



UHASSELT



Maastricht University

KNOWLEDGE IN ACTION

Faculty of Sciences
School for Information Technology

Master of Statistics and Data Science

Master's thesis

Developing a tool to find the optimal neighborhood structure in a spatially discrete spatial analysis using Covid-19 mortality data

Maxime Busselen

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Quantitative Epidemiology

SUPERVISOR :

Prof. dr. Thomas NEYENS

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

2022
2023



Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics and Data Science

Master's thesis

Developing a tool to find the optimal neighborhood structure in a spatially discrete spatial analysis using Covid-19 mortality data

Maxime Busselen

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science,
specialization Quantitative Epidemiology

SUPERVISOR :

Prof. dr. Thomas NEYENS

HASSELT UNIVERSITY

Abstract

Master of Statistics and Data Science

Developing a tool to find the optimal neighborhood structure in a spatially discrete spatial analysis using Covid-19 mortality data

by Maxime BUSSELEN

Introduction: In spatial epidemiology, discrete spatial models require the construction of a neighborhood structure. These neighborhood structures have an order which represents the distance of spatial correlation. This study aims to develop a method that makes selecting the best-fitting neighborhood structure order easier.

Methods: Using the 2020 Covid-19 mortality data of Belgium, continuous spatial models are fitted to estimate the practical correlation range of spatial correlation present in the data. Next, using this estimation, a neighborhood structure order for discrete spatial modelling is selected that aligns with the estimated practical correlation range from the spatially continuous analysis. The analysis is then repeated using simulated data.

Results: The results are presented as a comparison of the estimated practical correlation range by the continuous model and the WAIC values of the discrete spatial models with varying neighborhood structure order. The results of the small-scale simulation study are presented analogously.

Discussion: The method performs well in some cases; in others, the model with the suggested neighborhood structure order is not the best-fitting model. The correlation range is often overestimated by the spatially continuous analysis. This results in the selection of a model with a high-order neighborhood structure, usually not the model with the lowest WAIC.

Keywords: Spatial epidemiology, spatial correlation, practical correlation range, neighborhood structure

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Dr. Thomas Neyens for providing guidance and support throughout the thesis project. This thesis would not have been possible without his advice, knowledge, and patience. Next, I would like to thank Prof. Dr. Christel Faes and Dr. Thomas Neyens for introducing me to the world of spatial statistics and spatial epidemiology. Their course has taught me spatial analysis methods that have proven essential for this project.

I would not have been where I am now without the help of my fellow classmates during this Master's. They have not only been good colleague students, but good friends as well. For that, I owe them my most sincere gratitude. Finally, I would like to mention my family and partner, who have always been supportive and encouraging during the completion of my thesis.

Contents

Abstract	iii
Acknowledgements	v
1 Introduction	1
1.1 Background	1
1.2 Objectives and research questions	4
1.3 Societal relevance and stakeholder awareness	4
1.4 Ethical considerations	5
2 Dataset	7
2.1 2020 mortality data	7
2.1.1 Data description	7
2.1.2 Data exploration	8
3 Methodology	11
3.1 Workflow	11
3.2 Spatially Continuous Model	12
3.2.1 Analysis procedure in R	14
3.3 Spatially Discrete Model	14
3.3.1 Analysis procedure in R	15
3.4 Comparing neighborhood structures	15
3.5 Simulated data	17
4 Results	19
4.1 Case study	19
4.1.1 Comparative results	19
4.1.2 Sensitivity Analysis	20
4.1.3 Results for Limburg	20
4.2 Simulation study	22
4.2.1 Exploration of simulated data	22
4.2.2 Results for the simulated datasets	23
5 Discussion and conclusions	25
5.1 Spatially continuous vs. spatially discrete model	25
5.2 Conclusion	26
5.3 Limitations	26
5.4 Further Research	27
Bibliography	27
Appendices	33

A Variograms	35
A.1 Data on 16 km^2 scale	35
A.2 Data on 4 km^2 scale	37
B BYM model results	39
B.1 Order $k = 1$	39
B.1.1 Data on 4 km^2 scale	39
B.1.2 Data on 16 km^2 scale	41
B.2 Order $k = 2, 3, 4, 5$	42
B.2.1 Data on 4 km^2 scale	42
B.2.2 Data on 16 km^2 scale	45
C R Code	49

Chapter 1

Introduction

1.1 Background

The field of data analysis and statistics has witnessed remarkable advancements in recent decades, leading to the development of numerous statistical models designed to handle the presence of correlation in data. Traditional methods, such as classic linear regression, which once dominated data analysis, often assumed independence among data points, a premise that is often far from reality (Chan, 2004). As a response to this limitation, researchers adapted and extended these models to create generalized linear models, which introduced enhanced capabilities to explicitly address correlation structures and implement weighted least squares (Dunn, 2018). Moreover, the emergence of mixed effects models, also known as hierarchical models, revolutionized the way dependencies within groups were handled by simultaneously estimating fixed effects at the population level and random effects to capture group-specific variations (Gelman, 2006). Temporal correlation, a distinctive form of correlation where data points observed in close temporal proximity tend to be more similar than those further apart, has also been a subject of intensive investigation. In this context, specialized statistical models like Generalized Estimating Equations (GEE) and mixed effects models have been developed in longitudinal data analyses to cater to the unique challenges of temporal correlation (Garcia, 2017).

Integrating geographical location and distances between data points in datasets gives rise to a dimension of data correlation known as spatial correlation (Chou, 1995). This form of correlation arises from spatial patterns in the data, as explained by Tobler's first law, emphasizing that "everything is related to everything else, but close things are more related than distant things" (Tobler, 1970). Over the past few decades, spatial statistics and geostatistics have developed remarkably, solidifying their status as crucial branches of modern statistics specializing in datasets, including spatial information. Geostatistics, in particular, focuses on unveiling average spatial trends and studying the correlations between data points across the spatial domain (Diggle, 2019).

The significance of spatial statistics is evident across diverse scientific fields, where spatial models have proven essential in unlocking new insights. In economics, spatial models have found utility in analyzing regional economic growth, using the spatial locations of economic organizations and the interdependence they share in exchanging information (Huggins, 2017). In agriculture and urban planning, spatial models are instrumental in monitoring and managing land use, optimizing resource allocation, and shaping sustainable development practices (Moura, 2020). Ecology benefits greatly from spatial statistics, as these models enable researchers to study animal movement patterns, examine the distribution of plants across vast areas, and understand ecosystem dynamics in

response to environmental changes (Lichstein, 2002; Thorup, 2014). Environmental science leverages spatial models to study the distribution of environmental variables, assess spatial trends and changes, and study the impacts of human activities on the delicate balance of our ecosystems (Li, 2014).

Arguably one of the most impactful applications of spatial statistics lies in the field of public health and epidemiology of infectious diseases (Lawson, 2018). In this domain, spatial modelling serves as an essential tool for analyzing the spread of diseases, identifying disease clusters, and studying the profound influence of environmental factors on health outcomes. By employing spatial modelling techniques, researchers gain a comprehensive understanding of disease dynamics, enabling more effective public health interventions and mitigating the risks of infectious diseases.

Modelling spatial data can be done in multiple ways, of which the most common are point process analysis, spatially discrete models, and spatially continuous geostatistical models. Spatial point process models are used to describe point patterns in a spatial field (Baddeley, 2007). The modelling process involves estimating parameters that describe the spatial intensity function and interaction effects. Some statistical functions that are often used for point processes are the K-function, G-function, or Ripley's K-function. In this thesis, however, the focus lies on the other two ways of spatial modelling. The main difference between spatially discrete and spatially continuous modelling is how they represent the spatial information of the data. In spatially continuous models, the spatial field is seen as a continuum at which a discrete number of locations are sampled. These models measure spatial trends across a continuous spatial field and can thus capture gradual variation between measurements. For example, one can estimate the disease prevalence continuously over the study region by sampling at a limited number of locations. By doing this, the prevalence of unsampled locations can be estimated thanks to the information gathered in surrounding locations. The observations are correlated to a certain degree, meaning closer observations are assumed to be more alike than observations further away. For instance, an area with a high prevalence of disease cases is likely to be surrounded by areas that have similar prevalence numbers. This correlation is included in the modelling by a spatial process. Examples include spatial autoregressive models and generalized geostatistical linear models, abbreviated as GLGM (Giorgi, 2018).

The other investigated type of spatial models consists of spatially discrete models. Examples include spatial lattice models and conditional autoregressive models (CAR) (Besag, 1974). These models aggregate data in spatial units, separated by pre-specified boundaries. These units or sub-regions can be regular over a raster with the same area size, or irregular in shape or size. An example of irregular sub-regions is municipalities or provinces. When modelling data discretely, the spatial relationship between the units must be explicitly specified in a neighborhood structure. These models treat the spatial domain as a network of interconnected units and acknowledge the presence of neighbors in the form of adjacency or contiguity. Specifying the neighborhood structure is a critical aspect of spatially discrete models, as it directly impacts the estimated spatial correlation.

Despite the utility of spatially discrete models, the choice of neighborhood structure is often subjective and can significantly influence model outcomes. The selection of neighbors can vary based on different criteria, resulting in distinct correlation patterns and potentially leading to contrasting conclusions. Some methods of defining neighboring

spatial units are the following (Moraga, 2023). Contiguity-based neighbors are spatial units that share a common border, known as rook's contiguity, or a common border and/or a common vertex, known as queen's contiguity. Distance-based neighbors are defined based on spatial proximity and a prespecified distance threshold. Neighbors can also be specified using weights. A spatial weights matrix commonly represents the neighborhood structure using some or all spatial units where each entry represents the weight or connectivity between two spatial units. In this analysis, neighbors were specified using a contiguity-based structure that allows for lags using an order parameter. The model considers spatial relationships beyond the immediate neighboring units to allow indirectly linked spatial units to be neighbors with a lag that aligns with the order number. First-order neighbors are spatial units that are directly linked. Second-order neighbors are indirectly connected by being separated by one spatial unit. Neighborhood structures can be defined using any higher-order structure, making it as extensive as necessary. Higher-order structure models use more information on the farther regions than models using only information on the first or second-order neighboring spatial units. If the correlation of the spatial phenomenon is thought to reach further than the directly surrounding regions, higher-order neighborhood structures are more appropriate.

The neighborhood structure order can impact the validation of discrete spatial models. Choosing an appropriate structure that aligns with empirical observations or domain knowledge is crucial for ensuring the model accurately represents real-world phenomena. A mismatch between the neighborhood structure assumed in the model and the actual spatial relationships in the system being studied may lead to inaccurate results. Especially in an epidemiological setting, when modelling infectious diseases, the neighborhood structure should reflect the spatial correlation of the disease cases. The optimal order depends on the spatial correlation of the data and, thus, the practical correlation range of the disease.

The spatial dependence term in geostatistics has been extensively discussed in the literature; however, there remains a lack of clarity regarding the definition of spatial association between data points within a neighborhood structure. This ambiguity poses a potential challenge in spatial modelling, particularly in infectious disease modelling. The Covid-19 pandemic has highlighted the variable nature of spatial correlation due to its initial strong outbreaks (Franch-Pardo, 2020). In the early stages of a Covid-19 outbreak, we often observe small clusters of infections that can be effectively captured by employing single-order neighborhood structures. However, as the infections spread geographically, higher-order neighborhood structures may become more suitable for accurately representing the disease's broader coverage. This is the rationale for utilizing Covid-19 mortality data to test and validate the selection of an appropriate neighborhood structure in this analysis.

It is important to note that determining the optimal order of dependence structure in a disease scenario where the number of infections has reached saturation is not always straightforward. Additionally, the spatial correlation effect may be less pronounced due to the nature of the data being focused on mortality cases rather than infections. By acknowledging these complexities and investigating the relationship between the neighborhood structure and the Covid-19 mortality data, this thesis aims to enhance our understanding of spatial dependencies and their impact on infectious disease modelling.

Such insights can contribute to more accurate and reliable predictions, ultimately effectively assisting public health efforts in combating future outbreaks.

