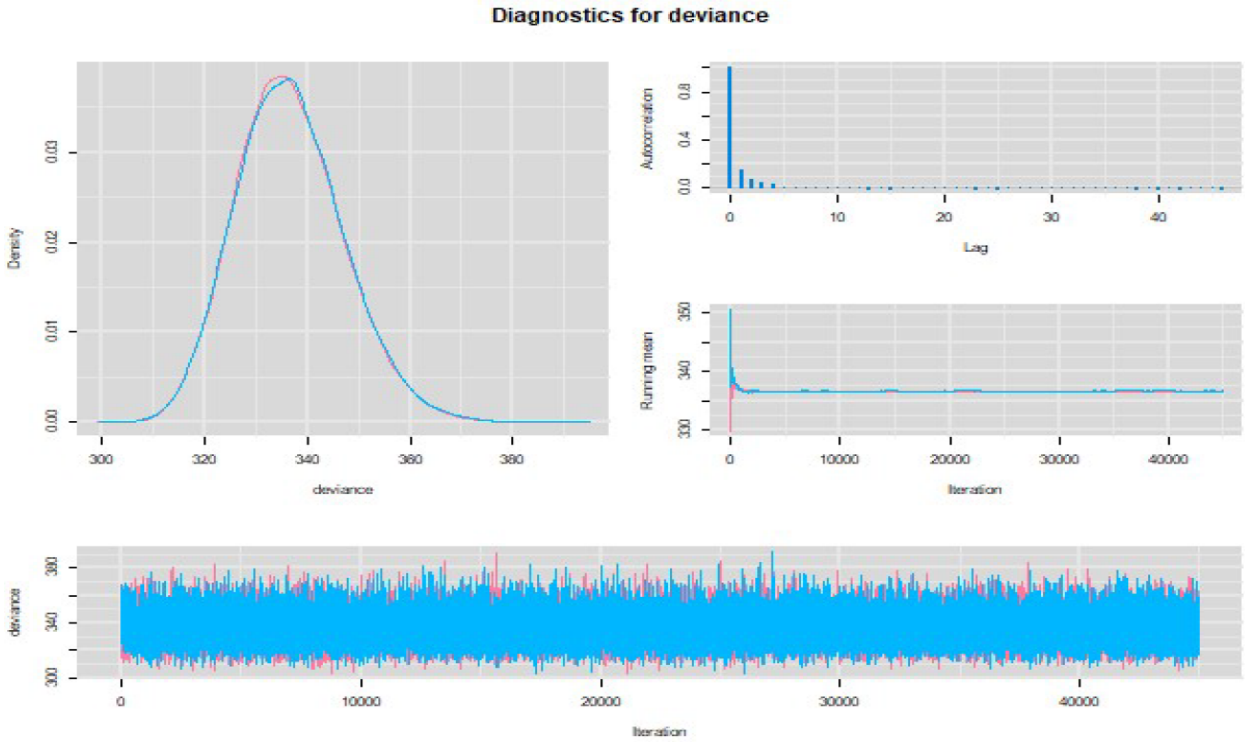


Figure B.11: Diagnostics plots for the Poisson-lognormal model

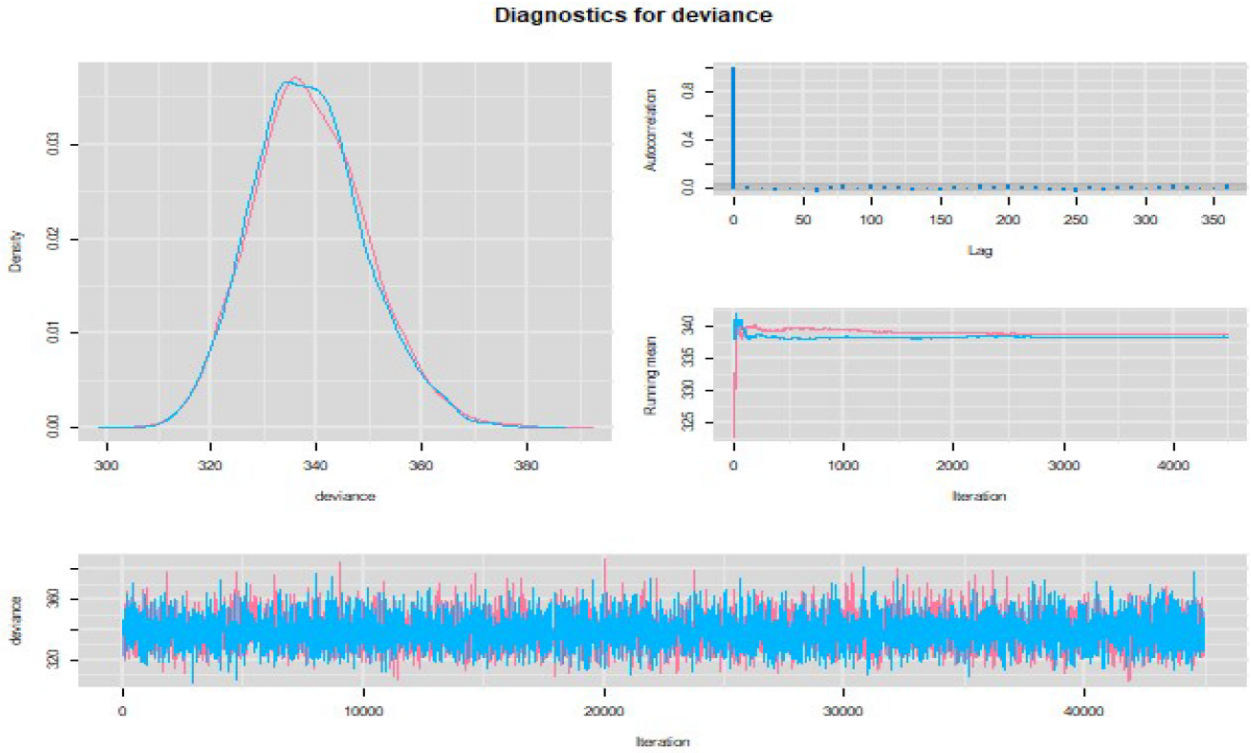


Figure B.12: Diagnostics plots for the CAR (CH) model

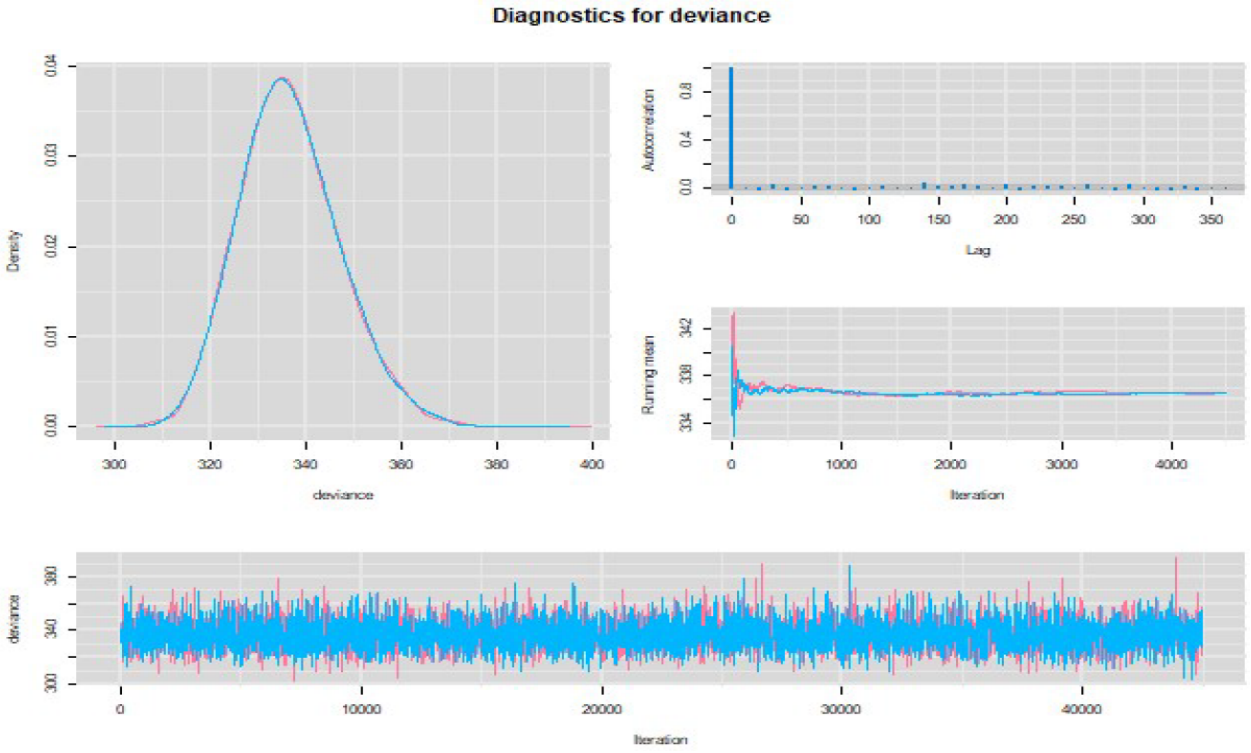


Figure B.13: Diagnostics plots for the CAR convolution model

Basic WinBUGS codes to run from R in the data analyses

```
#=============================================================================#
                          # spatial analysis
#=============================================================================#
setwd("E:\\Hasselt University\\2nd Year\\Spatio-Temporal\\Data")
s_sir<-read.csv("sir_spatial.csv", sep= ",", header = T)
O<-s_sir$Y## observed incidence counts
E<-s_sir$E## standardized expected counts
library(R2WinBUGS)
#=====================================================================
                         #Poisson-gamma model
#=====================================================================
sink("Pgamma_model.bug")
cat("model {
for (i in 1 : N)
{ # Poisson likelihood for observed counts
O[i] ~ dpois(mu[i])
mu[i] <- E[i]*theta[i]
# Relative Risks
theta[i] ~ dgamma(a, b)
## Goodness of fit
Opred[i]~dpois(mu[i])
pres[i]<-O[i] - Opred[i]
SPE[i]<-pow(pres[i],2)
APE[i]<-abs(pres[i])}
# Vague prior distributions
a ~ dexp(0.01)
b ~ dexp(0.01)
# Additional estimates
m <- a/b
var <- a/pow(b,2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[])
MAPE<-mean(APE[])
}", fill=TRUE)
sink()

model_PG = bugs(data= data_PG, inits= inits_PG,parameters_PG,
                model.file="Pgamma_model.bug", n.chains=2,
                n.iter=90000,n.burnin=45000, n.thin=10,
                working.directory = "E:\\Hasselt University\\2nd Year
                \\Spatio-Temporal\\Data",
                bugs.directory="C:\\Users\\user\\WinBUGS14",
                debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
                program="WinBUGS",codaPkg=F)
#=====================================================================
                        # Poisson-lognormal model
#=====================================================================
sink("Plognormal_model.bug")
cat("model {
for (i in 1:N){
# Poisson-likelihood for observed counts
O[i]~dpois(mu[i])
log(mu[i])<-log(E[i])+a0+v[i]
#heterogeneity random effects