This thesis addresses the challenge of defining the most appropriate neighborhood structure using mortality data from Flanders, Belgium. Specifically, this study focuses on the data collected in 2020, a year significantly impacted by the Covid-19 pandemic, which greatly influenced mortality rates. While this research utilizes Covid-19 mortality data, it is essential to note that the primary objective is methodological rather than an epidemiological assessment of Covid-19 mortality itself. The main goal is to develop a tool to effectively determine the optimal neighborhood structure. The proposed tool encompasses a method of analyzing mortality data and its geographic information in R to identify the most suitable order of neighborhood structure for discrete modelling. A Gaussian spatial process with a Matérn correlation function is employed to achieve this (Guttorp, 2006). The mortality data are fitted to a spatially continuous model, utilizing the coordinates of the centroids of municipalities as spatial locations. A model-based estimation of the practical correlation range of spatial correlation is obtained by utilizing the Gaussian process. This estimation provides valuable insight into the spatial relationships present in the mortality data. Moreover, it helps define the order k of the neighborhood matrix in the subsequent spatially discrete analysis. The aim is to ensure that the estimated practical correlation range best aligns with the mean distance between municipalities and their k th-order neighbors.

1.2 Objectives and research questions

The primary objective of this project is to develop an innovative method to optimize the process of selecting a neighborhood structure. Specifically, this method uses 2020 mortality data from Flanders, focusing on the region of Limburg, to enhance the accuracy and efficiency of spatial modelling. To streamline the utilization of this approach, the R code utilized in generating the results is shared in the appendices. If this method were optimized in future research, it could be consolidated into a user-friendly and convenient function within the R programming environment. By simply providing the required input data and specifying relevant parameters, researchers and practitioners could use this tool to significantly save time, effort, and computational resources when constructing discrete spatial models that necessitate carefully constructing a neighborhood structure.

1.3 Societal relevance and stakeholder awareness

Developing a tool that efficiently selects a neighborhood structure in a matter of minutes has significant implications for spatial epidemiology and early warning systems in infectious disease surveillance. This advancement is particularly crucial during times of crisis, such as the global pandemic exemplified by Covid-19, where the rate of disseminating results plays a pivotal role in disease management. By speeding up the selection process, this method offers a valuable tool to optimize the choice of a neighborhood structure a priori. Unlike traditional approaches that involve running multiple analyses to compare various neighborhood structures, this novel approach performs a single continuous spatial analysis, significantly reducing the time required, especially when dealing with large numbers of observed cases.

The potential beneficiaries of this research extend to a range of stakeholders, including governmental public health institutions, policymakers, scientists, and non-statisticians who rely on these models for decision-making processes and require enhanced efficiency. The implications also extend to researchers across disciplines utilizing spatial data analysis in R, as the availability of such a tool would streamline their workflow and increase the effectiveness of their research endeavors.

In spatial epidemiology, the ability to efficiently identify the most appropriate neighborhood structure contributes to the timely identification of disease hotspots, spatial patterns, and clusters. This knowledge empowers public health institutions to make informed decisions regarding resource allocation, targeted interventions, and the implementation of effective control measures. Moreover, policymakers can benefit from a more efficient selection process, enabling them to respond promptly to emerging threats, allocate resources effectively, and develop evidence-based policies to mitigate the impact of infectious diseases.

Furthermore, the research community is expected to gain significant advantages from the development of this function. Researchers can devote more time and resources to analyzing and interpreting results by optimizing the process of selecting neighborhood structures. This ultimately promotes a deeper understanding of the underlying spatial dynamics of diseases, facilitates the discovery of novel insights, and encourages the development of innovative disease prevention and control strategies.

1.4 Ethical considerations

The aggregated dataset under consideration contains geographical information on the population and mortality numbers of the Belgian population. Even though the data is aggregated, it is important to prioritize the privacy and confidentiality of the individuals represented within the dataset. Ensuring privacy protection can prevent unauthorized access or disclosure of sensitive information.

Spatially discrete models emerge as a suitable approach in situations where data is presented in a spatially aggregated manner. The use of such models allows for the exploration of spatial patterns and the extraction of meaningful insights while upholding privacy standards. By conducting analyses at the aggregated level, the risk of potential re-identification or disclosure of personal information is significantly minimized. The use of spatial discrete models represents an ethical approach to handling sensitive data. By ensuring the confidentiality of observed individuals, researchers and policymakers can navigate the balance between data utility and privacy preservation. This methodological framework provides a robust platform for conducting the best possible analysis in situations where privacy concerns require the use of spatially aggregated data.

Chapter 2

Dataset

2.1 2020 mortality data

2.1.1 Data description

The dataset used for the analyses (Table 2.1) includes aggregated information on population and mortality numbers for all provinces in Belgium. The population number for 2020 is given in the `pop2020` variable, and the number of deaths is given in the `dth2020` variable column. For the development of the proposed methodology, data is used that is aggregated in regions of 4 and 16 square kilometers. The analysis is performed on both scales to explore the differences in results, which could provide valuable insights into the implications of data aggregation level in spatial statistics. The data includes geographical information on the country's borders and the boundaries for each subregion. The geometry variable describes the polygon of the aggregation and its location within the spatial field. Data analysis was first done using the 4 by 4 grid and later with the more precise 2 by 2 grid. The analyses are performed separately per province. The expected number of observations, E_{2020} in the data, is calculated by indirect standardization for the age and sex of the population. In the models, the observed number of cases is written as y and the expected number of cases as E . In the models, the standardized mortality ratio (SMR) of subregion i is calculated as the number of observed events divided by the number of expected events.

$$SMR_i = \frac{dth_{2020_i}}{E_{2020_i}}$$

A subregion with a high SMR value (> 1) indicates that there are more cases observed in that area than expected based on the age and sex-specific incidence proportions from the standard population.

TABLE 2.1: Description of Variables in the data

Variable	Description
<code>province</code>	Province in which the subregion is located
<code>geometry</code>	Simple feature geometry list column describing the polygon
<code>X</code>	X coordinate of the subregion centroid
<code>Y</code>	Y coordinate of the subregion centroid
<code>pop2020</code>	Population number in 2020
<code>dth2020</code>	Mortality number in 2020
<code>E2020</code>	Expected number of deaths
<code>SMR2020</code>	Standardised mortality ratio

2.1.2 Data exploration

The aggregated data can be visualized in a number of different ways. Here, the subset containing the information on the province of Limburg is visualized by plotting the spatial distribution of deaths in 2020 in blue, and the SMR (standardized mortality ratio) per aggregation in red. The same visualization was done on the 16km^2 data and on the 4km^2 data. Figures 2.1 and 2.2 show that the 4km^2 data contains more variability in terms of SMR values. This is explained by the fact that the data is aggregated on a smaller scale. There are therefore fewer observations per subregion. The spatial units with a zero population, such as in national parks or military domains, were left out of the analysis. These subregions appear as blank spaces in the figures.

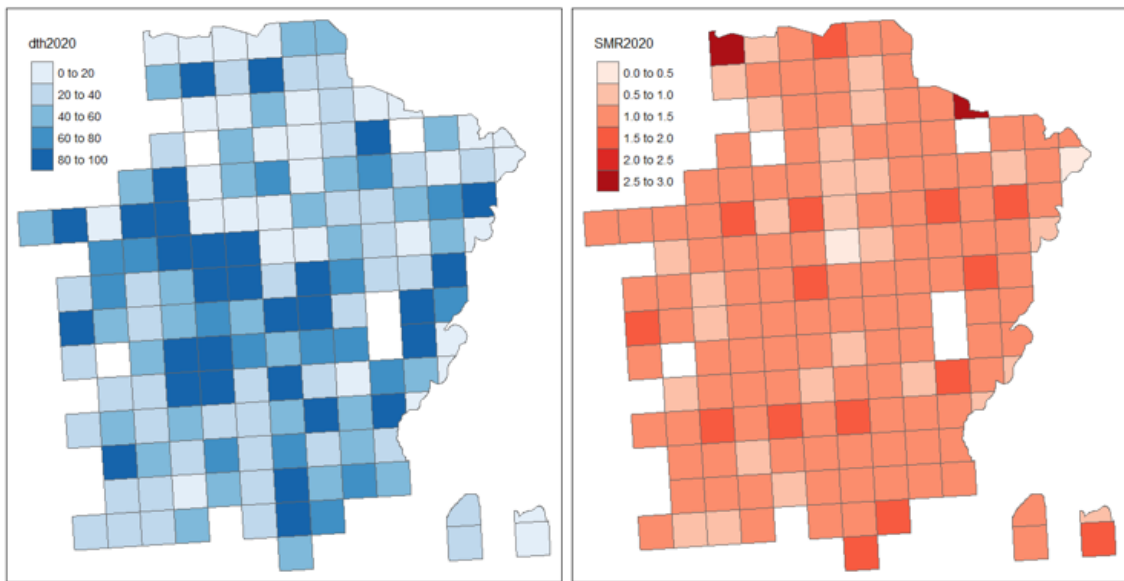


FIGURE 2.1: Left: mortality number of 2020 in Limburg. Right: SMR per aggregated subregion of Limburg. These results are shown for the 16km^2 data.

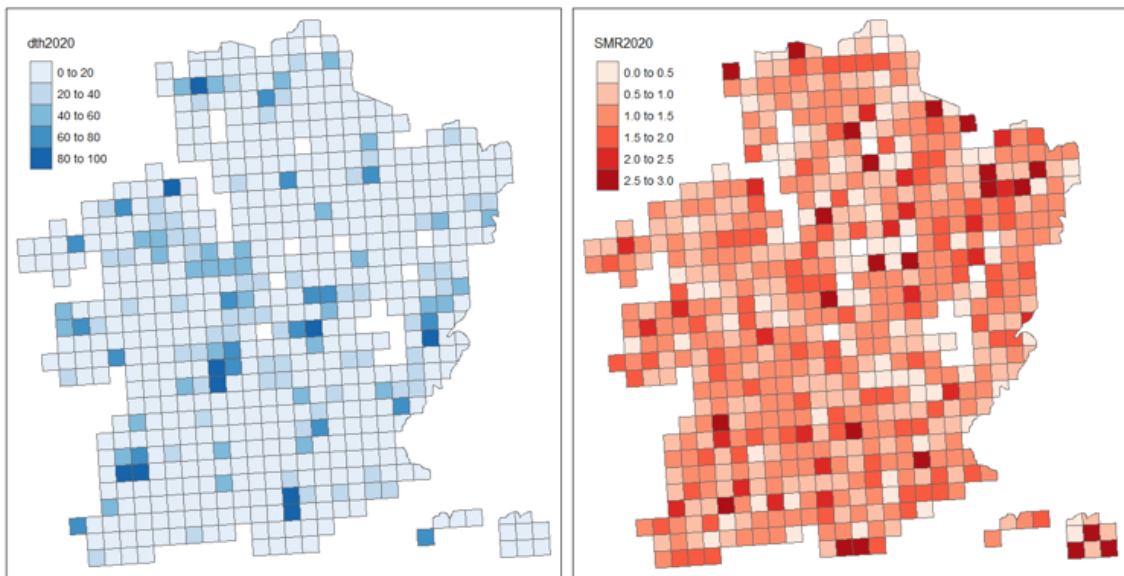


FIGURE 2.2: Left: mortality number of 2020 in Limburg. Right: SMR per aggregated subregion of Limburg. These results are shown for the $4km^2$ data.