v[i]~dnorm(0.0, tau.v)
```

```
# relative risks
theta[i]<-exp(a0+v[i])
## Goodness of fit
Opred[i]~dpois(mu[i])
pres[i]<-O[i] - Opred[i]
SPE[i]<-pow(pres[i],2)
APE[i]<-abs(pres[i])}
# Vague prior distribution for intercept
a0~dnorm(0.0, 1.0E-05)
mean<-exp(a0)
# Hyperprior distibutions on inverse variance parameter
tau.v~dgamma(0.01, 0.01)
sd.v<-1/pow(tau.v,2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[])
MAPE<-mean(APE[])
}", fill=TRUE)
sink()

model_plog = bugs(data= data_plog,inits_plog, parameters_plog,
             model.file="Plognormal_model.bug", n.chains=2,
             n.iter=90000,n.burnin=45000, n.thin=10,
              working.directory = "E:\\Hasselt University\\2nd Year
              \\Spatio-Temporal\\Data",
              bugs.directory="C:\\Users\\user\\WinBUGS14",
              debug=T, DIC=TRUE, digits=5,
              program="WinBUGS", bugs.seed = 1234,codaPkg=F)
#=============================================================================
                        # CAR model (CH)
#=============================================================================
# Convolution model: CH
sink("CH_model.bug")
cat("model
## Poisson likelihood{
for (i in 1:N){
O[i]~dpois(mu[i])
log(mu[i])<-log(E[i])+a0+u[i]
theta[i]<-exp(a0+ u[i])
## Goodness of fit
Opred[i]~dpois(mu[i])
pres[i]<-O[i]-Opred[i]
SPE[i]<-pow(pres[i],2)
APE[i]<-abs(pres[i])
#Moran's I
eres[i]<-(O[i]-mu[i])/sqrt(mu[i]) ## estimate of residuals
estar[i]<-sum(we[cum[i]+1:cum[i+1]])
de[i]<-eres[i]-mean(eres[])
d.estar[i]<-estar[i]-mean(estar[])
dt[i]<-de[i]*d.estar[i]
db[i]<-pow(d.estar[i], 2)}
# CAR prior distribution for random effects:
u[1:N]~car.normal(adj[], weights[], num[], tau.u)
for(k in 1:sumNumNeigh) {
weights[k]<-1
we[k]<-eres[adj[k]] }
## Other hyperprior distributions
a0~dflat()
```

```
mean<-exp(a0)
tau.u~dgamma(0.5,0.0005)
sd.u<-1/pow(tau.u,2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[])
MAPE<-mean(APE[])
rho<-sum(dt[])/sum(db[])
 }", fill=TRUE)
sink()

model_CH = bugs(data= data_CH,inits_CH,parameters_CH,
            model.file="CH_model.bug", n.chains=2,
            n.iter=90000,n.burnin=45000, n.thin=10,
            working.directory = "E:\\Hasselt University\\2nd Year
            \\Spatio-Temporal\\Data",
           bugs.directory="C:\\Users\\user\\WinBUGS14",
          debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
          program="WinBUGS",codaPkg=F)
#=============================================================================
                            # CAR convolution model (CH+ UH)
#=============================================================================
sink("Convlution_model.bug")
cat("model {
# Likelihood
for (i in 1:N) {
O[i]~dpois(mu[i])
log(mu[i])<-log(E[i]) + a0 + u[i]+ v[i]
theta[i]<-exp(a0+ u[i] + v[i])
v[i]~dnorm(0.0, tau.v)
## Goodness of fit
Opred[i]~dpois(mu[i])
pres[i]<-O[i] - Opred[i]
SPE[i]<-pow(pres[i],2)
APE[i]<-abs(pres[i])
# CAR prior distribution for random effects:
u[1:N]~car.normal(adj[], weights[], num[], tau.u)
for(k in 1:sumNumNeigh){
weights[k]<-1
we[k]<-eres[adj[k]]}
# Other priors
a0~dflat()
mean<-exp(a0)
tau.v~dgamma(0.5, 0.005)
tau.u~dgamma(0.5, 0.005)
sd.v<-1/pow(tau.v,2)
sd.u<-1/pow(tau.u,2)
ratio<-sd.u/sd.v
intraclass<-sd.u/(sd.u+sd.v)
## Overall MSPE and MAPE
MSPE<-mean(SPE[])
MAPE<-mean(APE[])
}",fill=TRUE)
sink()

model_conv = bugs(data= data_conv,inits_conv,
            parameters_conv,
             model.file="Convlution_model.bug", n.chains=2,
```

```
            n.iter=90000,n.burnin=45000, n.thin=10,
            working.directory = "E:\\Hasselt University\\2nd Year
            \\Spatio-Temporal\\Data",
            bugs.directory="C:\\Users\\user\\WinBUGS14",
           debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
            program="WinBUGS",codaPkg=F)
#=============================================================================
                            # Mapping for all spatial models
#=============================================================================
## load the libraries
library(maps)
library(maptools)
library(spdep)
library(rgdal)
library("sp")
##============================================================================
## posterior mean of relative risk estimates
##============================================================================
shap<-rgdal::readOGR ("Limburg2.shp")
## RR for PG
RR.PG<-model_PG$mean$theta
RR.PG <- as.data.frame(RR.PG)
colnames(RR.PG) <- c("RR.PG")
## RR for PL
RR.PL<-model_plog$mean$theta
RR.PL <- as.data.frame(RR.PL)
colnames(RR.PL) <- c("RR.PL")
## RR for CH
RR.CH<-model_CH$mean$theta
RR.CH <- as.data.frame(RR.CH)
colnames(RR.CH) <- c("RR.CH")
## RR for convoultion
RR.con<-model_conv$mean$theta
RR.con <- as.data.frame(RR.con)
colnames(RR.con) <- c("RR.Con")
shap@data <- data.frame(shap@data,RR.PG, RR.PL,RR.CH, RR.con)
library(RColorBrewer)
## mapping
spplot(shap, c("RR.PG","RR.PL", "RR.CH","RR.Con"),
names.attr = c("PG","PL", "CH", "CH+UH"),
at=c(0, 0.25,0.5, 0.75, 1.0, 1.25, 1.5), as.table=T,
col.regions=brewer.pal(9, "Oranges"),
par.settings = list(fontsize = list(text = 16)))


#============================================================================================#
                                    # Spatio-temporal analysis
#============================================================================================#
setwd("E:\\Hasselt University\\2nd Year\\Spatio-Temporal\\Data")
library(R2WinBUGS)
st_sir<-read.csv("SIR_spatio-temporal.csv", sep= ",", header = T)
O<-matrix(st_sir$observed,nrow = 44, ncol = 10, byrow = T)
E<-matrix(st_sir$expected,nrow = 44, ncol = 10, byrow = T)
SIR<-matrix(st_sir$SIR,nrow = 44, ncol = 10, byrow = T)
#============================================================================
# Bernardinell et al (1995): UH + CAR for structured main and interaction effects
#============================================================================
```

```
sink("Bernardinell_1.bug")
cat("model{
for (i in 1:N){
for (j in 1:T){
# Poisson likelihood for observed counts
O[i,j]~dpois(mu[i,j])
log(mu[i,j])<-log(E[i,j])+a0+u[i]+v[i]+beta*t[j]+delta[i]*t[j]
# Relative Risk in each area and period of time
theta[i,j]<-exp(a0+u[i]+v[i]+beta*t[j]+delta[i]*t[j])
## Goodness of fit
Opred[i,j]~dpois(mu[i,j])
diff[i,j]<-O[i,j]-Opred[i,j]
SPE[i,j]<-pow(diff[i,j],2)
APE[i,j]<-abs(diff[i,j])
# deviance
D[i,j]<-2*(mu[i,j]-O[i,j] +
O[i,j]*(log(O[i,j]/mu[i,j])))
Dres[i,j]<-pow(D[i,j],0.5)*((O[i,j]-mu[i,j])/abs(O[i,j]-mu[i,j]))}
## the posterior mean of the trend effect
Trend[i]<-exp(beta+delta[i])}
# CAR prior distribution for spatial correlated heterogeneity
u[1:N]~car.normal(adj[],weights[],num[],tau.u)
delta[1:N]~car.normal(adj[],weights[],num[],tau.delta)
# Prior distributions for the Uncorrelated Heterogeneity
for(i in 1:N){
v[i]~dnorm(0,tau.v)}
# Weights
for(k in 1:sumNumNeigh){
weights[k]<-1}
a0~dflat()
mean<-exp(a0)