Chapter 3

Methodology

3.1 Workflow

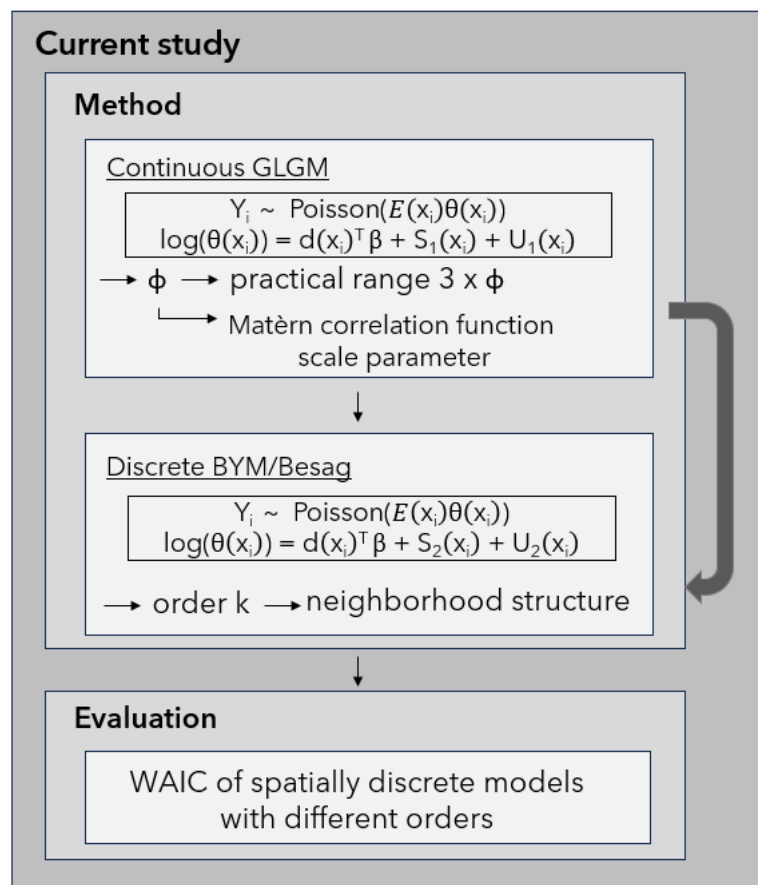


FIGURE 3.1: Workflow of the neighborhood structure selection method. The ϕ parameter refers to the scaling parameter of the Matèrn correlation function, given by $\rho(u; \phi, \kappa) = \{2^{\kappa-1} \Gamma(\kappa)\}^{-1} (u/\phi)^\kappa K_\kappa(u/\phi)$. In both models, S represents spatial correlation and U represents non-spatial heterogeneity. Subscript 1 is for the continuous spatial model and 2 for the spatially discrete model.

The method under development, illustrated in Figure 3.1, involves an innovative approach to tackling spatial data analysis by combining continuous and discrete spatial

modelling. The key idea is to utilize continuous spatial modelling to estimate the practical correlation range of spatial correlation within the data. This estimated range is then used to select an appropriate neighborhood structure for discrete spatial modelling. In the evaluation phase of the research, various discrete models with different orders of neighborhood structures are fitted, and their performance is compared using WAIC (Watanabe-Akaike Information Criterion) values (Wanatable, 2009; Gelman, 2014).

First, a spatially continuous model is fitted to the data to estimate the practical correlation range of the infectious disease of interest. The practical correlation range estimation is based on the ϕ parameter estimate of the GLGM. Then, a spatially discrete model is fitted with a neighborhood structure order closest to the previously estimated practical correlation range, meaning the maximum distance between two correlated centroids that aligns with the range parameter ϕ . It should be noted that the distances between spatial units can be variable, either on a regular or irregular grid. The mean distance between the centroids could be compared to the estimated practical correlation range of irregularly spaced spatial units. In this analysis, regular grids of 16km^2 and 4km^2 are used in which the distance between centroids are considered to be 3 and 6 km, respectively. This is the diagonal distance between centroids. Taking these distances will ensure that the continuous analysis selects the smallest suitable neighborhood structure order.

3.2 Spatially Continuous Model

A generalized linear geostatistical model or GLGM (Giorgi, 2018) was fitted to the data, which is a generalized linear model that incorporates the spatial dependence between data points. The GLGM consists of three components. The first component is the fixed component; however, in this case, no covariates were included in the model besides the X and Y coordinates of the data points, such that the fixed component only consists of an intercept coefficient in our scenario. The second component is the component that accounts for the spatial process, expressed as $S(x_i)$ in the model. It captures the unexplained spatial variation in Y_i and uses a spatial correlation function to explain how the data points are spatially correlated. In this case, the Matérn correlation function is used (Guttorp, 2006). The nugget effect U_i represents short-scale randomness or noise in the regionalized variable (Morgan, 2011). It captures the variability of explanatory variables that were not measured that are either not spatially structured or are spatially structured at a scale smaller than the minimum distance observed. The third component is the link function that links the linear predictor to the response variable via the log link since the Poisson distribution is used. Model parameter estimation is done via maximum likelihood estimation and MCML simulations.

Monte Carlo maximum likelihood (MCML) is a computational method to approximate the model parameters' maximum likelihood estimates (MLEs). Using Monte Carlo sampling, a huge number of simulated data sets are produced, each of which is based on a unique set of parameter values. The MLEs are calculated using maximum likelihood estimation methods for each simulated data set. The MCML estimates for the model parameters are then derived using the average values of the estimated parameters across all simulated data sets.

The initial values, including the initial nugget effect value for the maximum likelihood

estimation of the generalized linear geostatistical model, are based on the empirical variograms constructed for each model. These variograms, which are included in Appendix A, give a first look at the residual spatial correlation present in the data. A variogram measures the relationship between the variance or covariance of the variable and the distance between locations. The variogram is depicted by a line graph, where the y-axis is the estimated measure of spatial dependency and the x-axis indicates the distance. Positive spatial dependency can be observed by an increasing variogram, which suggests that values at sites near one another tend to be similar in observed values.

The Poisson log-linear geostatistical model assumes that the counts of observed deaths Y_i at location x_i are independent Poisson variables with mean $\lambda(x_i)$ (Diggle, 2019). Y_i are conditional on the Gaussian process $S(x_i)$ and on a set of independent zero-mean normally distributed random variables U_i with conditional expectations $\lambda(x_i)$, with

$$Y_i \sim \text{Poisson}(\lambda(x_i))$$

$$\log(\theta(x_i)) = d(x_i)^T \beta + S_1(x_i) + U_1(x_i)$$

In this study, the simpler version of the model becomes

$$\log(\lambda(x_i)) = \beta_0 + S_1(x_i)$$

The Matérn correlation function was used for the Gaussian spatial process. It is defined as

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1} \Gamma(\kappa)\}^{-1} (u/\phi)^\kappa K_\kappa(u/\phi)$$

- ϕ is the scaling parameter
- κ is the shape parameter, set at 0.5
- K_κ is the Bessel function of the third kind of order κ

The spatially continuous model gives an estimated value for the ϕ parameter. This parameter is associated with the spatial correlation within the data and the practical correlation range of the phenomenon that is studied, in this case, Covid-19 related mortality. The practical correlation range is the distance at which $\rho(u) = 0.05$ and can be estimated as $3 \times \phi$ (Diggle, 2019). Ideally, the estimated practical correlation range of the spatially continuous model agrees with the spatially discrete model that fits the data the best. The practical correlation range should somewhat agree with the distance between the centroids of subregions defined by the selected neighborhood structure k . The discrete models with varying neighborhood structure orders can be compared using their WAIC values.

The WAIC, or Watanabe–Akaike information criterion, is a generalized version of the Akaike information criterion known as AIC (Bevans, 2023). It estimates the effective number of parameters to adjust for overfitting. Models with lower WAIC values are preferred. By prioritizing models with lower WAIC values, researchers can have more confidence in the selected model's ability to generalize well to new data and make robust predictions.

3.2.1 Analysis procedure in R

The continuous Poisson generalized geostatistical linear models are fitted using the `PrevMap` package in R (Giorgi, 2017; R Core Team (2022); RStudio Team (2020)). To initiate the model fitting process, the `spat.corr.diagnostic` function is used to generate informative variogram plots. These variograms serve as valuable diagnostic tools, enabling the estimation of initial values that are then utilized in the `glgm.LA` function. This function provides preliminary parameter estimates to incorporate into the `poisson.log.MCML` function. This function serves as a key component in the estimation of the final model estimates. The `poisson.log.MCML` function performs MCML simulations with 20000 iterations, a burn-in period of 1000, and a thinning factor of 8.

3.3 Spatially Discrete Model

The Besag, York, and Mollié (BYM) model (Lawson, 2006) was fitted to the data to model the prevalence of mortality across the spatial units over a regular grid. The structured component captures and represents the spatial heterogeneity across the study area, providing insights into the systematic spatial patterns. On the other hand, the unstructured component accounts for the random and unpredictable variation that remains unexplained after incorporating the structured component. This separation allows for a comprehensive analysis encompassing the known spatial patterns and the inherent randomness in the data. The BYM model relies on the specification of a neighborhood structure, which defines the spatial relationships between units and can be adjusted to different orders or levels of proximity, allowing a flexible and customizable modelling approach.

The BYM model is defined as

$$Y_i \sim \text{Poisson}(\lambda(x_i))$$

$$\log(\theta(x_i)) = d(x_i)^T \beta + S_2(x_i) + U_2(x_i)$$

- $\theta(x_i)$ is the relative risk of spatial unit i
- $U_2(x_i)$ is the uncorrelated heterogeneity $U(x_i) \sim N(0, \sigma_U^2)$
- $S_2(x_i)$ represents the correlated heterogeneity $[S(x_i)|S(x_j), i \neq j, \tau_S] \sim N(S(x_i), \sigma_i^2)$

Hyperpriors are the prior distributions given to a model's parameters in the context of Bayesian modelling. In the BYM model, hyperpriors are used to provide the prior distributions for the parameters that regulate the spatial structure and the unstructured variation. These precision parameters regulate the strength or smoothness of the spatial dependence, influencing the degree to which neighboring areas are expected to be similar. In this analysis, Gamma uninformative hyperpriors were chosen to allow the data to drive the inference as much as possible. The hyperpriors for α , σ_S^2 and σ_U^2 were assigned as follows:

- $\tau_S \sim \text{Gamma}(0.001, 0.001)$
- $\tau_U \sim \text{Gamma}(0.001, 0.001)$
- $\tau_\alpha \sim \text{Gamma}(0.001, 0.001)$

3.3.1 Analysis procedure in R

The BYM models are fitted using the NIMBLE package in R (NIMBLE Development Team, 2023). NIMBLE (Numerical Inference for Statistical Models with Bayesian and Likelihood Evaluation) is a package for conducting Bayesian analysis, and it provides functions to fit complex spatial models. First, the model structure, including the response variable, coordinates, and neighborhood structure is defined. The priors and hyperpriors are also specified beforehand in the syntax. The model is then compiled and maximum likelihood estimation as well as Markov chain Monte Carlo (MCMC) sampling is used to estimate the model parameters. The number of iterations is set at 10000 and the burn-in period to 5000. Two chains are run using a thinning factor of 10. From the produced posterior samples, estimates for the parameters are obtained. The model produces relative risk values for each spatial unit, estimates for the unstructured and structured variance components, and a WAIC value for the model.

3.4 Comparing neighborhood structures

This research study utilizes the 2020 COVID-19 mortality dataset to develop a robust and efficient tool to optimize the selection of the most appropriate k th-order neighborhood structure when fitting a spatially discrete geostatistical model. By developing this tool, researchers and practitioners can make more informed decisions regarding the choice of neighborhood structure, leading to improved modelling outcomes and enhanced understanding of the spatial patterns of Covid-19 mortality.

A comparative analysis of the models using different neighborhood structures was conducted to determine the best-fitting order of the neighborhood structure. This analysis evaluated each model's estimated Watanabe-Akaike Information Criterion (WAIC) values. The WAIC serves as an essential criterion for model selection, as it balances the model's goodness of fit with its complexity, effectively addressing the issue of overfitting. It is computed from the log point-wise predictive density. Lower WAIC values indicate better model performance, suggesting a better agreement between the model and the data.

The WAIC value of a model is defined as follows (Gelman, 2014):

$$WAIC = lppd - pWAIC.$$

The lppd (log point-wise predictive density) is defined as

$$lppd = \sum_{i=N}^G \log\left(\sum_{g=1}^G p(y_i|\theta^g)\right),$$

where N is the number of areas and G is the sample size of area I and θ are the model parameters to be estimated.

The penalty term $pWAIC$ is defined as the variance of the lppd

$$pWAIC = \sum_{i=1}^N Var(\log(p(y_i|\theta^g))).$$

There is not one correct way to evaluate the "significance" of the difference in WAIC values of two models. In general, it is not possible to provide a principled threshold of difference that makes one model "significantly" better than another (McElreath, 2016). However, in this analysis, a difference in WAIC values of three or more was seen as evidence in favor of the model with the lower WAIC (Tillman, 2017). Models with a WAIC difference of three or less to the model with the lowest WAIC, are considered to have a similar fit.