# Hyperprior distributions on inverse variance parameter of random effects
beta~dnorm(0,1.0E-5)
tau.v~dgamma(0.5,0.0005)
sd.v<-1/pow(tau.v,2)
tau.u~dgamma(0.5,0.0005)
sd.u<-1/pow(tau.u,2)
tau.delta~dgamma(0.5,0.0005)
sd.delta<-1/pow(tau.delta,2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[,])
MAPE<-mean(APE[,])
}", fill=TRUE)
sink()

model_Bernar_1 = bugs(data= data_bernard,inits1_B1 ,
                 parameters.to.save = parameters1_B1,
                 model.file="Bernardinell_1.bug", n.chains=2,
                n.iter=100000,n.burnin=50000, n.thin=10,
                working.directory = "E:\\Hasselt University\\2nd Year
                \\Spatio-Temporal\\Data",
                bugs.directory="C:\\Users\\user\\WinBUGS14",
                debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
                program="WinBUGS",codaPkg=F)
#============================================================================
                        # Knorr-Held Type I interaction model (model K1)
#============================================================================
```

```
sink("Knorr_model_1.bug")
cat("model {
for (i in 1:N){
for (j in 1: T){
# Poisson likelihood for observed counts
O[i,j]~dpois(mu[i,j])
log(mu[i,j])<-log(E[i,j]) + a0 + u[i] + v[i] + g[j] + psi[j]+  d[i,j]
# Relative Risk in each area and period of time
theta[i,j]<-exp( a0 + u[i] +v[i]+ g[j]  + psi[j]+ d[i,j])
theta_exced[i,j]<-step(theta[i,j]-1)
## Goodness of fit
Opred[i,j]~dpois(mu[i,j])
diff[i,j]<-O[i,j]-Opred[i,j]
SPE[i,j]<-pow(diff[i,j],2)
APE[i,j]<-abs(diff[i,j])
d[i,j]~dnorm(0,tau.d)
# deviance
D[i,j]<-2*(mu[i,j]-O[i,j] +
O[i,j]*(log(O[i,j]/mu[i,j])))
Dres[i,j]<-pow(D[i,j],0.5)*((O[i,j]-mu[i,j])/abs(O[i,j]-mu[i,j]))
## PPO for outlier detection
PPO[i,j] <- exp(-mu[i,j]+O[i,j]*log(mu[i,j])-logfact(O[i,j]))
CPO[i,j]<- 1/PPO[i,j] }
## spatial  uncorrelated heterogeneity main effects
v[i]~dnorm(0,tau.v)
#deviance residual
resdev[i]<-sum(D[i,1:T])}
for (j in 1:T){
psi[j]~dnorm(0, tau.psi)}
# Hyperprior distributions on inverse variance parameter of random effects
# CAR prior distribution for spatial correlated heterogeneity
u[1:N]~car.normal(adj[], weights[], num[], tau.u)
for(k in 1:sumNumNeigh) {
weights[k] <- 1  }
# prior for temporal correlated effects
g[1]~dnorm(0,0.0001)
time1[1]<-time[1]-1995
for (j in 2: T){