The ultimate objective is to identify the model with the lowest WAIC value, indicating the best fit to the data. It is hypothesized that this optimal model, in theory, should align closely with the neighborhood structure that agrees the most with the practical correlation range of spatial correlation observed in the mortality cases. To explore a comprehensive range of possibilities, neighborhood structures up to order 5 are considered in the analysis, enabling a broad examination of the spatial relationships. Next to the order, it is also required to specify the scope of the neighborhood. In this analysis, the "queen" matrix, neighboring spatial units are those having at least one vertex or border in contact with the reference element (Moura, 2020). This is a generally accepted way of defining neighbors as it is the broadest scope of directly linking spatial units.

To ensure the accuracy and reliability of the modelling process, regions with zero population were excluded from the data. This exclusion was implemented to mitigate potential computational challenges that could arise when modelling in areas with no population, thus ensuring a more focused and meaningful analysis specifically related to the areas with relevant data.

The analysis was done twice, once using the 4 km^2 grid and once using the larger 16 km^2 grid. The results are then compared to see at what level they are agreeable. Finer grids contain fewer observations per spatial unit. Smaller spatial units encompassed by the finer grid cover a smaller geographic area, leading to a sparser distribution of observations within each unit. However, they provide better spatial resolution between observations by giving a more detailed perspective. On the other hand, the larger grid, encompassing fewer spatial units, may yield a higher number of observations per unit due to the more considerable geographic coverage of each unit.

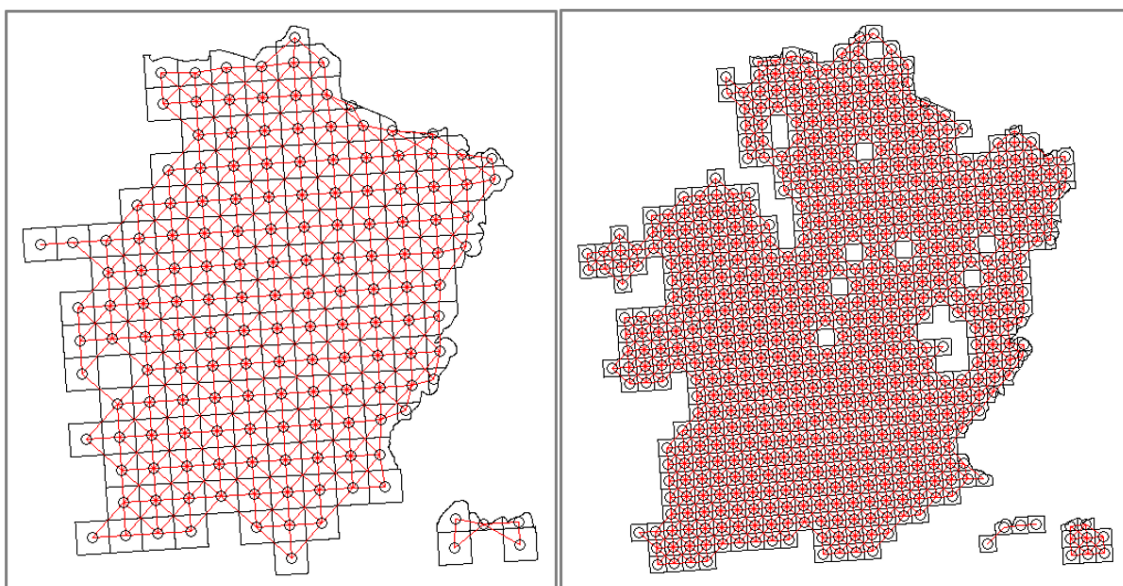


FIGURE 3.2: Comparison of 1st order neighborhood structures of Limburg using the 16 km² data on the left and the 4 km² on the right.

3.5 Simulated data

The analysis procedure described above was repeated on simulated data. By using artificial spatial data, it becomes possible to gain more control over the spatially structured heterogeneity, thereby enhancing the understanding of the underlying processes. In particular, the simulation of spatial correlation and practical correlation range throughout the entire spatial field would provide a valuable opportunity to examine the consistency and efficacy of the newly developed method.

Although this approach may lack realism in terms of replicating actual scenarios, it represents an ideal testing ground for the new method. By imposing a pre-specified practical correlation range for the disease, the simulated data would enable a more comprehensive evaluation of the method's performance. Consequently, the results obtained from such an experiment would likely offer a clearer understanding of the functionality and accuracy of the method, specifically in terms of suggesting the correct neighborhood structure order based on the predetermined practical correlation range by keeping the random noise within the data to a minimum.

Using simulations provides advantages when drawing conclusions regarding the adequacy of the tool. With the absence of confounding factors present in real-world data, it becomes easier to isolate the specific effects of the method itself. This controlled environment enables researchers to examine the tool's capabilities more effectively, providing valuable insights into its strengths and limitations.

A total of 20 datasets were generated with a neighborhood structure order of 3. The intercept α was set to 0.15 and the spatial random effect variance σ_{ξ}^2 to 0.5. These parameter values were obtained from the analyses of the case study data. These are shown in the Results section. The data were generated using multinomial counts to simulate

case counts and relative risks for each spatial unit, considering the spatial correlation corresponding with the neighborhood structure order of 3. Each of these datasets was analyzed using a continuous GLGM to estimate a practical correlation range parameter ϕ and then analyzed discretely with neighborhood structure k going from 1 up to 5 and with the order suggested by the continuous model.

Chapter 4

Results

4.1 Case study

4.1.1 Comparative results

The estimated scale parameter ϕ from the spatially continuous generalized linear geostatistical model is listed in the second column in the tables below. Tables 4.1 and 4.2 present the estimations for the data on the 4 km^2 and 16 km^2 scale, respectively. The centroids are 2 km spaced from each other for the 4 km^2 data, and 4 km in the 16 km^2 data. Diagonally, the centroids are approximately 3 and 6 km apart, since the spatial units are squares. To select the neighborhood order, these larger distances will be used to select the smallest order possible from the continuous analysis results. The practical range, $3 \times \phi$ is then calculated by dividing it by 3 or 6 and rounded to the next integer. This is the suggested neighborhood structure order, shown in brackets. The other columns show the WAIC values of the spatially discrete BYM models using neighborhood structures with the order going from $k=1$ to $k=5$. The last column gives the WAIC of the model using the order k suggested by the estimated practical correlation range. The significantly lowest WAIC values for each province are written in bold.

Results on the 4 km^2 scale

TABLE 4.1: Comparison of result per province of estimated practical correlation range and WAIC value of the BYM model with k 'th order neighborhood structure using the 4 km^2 data.

	$3 \times \phi$ (order)	k=1	k=2	k=3	k=4	k=5	suggested k
Limburg	5.19 (2)	2715.88	2713.63	2713.05	2714.37	2714.46	/
Antwerpen	50.31 (17)	3409.44	3410.39	3408.86	3409.99	3410.49	3402.63
Vlaams-Brabant	34.77 (12)	3196.93	3200.51	3199.68	3200.99	3198.84	3201.20
Oost-Vlaanderen	17.61 (6)	3818.12	3817.79	3817.46	3819.17	3818.98	3816.75
West-Vlaanderen	8.07 (3)	3262.18	3261.40	3261.23	3259.92	3264.43	/

Results on the 16 km² scale

TABLE 4.2: Comparison of result per province of estimated practical correlation range and WAIC value of the BYM model with k-th order neighborhood structure using the 16 km² data.

	$3 \times \phi$ (order)	k=1	k=2	k=3	k=4	k=5	suggested k
Limburg	47.37 (8)	991.35	990.78	990.98	990.04	991.13	992.97
Antwerpen	151.2 (25)	1347.94	1344.92	1344.60	1346.19	1345.46	1342.11
Vlaams-Brabant	32.4 (6)	1104.57	1102.77	1102.96	1102.94	1103.48	1103.36
Oost-Vlaanderen	117.75 (20)	1312.57	1294.37	1295.26	1294.66	1294.91	1293.46
West-Vlaanderen	24.12 (4)	1399.95	1399.53	1400.40	1399.85	1400.67	/

4.1.2 Sensitivity Analysis

A sensitivity analysis was performed on the discrete model using different gamma hyperpriors to check the robustness of the results (Vranckx et. al., 2023). The models were fitted both using Gamma(0.01, 0.01) priors and Gamma(0.001, 0.001) priors for the σ_S parameter. The tables below show the models' comparison using the different priors and on the four-by-four and two-by-two data for Limburg. The priors suggest that they influence the results to some degree. Because of this, the Gamma priors for all analyses were changed from Gamma(0.01, 0.01) to Gamma(0.001, 0.001), since these priors are less informative.

TABLE 4.3: 4 km²

		Mean	Median	Std.Dev.	95%CI L	95%CI U
Gamma(0.01, 0.01)	alpha	0.1610	0.1612	0.0148	0.1319	0.1895
	sigma.u	0.0068	0.0059	0.0034	0.0026	0.0158
	sigma.v	0.0226	0.0224	0.0051	0.0131	0.0336
Gamma(0.001, 0.001)	alpha	0.1569	0.1568	0.0144	0.1284	0.1840
	sigma.u	0.0035	0.0015	0.0042	0.0003	0.0162
	sigma.v	0.0231	0.0230	0.0057	0.0122	0.0348

TABLE 4.4: 16 km²

		Mean	Median	Std.Dev.	95% CI L	95% CI U
Gamma(0.01, 0.01)	alpha	0.1703	0.1696	0.0138	0.1446	0.1984
	sigma.u	0.0105	0.0088	0.0064	0.0031	0.0261
	sigma.v	0.0061	0.0057	0.0025	0.0026	0.0122
Gamma(0.001, 0.001)	alpha	0.0275	0.0270	0.0048	0.0194	0.0382
	sigma.u	0.0037	0.0024	0.0037	0.0005	0.0142
	sigma.v	0.0042	0.0037	0.0026	0.0008	0.0108

4.1.3 Results for Limburg

Results on the 4 km² scale

The relative risks compare the risk of higher mortality cases than expected to the overall risk. Values lower than 1 indicate a small relative risk, while values larger than 1 indicate

higher relative risks. The exceedance probability is the probability of the relative risk being larger than 1. A high exceedance probability means that there is a high probability that the relative risk of a specific spatial unit is above 1, meaning that the subregion has a higher risk of mortality as compared to the standardization region. The relative risks and exceedance probabilities for the Limburg data are shown in Figure 4.1.

These results show that most spatial units have relative risks of around 1 to 1.4. These spatial units have more mortality cases than expected. A few spatial units have values smaller than 0.9 and values over 1.5. The exceedance probabilities range from 0.6 to 1, suggesting that some areas are more likely to have relative risks exceeding 1 than others. The exceedance probability plot also suggests that areas with similar values often surround high-probability areas. The province of Limburg struggled with large numbers of Covid-19 cases in 2020, more so than some other provinces. This explains the high relative risk values and high exceedance probabilities across Limburg, since these results are standardized over Belgium as a whole.

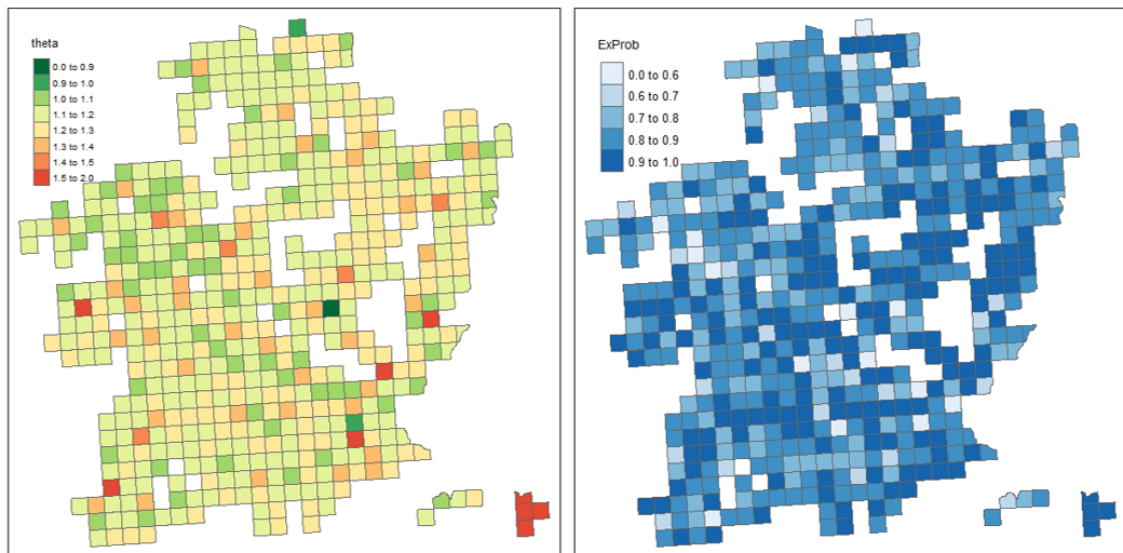


FIGURE 4.1: Relative risks and exceedance probabilities of the 4 km² spatial units

Results on the 16 km² scale

These results, shown in Figure 4.2, show much less heterogeneity than the results on the smaller spatial scale. This shows that the variability between spatial units decreases when observations are aggregated together on a larger scale. The spatial units have relative risks ranging from 1.1 to 1.3, with a few exceptions in the west and southeast (Voeren region). This homogeneity of results is also reflected in the exceedance probability plot. This plot suggests that the spatial units have high exceedance probabilities.

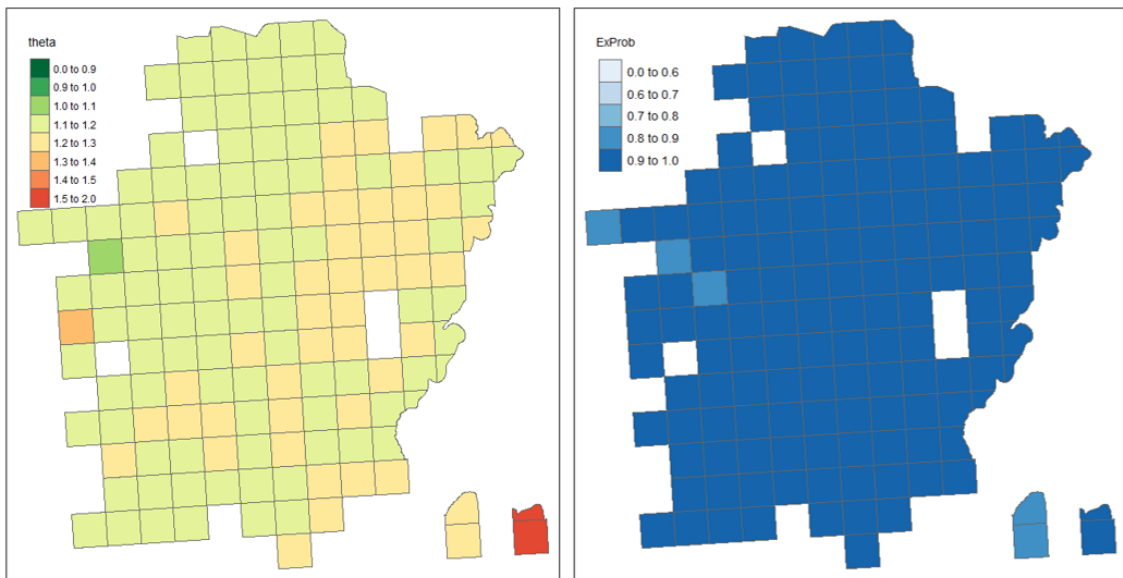


FIGURE 4.2: Relative risks and exceedance probabilities of the 16 km² spatial units

4.2 Simulation study

4.2.1 Exploration of simulated data

The simulations were done only on the smaller 4 km² scale. The distribution of the mortality cases and the SMR values of the spatial units show mortality case clusters that are more clearly distinguishable than in the actual mortality data. These SMR values are standardized over the province of Limburg, not over Belgium as is the case for the case study data.

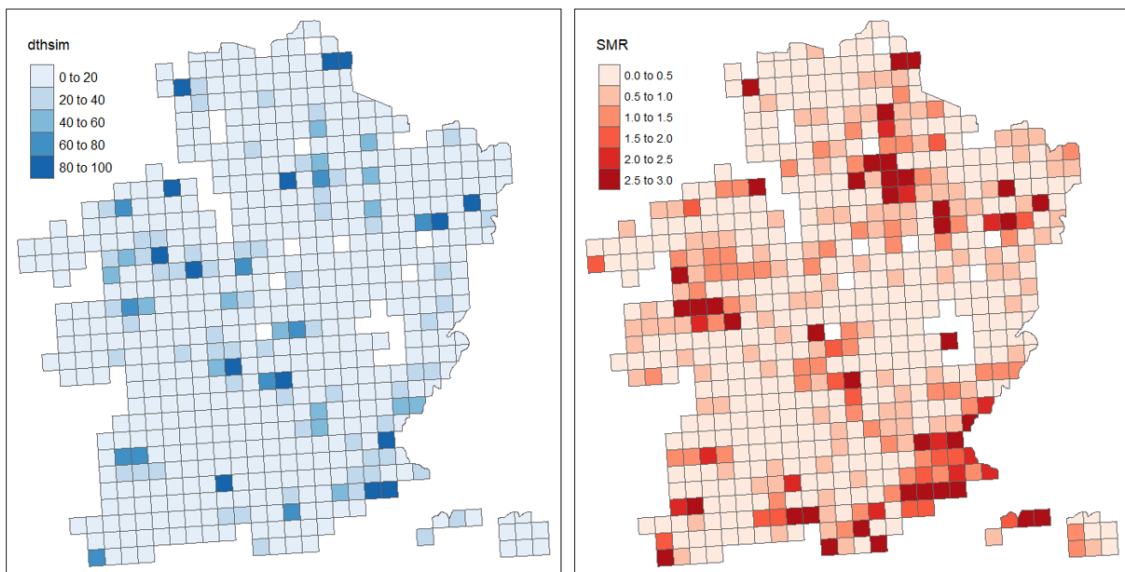


FIGURE 4.3: Left: mortality numbers of one of the simulated datasets. Right: SMR of the spatial units in one of the simulated datasets.

4.2.2 Results for the simulated datasets

Figure 4.4 shows the relative risks, standardized over the province of Limburg itself, and the probabilities of the relative risk exceeding 1. These plots show that there is a clear spatial correlation present in the data as suggested by the clustering of the cases.

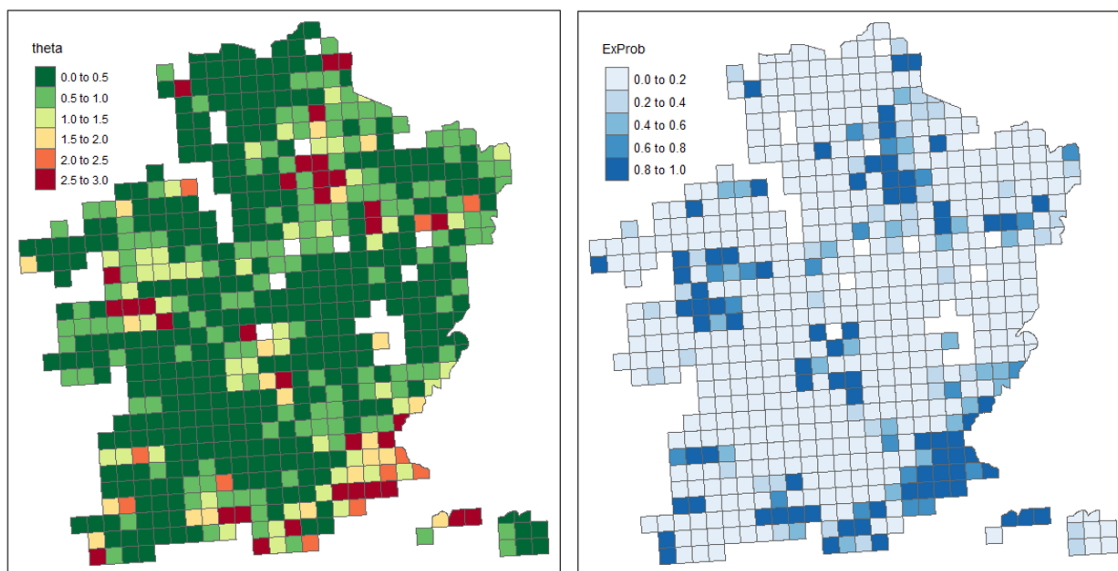


FIGURE 4.4: Relative risks and exceedance probabilities for one of the simulated datasets.

Table 4.5 gives an overview of the results of the analysis on the 20 simulated datasets. They were first analyzed continuously with a geostatistical poisson model. MCML estimation was used for the $3 \times \phi$ parameter estimation which is the practical correlation range parameter. In brackets, the neighborhood structure order is given which approaches this practical correlation range. Then, discrete spatial modelling was done on all 20 simulated datasets using order 1 up to 5 and the order which was suggested in the second column. The WAIC values are listed in the table. The significantly lowest WAIC values are written in bold, those being the lowest WAIC's per row with a difference of at least 3.

TABLE 4.5: Comparison of result per simulated dataset of estimated practical correlation range and WAIC value of the BYM model with k 'th order neighborhood structure on the 4 km² scale

Dataset	$3 \times \phi$ (order)	k=1	k=2	k=3	k=4	k=5	suggested k
1	20.7 (7)	3603.94	3610.60	3612.59	3607.24	3610.24	3609.75
2	30.51 (11)	3510.78	3540.76	3537.70	3539.80	3538.68	3539.71
3	19.26 (7)	3629.19	3605.07	3602.71	3604.04	3600.88	3601.89
4	20.88 (7)	3682.09	3653.54	3652.90	3653.59	3654.24	3652.04
5	13.92 (5)	3555.07	3574.62	3575.14	3573.96	3577.71	/
6	12.93 (5)	3538.73	3606.08	3606.14	3610.19	3607.50	/
7	25.74 (9)	3557.66	3608.73	3610.04	3605.57	3604.46	3608.75
8	27.63 (10)	3495.30	3522.12	3524.46	3519.24	3521.94	3521.26
9	26.28 (9)	3665.28	3637.84	3635.35	3633.17	3633.35	3634.25
10	19.08 (7)	3542.87	3612.72	3610.56	3610.94	3608.20	3612.39
11	24.24 (9)	3817.17	3761.77	3763.95	3762.71	3761.59	3764.41
12	24.93 (9)	3703.19	3651.57	3651.20	3650.06	3648.60	3650.26
13	36.51 (13)	3702.47	3632.25	3628.94	3631.80	3634.46	3630.70
14	27.90 (10)	3717.46	3671.71	3675.49	3674.54	3675.97	3673.09
15	26.28 (9)	3706.44	3658.06	3660.76	3657.86	3658.78	3659.15
16	18.30 (7)	3595.83	3628.12	3627.19	3628.74	3626.33	3630.11
17	22.53 (8)	3558.40	3573.58	3574.41	3575.26	3577.41	3573.97
18	20.01 (7)	3535.64	3596.522	3600.35	3599.17	3599.10	3600.31
19	21.09 (8)	3742.12	3673.66	3675.33	3675.82	3677.07	3675.86
20	28.47 (10)	3651.42	3680.49	3684.18	3682.64	3677.25	3682.89

Chapter 5

Discussion and conclusions

5.1 Spatially continuous vs. spatially discrete model

The method under development centers on harnessing continuous spatial modelling to gauge the practical correlation range of spatial correlation inherent in spatial data. Leveraging this estimation, the method determines an appropriate neighborhood structure for discrete spatial modelling. This technique's efficacy is assessed by fitting discrete models with varying orders of neighborhood structures and subsequently measuring their performance using the WAIC criterion.