g[j]~dnorm(g[j-1],tau.g)
time1[j]<-time[j]-1995}
# other priors
a0~dnorm(0,1.0E06)
mean<-exp(a0)
tau.u~dgamma(0.5,0.0005)
sd.u<-1/pow(tau.u,2)
tau.v~dgamma(0.5,0.0005)
sd.v<-1/pow(tau.v,2)
tau.psi~dgamma(0.5,0.0005)
sd.psi<-1/pow(tau.psi,2)
tau.g~dgamma(0.5,0.0005)
sd.g<-1/pow(tau.g,2)
tau.d~dgamma(0.5,0.0005)
sd.d<-1/pow(tau.d, 2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[,])
MAPE<-mean(APE[,])
#total residual deviance
```

```
totresdev<-sum(resdev[])
}", fill=TRUE)
sink()

model_knorr_1 = bugs(data_knorr_1,inits_knorr_1,parameters_knorr_1,
                 model.file="Knorr_model_1.bug", n.chains=2,
                 n.iter=120000,n.burnin=60000, n.thin=10,
                 working.directory = "E:\\Hasselt University\\2nd Year
                 \\Spatio-Temporal\\Data",
                 bugs.directory="C:\\Users\\user\\WinBUGS14",
                 debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
                 program="WinBUGS",codaPkg=F)
#============================================================================
                       # Knorr-Held Type II interaction model (model k4)
#============================================================================
sink("Knorr_model_2.bug")
cat("model {
for (i in 1:N){
for (j in 1: T){
# Poisson likelihood for observed counts
O[i,j]~dpois(mu[i,j])
log(mu[i,j])<-log(E[i,j]) + a0+ v[i]+ u[i]+ g[j]+ psi[j]+ d[i,j]
# Relative Risk in each area and period of time
theta[i,j]<-exp(a0+ v[i]+ u[i]+ g[j]+  psi[j] + d[i,j])
## RR for interaction only
theta.int[i,j]<-exp(d[i,j])
## relative risks exceeding 1
theta.ex[i,j]<-step(theta[i,j]-1)
## Goodness of fit
Opred[i,j]~dpois(mu[i,j])
diff[i,j]<-O[i,j]-Opred[i,j]
SPE[i,j]<-pow(diff[i,j],2)
APE[i,j]<-abs(diff[i,j])
# deviance
D[i,j]<-2*(mu[i,j]-O[i,j] +
O[i,j]*(log(O[i,j]/mu[i,j])))
Dres[i,j]<-pow(D[i,j],0.5)*((O[i,j]-mu[i,j])/abs(O[i,j]-mu[i,j]))
## PPO for outlier detection
PPO[i,j] <- exp(-mu[i,j]+O[i,j]*log(mu[i,j])-logfact(O[i,j]))
CPO[i,j]<- 1/PPO[i,j]}
## spatio-temporal effect for the first period
d[i,1]~dnorm(0,tau.d)
#Spatio-temporal effect for the subsequent periods
for (j in 2:T){
d[i,j]~dnorm(d[i,j-1],tau.d)}
## spatial heterogeniety effect
v[i]~dnorm(0,tau.v)
#deviance residual
resdev[i]<-sum(D[i,1:T])}
# CAR prior distribution for spatial correlated heterogeneity
u[1:N]~car.normal(adj[], weights[], num[], tau.u)