Despite its conceptual promise, the method has encountered substantial challenges in achieving satisfactory outcomes. Notably, the neighborhood structure order inferred from the practical correlation range estimate often does not align with the optimal model choice, as indicated by comparing WAIC values. This discrepancy is often caused by a tendency for the method to frequently overestimate the spatial correlation range, even when tested on simulated data. Given these limitations, it is evident that further extensive research is imperative to fine-tune and optimize the method. Ultimately, the overarching goal of this approach is to streamline the process of selecting an appropriate neighborhood structure order by relying on a solitary continuous model. Based on the outcomes of this investigation, it is apparent that the method, in its current state, is not yet suitable for practical use in selecting neighborhood structures, as it tends to overestimate spatial correlation and consequently recommends higher-than-optimal neighborhood structure orders.

Although it is not yet clear at this point what causes this discrepancy in the results, it could be that the spatially discrete and spatially continuous models are parameterized in a way that makes them too different regarding how they model the data. Although the sensitivity analysis section only covers part of the total dataset, it suggests that the spatially discrete BYM model is somewhat sensitive to changes set to the hyperparameter Gamma priors. The generalized linear geostatistical model is also known to be sensitive to their initial values for ϕ and the variance of the nugget effect, as the results of the MCML estimation for the geostatistical Poisson model can vary when these initial values are changed.

The discrete spatial models fitted onto the 2020 mortality data also show little difference in WAIC values, especially on the 16 km^2 scale. In most cases, the model with the neighborhood structure order suggested by the estimated practical correlation range parameter has a good fit according to the WAIC values. In the larger 16 km^2 scale data, there is little to no significant differences between the WAIC's of the discrete spatial models with

different orders. This could partly be explained by the more homogeneous nature of the case distribution across larger spatial units compared to the smaller scale data. These results suggest that there is not much to gain from optimizing the neighborhood structure selection, as there is often not one preferable model. More research into the model selection method is needed.

The inconsistent results could also be partly a result of the second-order stationarity assumption (Fouedjio, 2017). This is the assumption that regardless of the two points you select, the covariance is the same between any two that are spaced in the same way and in the same direction. The distance between any two values determines the covariance, not where the values are. This assumption might not be accurate because of regional impacts.

5.2 Conclusion

These results have shown that developing a new method for selecting the neighborhood structure order can be challenging when using real as well as simulated data. The practical correlation range of the cases is inconsistent across regions and between simulated data sets according to the spatially continuous model. This leads to issues when relying on this practical correlation range to pick the order of neighborhood structure for spatially discrete modelling, since the selection of the discrete spatial model is dependent on the estimation of the range of spatial correlation by the spatially continuous model. There are also optimization steps to be taken at the evaluation step.

5.3 Limitations

The development of the neighborhood structure optimization tool is met with some limitations. The data is aggregated into two-by-two or four-by-four kilometers subregions. These aggregated cases can still be used for continuous modelling, but any granularity within these subregions is lost. Furthermore, modelling the data using a wider grid of four by four kilometers, give different results regarding the suggested order of neighborhood structure and best-fitting models.

Another limitation is that a choice has to be made as how to measure practical correlation range in the spatial continuous analysis. In this thesis, the mean was used, but one could argue that confidence intervals are more informative and thus better to base the practical range correlation on. However, confidence intervals would lead to multiple possible neighborhood structure orders. This would then complexify the analysis. Also, the problem arises that the practical range from the spatial continuous analysis actually assumes a circular "correlation zone," while the queen neighborhood is a square.

After analyzing the Covid-19 mortality dataset, it is clear that working on a different scale can suggest other preferred models. This is because of the practical correlation range that has a different estimate in the continuous model and varying WAIC values of the discrete models. Thus, it is important to remember that this method can not always pick the definitive best-fitting model because of these limiting factors. One should always keep in mind the context of the research question and the plausibility of the estimated practical correlation range in the field. A sensitivity analysis is a helpful extra step to ensure the

chosen model is sufficiently robust.

The dataset covers the mortality and population numbers in 2020. Whilst this is useful information when modelling the Covid-19 cases in Belgium, the mortality cases include all mortality in this year, and not only deaths following a Covid-19 infection. This means that the data is not very specific for modelling Covid-19.

Another limitation is the computational limits of the system that was used to run these models. The aggregated data was split into subsets per province of Flanders to make the running smoother. This does mean that neighboring regions that lie in different provinces were now not in the same subset, resulting in a loss of information. There are also no overall results available for the entire dataset covering the whole spatial field of Flanders.

5.4 Further Research

In future research, several steps could be explored to optimize the method and address its current limitations. To refine the practical correlation range estimation, alternative estimation techniques could be applied to potentially estimate the range of spatial correlation more accurately. This could involve exploring different spatial models or incorporating additional covariates to enhance the estimation process. The model selection criteria could also be further investigated. In this thesis, only the WAIC values were compared to select the best-fitting model. Instead of relying solely on WAIC values, one could consider incorporating other Bayesian model selection criteria such as DIC (Deviance Information Criterion), BIC (Bayesian Information Criterion), and MSPE (Mean Squared Prediction Error), which might provide different insights into model performance. Expanding the research by applying more extensive simulations would facilitate a more thorough exploration of various scenarios. In this analysis, only 20 datasets were generated, all with the same initial values and neighborhood structure order of 3. In future research, it would be a helpful step toward optimizing the method to simulate data with multiple spatial correlation ranges by varying the order of neighborhood structure. Researchers can assess the method's robustness and adaptability across various circumstances by systematically varying these parameters, such as the practical correlation range or values for the intercept and variance for the spatial effect. This can help identify scenarios where the method struggles and provide insights into potential improvements or adjustments needed to make the method more robust and reliable. This evaluation would enhance confidence in the tool's performance and its potential applicability in real-world situations. Future researchers could also opt for a hybrid approach that combines the continuous and discrete modelling techniques in more intricate ways, for example, a tool that adjusts the estimated practical correlation range based on the performance of discrete models, creating a feedback loop that iteratively adjusts the neighborhood structure selection.

Bibliography

- [1] Baddely, A. (2007). Spatial Point Processes and their Applications. Weil, W. (eds) Stochastic Geometry. Lecture Notes in Mathematics, vol 1892. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-38175-4_1
- [2] Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2), 192-225. <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x>
- [3] Bevans, R. (2023). Akaike Information Criterion | When & How to Use It (Example). Scribbr. <https://www.scribbr.com/statistics/akaike-information-criterion/>
- [4] Chan, Y. H. (2004). Biostatistics 201: linear regression analysis. *Age (years)*, 80, 140.
- [5] Chou, Y.H. (1995). Spatial pattern and spatial autocorrelation. In: Frank, A.U., Kuhn, W. (eds) *Spatial Information Theory A Theoretical Basis for GIS*. COSIT 1995. Lecture Notes in Computer Science, vol 988. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-60392-1_24
- [6] Dunn, P. K., & Smyth, G. K. (2018). *Generalized linear models with examples in R* (Vol. 53). New York: Springer.
- [7] Fouedjio, F. (2017). Second-order non-stationary modelling approaches for univariate geostatistical data. *Stoch Environ Res Risk Assess* 31, 1887–1906. <https://doi.org/10.1007/s00477-016-1274-y>
- [8] Garcia, T. P., & Marder, K. (2017). Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington’s Disease as a Model. *Current neurology and neuroscience reports*, 17(2), 14. <https://doi.org/10.1007/s11910-017-0723-4>
- [9] Diggle, P.J., & Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health: Methods and Applications* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315188492>
- [10] Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., & Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of the total environment*, 739, 140033.
- [11] Gelman, A., & Hill, J. (2006). *Data analysis using regression and multi-level/hierarchical models*. Cambridge university press.
- [12] Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2014). *"Bayesian Data Analysis, 3rd ed."*. CRC Press: Boca Raton, FL.
- [13] Giorgi E, Diggle P (2017). “PrevMap: An R Package for Prevalence Mapping.” *Journal of Statistical Software*, 78. doi:10.18637/jss.v078.i08.

- [14] Giorgi, E. On the goodness-of-fit of generalized linear geostatistical models. 2018. *Spatial Statistics*, Volume 28, Pages 79-83, ISSN 2211-6753, <https://doi.org/10.1016/j.spasta.2018.01.002>.
- [15] Guttorp, P., & Gneiting, T. (2006). Studies in the History of Probability and Statistics XLIX on the Matérn Correlation Family. *Biometrika*, 93(4), 989–995. <http://www.jstor.org/stable/20441340>
- [16] Huggins, R., & Thompson, P. (2017). Networks and regional economic growth: A spatial analysis of knowledge ties. *Environment and Planning A: Economy and Space*, 49(6), 1247–1265. <https://doi.org/10.1177/0308518X17692327>
- [17] Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173-189. <https://doi.org/10.1016/j.envsoft.2013.12.008>
- [18] Lichstein, J. W., Simons, T. R., Shriver, S. A., & Franzreb, K. E. (2002). SPATIAL AUTOCORRELATION AND AUTOREGRESSIVE MODELS IN ECOLOGY. *Ecological Monographs*, 72(3), 445-463. [https://doi.org/10.1890/0012-9615\(2002\)072\[0445:SAAAMI\]2.0.CO;2](https://doi.org/10.1890/0012-9615(2002)072[0445:SAAAMI]2.0.CO;2)
- [19] Moraga, Paula. (2023). *Spatial Statistics for Data Science: Theory and Practice with R*. Chapman & Hall/CRC.
- [20] Morgan, C. J. (2011). Theoretical and practical aspects of variography: in particular, estimation and modelling of semi-variograms over areas of limited and clustered or widely spaced data in a two-dimensional South African gold mining context (Doctoral dissertation).
- [21] Lawson, A.B. (2018). *Bayesian Disease Mapping: Hierarchical modelling in Spatial Epidemiology*, Third Edition (3rd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781351271769>
- [22] Lawson, A. B. (2006). *Statistical methods in spatial epidemiology*. John Wiley & Sons Ltd.
- [23] Lawson, A.B. (2021). *Using R for Bayesian Spatial and Spatio-Temporal Health modelling*. CRC Press, Taylor & Francis Group, LLC.
- [24] McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- [25] NIMBLE Development Team. (2023). *NIMBLE User Manual*. doi: 10.5281/zenodo.1211190. R package manual version 1.0.0. <https://r-nimble.org>.
- [26] R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [27] RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>

- [28] Thorup, K., Korner-Nievergelt, F., Cohen, E. B., & Baillie, S. R. (2014). Large-scale spatial analysis of ringing and re-encounter data to infer movement patterns: A review including methodological perspectives. *Methods in Ecology and Evolution*, 5(12), 1337-1350. <https://doi.org/10.1111/2041-210X.12258>
- [29] Tillman, Gabriel & Osth, Adam & van Ravenzwaaij, Don & Heathcote, Andrew. (2017). A Diffusion Decision Model Analysis of Evidence Variability in the Lexical Decision Task. *Psychonomic Bulletin & Review*. 24. 10.3758/s13423-017-1259-y.
- [30] Tobler, W. R. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region, *Economic Geography*, 46:sup1, 234-240, DOI: 10.2307/143141
- [31] Vranckx, M., Faes, C., Molenberghs, G., Hens, N., Beutels, P., Van Damme, P., Aerts, J., Petrof, O., Pepermans, K., & Neyens, T. (2023). A spatial model to jointly analyze self-reported survey data of COVID-19 symptoms and official COVID-19 incidence data. *Biometrical Journal*, 65 , 2100186. <https://doi.org/10.1002/bimj.202100186>
- [32] Watanabe, S. (2009). *Algebraic geometry and statistical learning theory* (Vol. 25). Cambridge university press.
- [33] Moura, Ana Clara & Fonseca, Bráulio. (2020). ESDA (Exploratory Spatial Data Analysis) of Vegetation Cover in Urban Areas—Recognition of Vulnerabilities for the Management of Resources in Urban Green Infrastructure. *Sustainability*. 12. 1933. 10.3390/su12051933.

Appendices

Appendix A

Variograms

A.1 Data on 16 km^2 scale

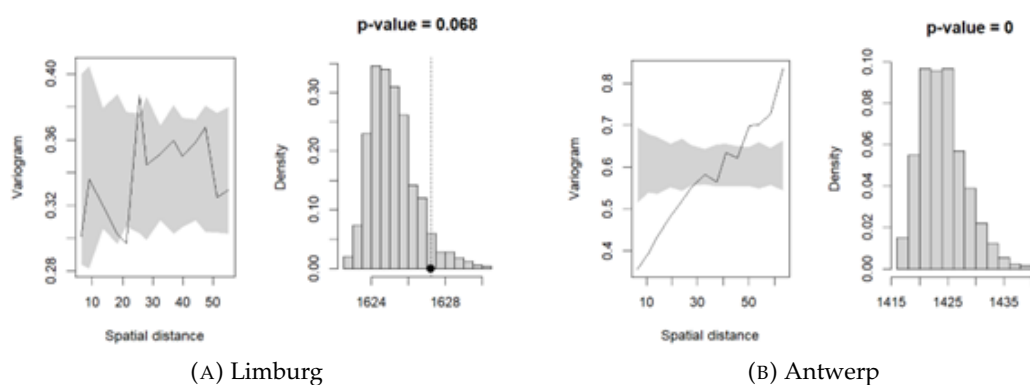


FIGURE A.1: Variograms for provinces Limburg and Antwerp

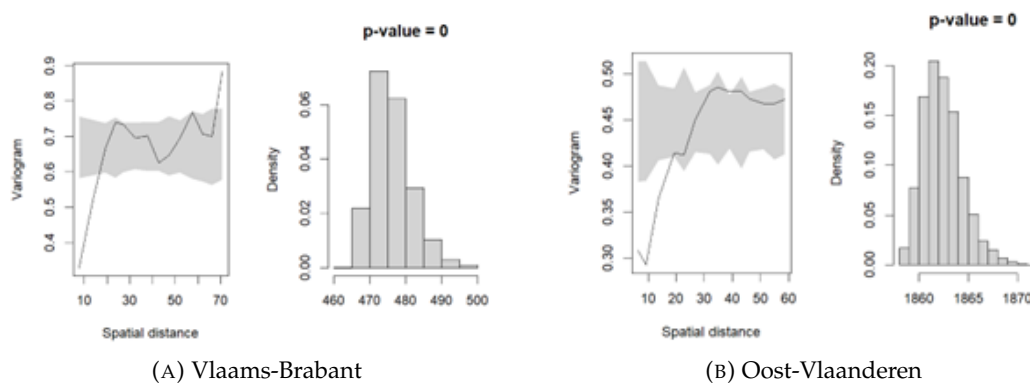


FIGURE A.2: Variograms for provinces Vlaams-Brabant and Oost-Vlaanderen

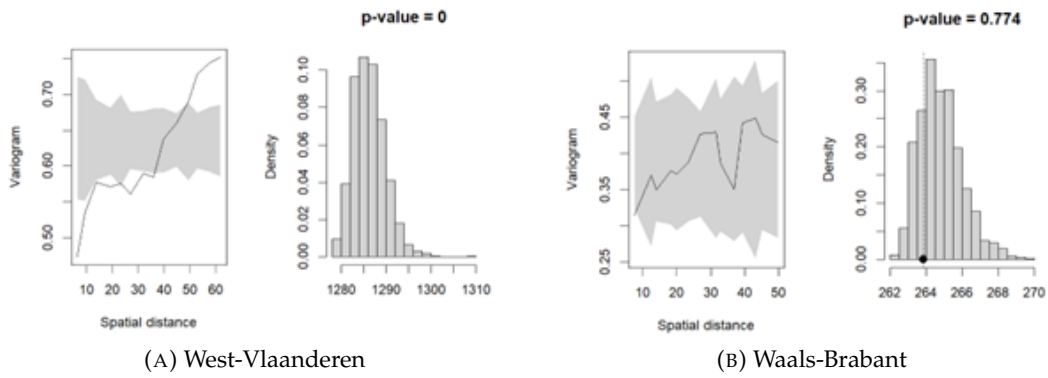


FIGURE A.3: Variograms for provinces West-Vlaanderen and Waals-Brabant

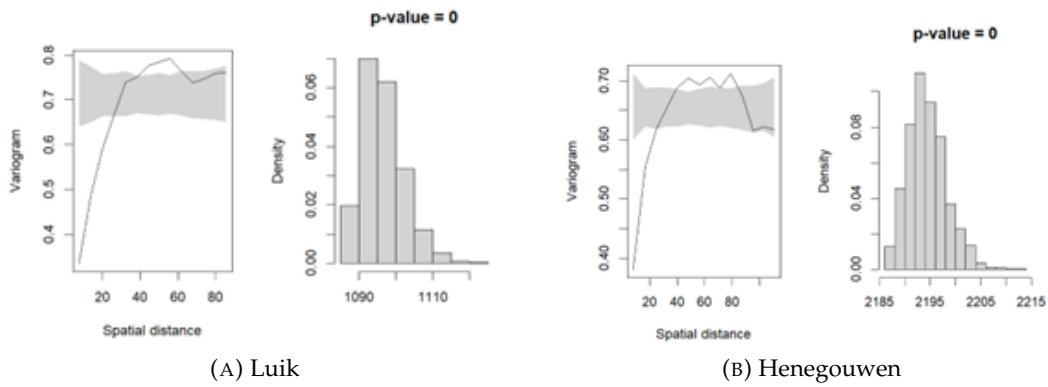


FIGURE A.4: Variograms for provinces Luik and Henegouwen

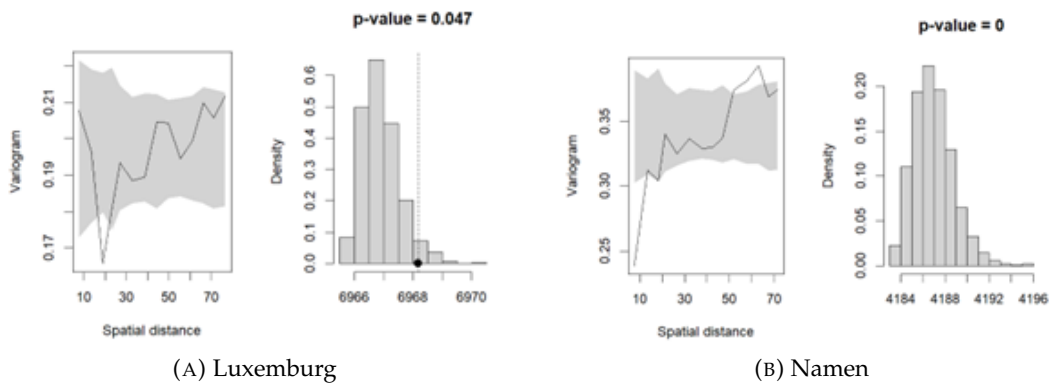


FIGURE A.5: Variograms for provinces Luxemburg and Namen

A.2 Data on 4 km² scale

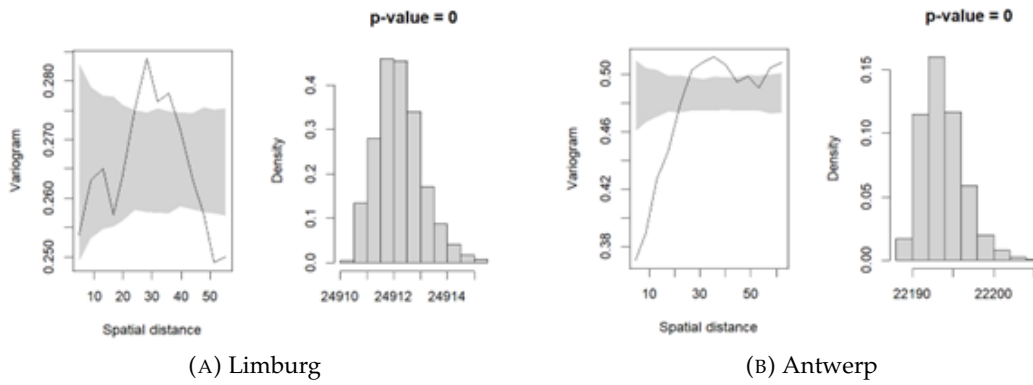


FIGURE A.6: Variograms for provinces Limburg and Antwerp

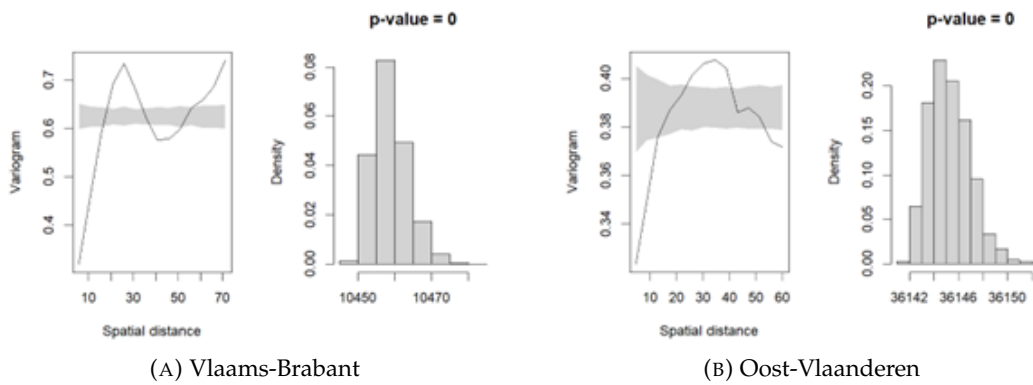


FIGURE A.7: Variograms for provinces Vlaams-Brabant and Oost-Vlaanderen

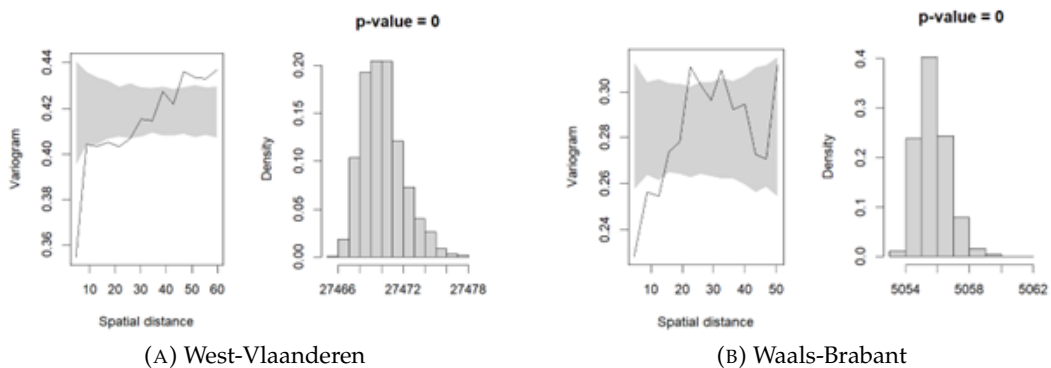


FIGURE A.8: Variograms for provinces West-Vlaanderen and Waals-Brabant

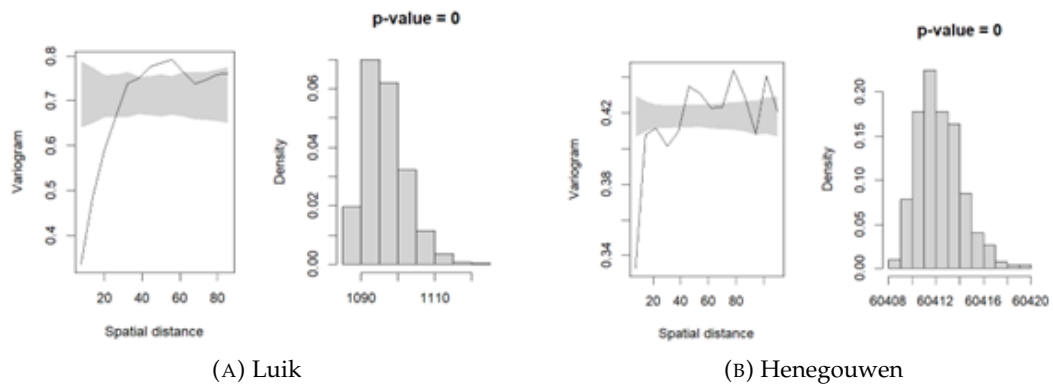


FIGURE A.9: Variograms for provinces Luik and Henegouwen

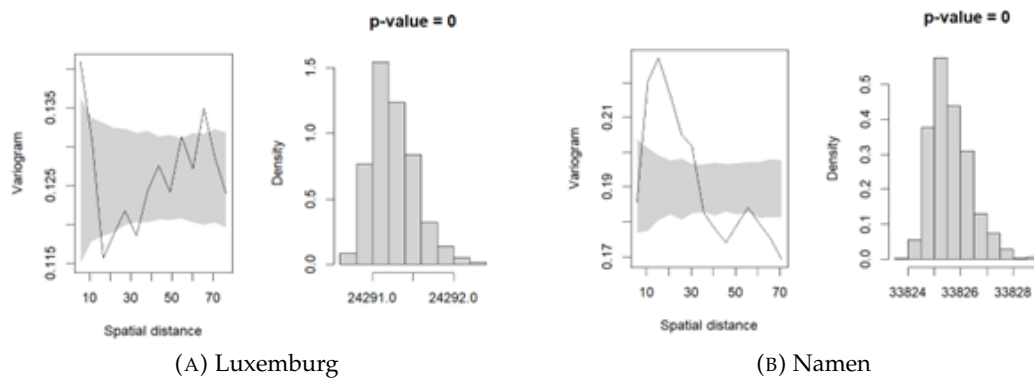


FIGURE A.10: Variograms for provinces Luxemburg and Namen

Appendix B

BYM model results

B.1 Order $k = 1$

B.1.1 Data on 4 km^2 scale

Limburg

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1572	0.1569	0.01487	0.1282	0.1855
ratio	0.1267	0.09623	0.09766	0.02401	0.3658
sigma.u	0.002694	0.002212	0.001735	0.0006064	0.007333
sigma.v	0.02305	0.02289	0.005669	0.01280	0.03482

Antwerp

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1197	0.1196	0.01192	0.09632	0.1431
ratio	0.1119	0.08905	0.08779	0.02084	0.3541
sigma.u	0.002539	0.002167	0.001762	0.0005053	0.006775
sigma.v	0.02390	0.02368	0.004258	0.01605	0.03303

Vlaams-Brabant

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.09768	0.09758	0.01093	0.07712	0.1193
ratio	1.4343	0.9695	1.6303	0.2384	5.5459
sigma.u	0.01101	0.008873	0.007546	0.002726	0.03327
sigma.v	0.009947	0.009139	0.002767	0.004805	0.01556

Oost-Vlaanderen

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1058	0.1056	0.01219	0.08300	0.1293
ratio	0.1906	0.1472	0.1550	0.02035	0.5922
sigma.u	0.004511	0.003508	0.003503	0.0004978	0.01346
sigma.v	0.0247	0.02454	0.004233	0.01709	0.03369

West-Vlaanderen

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1055	0.1055	0.01175	0.08285	0.1284
ratio	20.4312	2.4599	54.9936	0.1685	226.3986
sigma.u	0.2713	0.06477	0.5117	0.004304	2.0706
sigma.v	0.02391	0.02430	0.005801	0.009415	0.03410

B.1.2 Data on 16 km^2 scale**Limburg**

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1697	0.1698	0.01385	0.1423	0.1968
ratio	3.2128	1.0913	9.2314	0.09464	15.8378
sigma.u	0.006169	0.004125	0.006632	0.0005898	0.02579
sigma.v	0.004214	0.003802	0.002536	0.0006427	0.01009

Antwerpen

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1316	0.1317	0.01251	0.1058	0.1551
ratio	1.2461	0.6028	2.1469	0.07052	6.2824
sigma.u	0.006335	0.004742	0.005294	0.0006517	0.02093
sigma.v	0.007779	0.007632	0.002897	0.002451	0.01395

Vlaams-Brabant

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1024	0.1027	0.01189	0.0787	0.1250
ratio	1.7813	0.8459	3.01278	0.1299	10.2950
sigma.u	0.007049	0.004877	0.006503	0.0009218	0.02477
sigma.v	0.006050	0.005801	0.002532	0.001700	0.01179

Oost-Vlaanderen

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1158	0.1160	0.01231	0.09215	0.1399
ratio	0.5365	0.3105	0.7479	0.04443	2.4821
sigma.u	0.004726	0.003039	0.005277	0.0004886	0.01874
sigma.v	0.01013	0.009853	0.002703	0.005663	0.01609

West-Vlaanderen

	Mean	Median	Std.Dev.	95% CI L	95% CI U
alpha	0.1436	0.1434	0.01370	0.1165	0.1705
ratio	0.6042	0.3471	0.8169	0.03295	2.7771
sigma.u	0.005376	0.003817	0.005216	0.0003992	0.01985
sigma.v	0.01098	0.01071	0.003478	0.004864	0.01867

B.2 Order $k = 2, 3, 4, 5$

B.2.1 Data on 4 km^2 scale

Limburg

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1594	0.1591	0.01428	0.1301	0.1870
	ratio	0.3721	0.2434	0.3873	0.02737	1.5064
	sigma.u	0.007901	0.005543	0.007790	0.0006907	0.03160
	sigma.v	0.02263	0.02238	0.005401	0.01306	0.03414
k = 3	alpha	0.1593	0.1592	0.01445	0.1303	0.1878
	ratio	1.6826	0.8361	2.2245	0.03681	7.9006
	sigma.u	0.03210	0.01852	0.03481	0.0008412	0.1296
	sigma.v	0.02233	0.02188	0.005548	0.01246	0.03389
k = 4	alpha	0.1583	0.1585	0.01435	0.1298	0.1863