for(k in 1:sumNumNeigh) {
weights[k] <-1}
## prior for temporal random walk 1
g[1]~dnorm(0,0.0001)
time1[1]<-time[1]-1995
for (j in 2: T){
```

```
g[j]~dnorm(g[j-1],tau.g)
time1[j]<-time[j]-1995}
# temporal independent effect
for (j in 1: T){
psi[j]~dnorm(0, tau.psi)}
a0~dnorm(0, 1.0E06)
mean<-exp(a0)
# Hyperprior distributions on inverse variance parameter of random effects
tau.u~dgamma(0.5,0.0005)
sd.u<-1/pow(tau.u,2)
tau.v~dgamma(0.5,0.0005)
sd.v<-1/pow(tau.v,2)
tau.psi~dgamma(0.5,0.0005)
sd.psi<-1/pow(tau.psi,2)
tau.g~dgamma(0.5,0.0005)
sd.g<-1/pow(tau.g,2)
tau.d~dgamma(0.5,0.0005)
sd.d<-1/pow(tau.d, 2)
## Overall MSPE and MAPE
MSPE<-mean(SPE[,])
MAPE<-mean(APE[,])
#total residual deviance
totresdev<-sum(resdev[])
}", fill=TRUE)
sink()

model_knorr_2 = bugs(data_knorr_2,inits_knorr_2,parameters_knorr_2,
                 model.file="Knorr_model_2.bug", n.chains=2,
                 n.iter=120000,n.burnin=60000, n.thin=10,
                 working.directory = "E:\\Hasselt University\\2nd Year
                 \\Spatio-Temporal\\Data",
                bugs.directory="C:\\Users\\user\\WinBUGS14",
                debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
                program="WinBUGS",codaPkg=F)
#=============================================================================
                      # Martinez  Autoregressive model (model AR(3))
#=============================================================================
st_sir<-read.csv("SIR_spatio-temporal.csv", sep= ",", header = T)
Obs<-matrix(st_sir$observed,nrow = 10, ncol = 44, byrow = FALSE)
Exp<-matrix(st_sir$expected,nrow = 10, ncol = 44, byrow = FALSE)
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
sink("Mart_AR3.bug")
cat("model{
for(i in 1:nmuni){
for(j in 1:nperiods){
Obs[j,i]~dpois(mu[j,i])
#Modelling of the mean for every municipality and period
log(mu[j,i])<-log(Exp[j,i])+mediainter+inter[j]+theta.ST[j,i]
#SIR for every municipality and period
theta[j,i]<-exp(mediainter+inter[j]+theta.ST[j,i])
# Goodness of fit
Opred[j,i]~dpois(mu[j,i])
diff[j,i]<-Obs[j,i]-Opred[j,i]
SPE[j,i]<-pow(diff[j,i],2)
APE[j,i]<-abs(diff[j,i])
# deviance
D[j,i]<-2*(mu[j,i]-Obs[j,i] +
```

```
Obs[j,i]*(log(Obs[j,i]/mu[j,i])))
Dres[j,i]<-pow(D[j,i],0.5)*((Obs[j,i]-mu[j,i])/abs(Obs[j,i]-mu[j,i]))}}
#Spatio-temporal effect for the first period
theta.S[1,1:nmuni]~car.normal(adj[],weights[],num[],prec.spat)
for(i in 1:nmuni){theta.ST[1,i]<-pow(1-ro*ro,-0.5)*theta.S[1,i]}
#Spatio-temporal effect for the subsequent periods
for(j in 2:nperiods){
for(i in 1:nmuni){
theta.ST[j,i]<-ro*theta.ST[j-1,i]+theta.S[j,i]}
theta.S[j,1:nmuni]~car.normal(adj[],weights[],num[],prec.spat)}
for(k in 1:sumNumNeigh) {
weights[k] <- 1  }
#Prior distribution for the mean risk for every municipality and period
mediainter~dnorm(0,0.01)
mean<-exp(mediainter)
#Prior distribution for the global time trend
inter[1:nperiods]~car.normal(adjT[],weightsT[],numT[],prec.inter)
# Specify weight matrix and adjacency matrix corresponding to RW(1) prior
# (Note - this could be given in the data file instead)
for(j in 1:1) {
weightsT[j] <- 1; adjT[j] <- j+1; numT[j] <- 1}
for(j in 2:(nperiods-1)) {
weightsT[2+(j-2)*2] <- 1; adjT[2+(j-2)*2] <- j-1
weightsT[3+(j-2)*2] <- 1; adjT[3+(j-2)*2] <- j+1; numT[j] <- 2}
for(j in nperiods:nperiods) {
weightsT[(nperiods-2)*2 + 2] <- 1; adjT[(nperiods-2)*2 + 2] <- j-1;
numT[j] <- 1}
#Prior distribution for the precision parameters in the model
prec.inter~dgamma(0.5,0.005)
sd.inter<-1/pow(prec.inter,2)
prec.spat~dgamma(0.5,0.005)
sd.spat<-1/pow(prec.spat,2)
#Prior distribution for the temporal dependence parameter
ro~dunif(-1,1)
#Overall MSPE and MAPE
MSPE<-mean(SPE[,])
MAPE<-mean(APE[,])
}", fill=TRUE)
sink()

model_mart3 = bugs(data= data_mart,inits_mart3 ,
            parameters.to.save = parameters1_mart3,
            model.file="Mart_AR3.bug", n.chains=2,
            n.iter=100000,n.burnin=50000, n.thin=10,
            working.directory = "E:\\Hasselt University\\2nd Year
            \\Spatio-Temporal\\Data",
            bugs.directory="C:\\Users\\user\\WinBUGS14",
            debug=T, DIC=TRUE, digits=5, bugs.seed = 1234,
            program="WinBUGS",codaPkg=F)
#=============================================================================#
#=============================================================================#
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
The spatio-temporal modeling of prostate cancer in Limburg

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,

Ewnetu, Worku Biyadgie

Datum: **18/06/2018**