	ratio	1.1559	0.4281	1.4902	0.01544	5.1798
	sigma.u	0.02361	0.008836	0.02866	0.0004001	0.09606
	sigma.v	0.02247	0.02198	0.005785	0.01261	0.03519
k = 5	alpha	0.1586	0.1588	0.01502	0.1282	0.1878
	ratio	3.2727	0.6481	10.2781	0.03925	32.6574
	sigma.u	0.05043	0.01457	0.1148	0.0007435	0.4815
	sigma.v	0.02215	0.02193	0.005614	0.01150	0.03416

Antwerpen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1215	0.1216	0.01233	0.09648	0.1449
	ratio	1.9992	1.4709	1.8928	0.2063	7.3522
	sigma.u	0.03943	0.03242	0.03003	0.005181	0.1248
	sigma.v	0.02213	0.02184	0.004536	0.01302	0.03156
k = 3	alpha	0.1186	0.1187	0.01172	0.09523	0.1409
	ratio	3.3136	0.9330	6.8945	0.06259	24.5147
	sigma.u	0.05398	0.02287	0.07577	0.001454	0.2975
	sigma.v	0.02303	0.02320	0.004960	0.01172	0.03220
k = 4	alpha	0.1201	0.1200	0.01167	0.09776	0.1431
	ratio	24.9233	12.6309	44.9867	1.1683	135.0562
	sigma.u	0.2950	0.2451	0.2410	0.03123	0.9236
	sigma.v	0.01903	0.01924	0.006085	0.005822	0.03045
k = 5	alpha	0.1198	0.1202	0.001104	0.09753	0.1409
	ratio	255.2283	29.5023	635.7182	4.9738	2376.154
	sigma.u	0.7083	0.5029	0.5317	0.1217	1.9509
	sigma.v	0.01520	0.01670	0.0007809	0.0006971	0.02838

Vlaams-Brabant

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.09602	0.09603	0.01057	0.07538	0.1166
	ratio	5.4530	2.5193	7.7633	0.3586	25.8727
	sigma.u	0.03761	0.02282	0.03649	0.004131	0.1321
	sigma.v	0.009201	0.008980	0.002847	0.004213	0.01552
k = 3	alpha	0.09665	0.09698	0.01027	0.07593	0.1170
	ratio	7.1115	1.9458	12.1335	0.1137	45.9269
	sigma.u	0.05141	0.02014	0.06852	0.001379	0.2423
	sigma.v	0.009967	0.009765	0.003034	0.004713	0.01679
k = 4	alpha	0.09669	0.09653	0.01013	0.07646	0.1162
	ratio	48.3725	11.8445	109.2030	0.5191	328.4168
	sigma.u	0.1795	0.1071	0.1868	0.006243	0.6712
	sigma.v	0.009394	0.008472	0.003518	0.001731	0.01505
k = 5	alpha	0.09649	0.09658	0.01025	0.07643	0.1164
	ratio	153.1125	45.3672	243.2441	3.5838	863.6119
	sigma.u	0.4286	0.3282	0.3073	0.03780	1.1450
	sigma.v	0.006997	0.007043	0.003928	0.001066	0.01485

Oost-Vlaanderen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1048	0.1045	0.01143	0.08267	0.1277
	ratio	0.6957	0.2895	0.9082	0.04748	3.7431
	sigma.u	0.01611	0.007036	0.01971	0.001229	0.08091
	sigma.v	0.02468	0.02442	0.004183	0.01712	0.03371
k = 3	alpha	0.1053	0.1054	0.01154	0.08262	0.1283
	ratio	2.7575	0.4488	5.3327	0.05026	19.1110
	sigma.u	0.05400	0.01115	0.08887	0.01126	0.3193
	sigma.v	0.02463	0.02451	0.004911	0.01511	0.03465
k = 4	alpha	0.1049	0.1048	0.01146	0.08301	0.1282
	ratio	35.6770	15.1297	65.8215	0.7917	273.7241
	sigma.u	0.3823	0.2974	0.3288	0.02165	1.2475
	sigma.v	0.01936	0.01967	0.006771	0.004478	0.03123
k = 5	alpha	0.1059	0.1056	0.01153	0.08341	0.1287
	ratio	150.2129	6.6703	372.3302	0.2207	1294.474
	sigma.u	0.6085	0.1622	0.7547	0.006244	2.3115
	sigma.v	0.01887	0.02130	0.009024	0.001495	0.03266

West-Vlaanderen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1424	0.1426	0.01217	0.1177	0.1656
	ratio	3.3432	2.1085	3.3958	0.5011	12.3533
	sigma.u	0.04880	0.03502	0.03963	0.01021	0.1472
	sigma.v	0.01707	0.01673	0.004422	0.009479	0.02684
k = 3	alpha	0.1408	0.1409	0.01227	0.1164	0.1640
	ratio	37.3722	8.3787	85.5043	0.7517	242.5648
	sigma.u	0.1879	0.1223	0.1691	0.01538	0.5791
	sigma.v	0.01401	0.01447	0.006944	0.001911	0.02719
k = 4	alpha	0.1412	0.1414	0.01914	0.1173	0.1637
	ratio	60.1374	20.7528	148.7379	0.8561	343.7418
	sigma.u	0.3451	0.2817	0.2611	0.01773	0.9667
	sigma.v	0.01354	0.01344	0.006152	0.002311	0.02607
k = 5	alpha	0.1414	0.1415	0.01188	0.1176	0.1641
	ratio	179.1779	21.7679	349.3416	0.7021	1260.631
	sigma.u	0.4692	0.2971	0.4250	0.01412	1.4361
	sigma.v	0.01315	0.01392	0.008026	0.0009242	0.02787

B.2.2 Data on 16 km^2 scale

Limburg

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1693	0.1694	0.01366	0.1415	0.1963
	ratio	12.8281	3.9367	32.1796	0.2709	80.3891
	sigma.u	0.02197	0.01510	0.02067	0.001497	0.08135
	sigma.v	0.004098	0.003605	0.002510	0.0005422	0.01009
k = 3	alpha	0.1686	0.1682	0.01322	0.1429	0.1046
	ratio	16.7240	5.8591	33.0877	0.4134	104.5138
	sigma.u	0.02706	0.01854	0.02630	0.001755	0.09914
	sigma.v	0.003751	0.003189	0.002601	0.0004515	0.009812
k = 4	alpha	0.1691	0.1692	0.01354	0.1413	0.1957
	ratio	23.3265	9.6149	36.6832	0.2867	128.1999
	sigma.u	0.03779	0.03113	0.02997	0.001978	0.1134
	sigma.v	0.003696	0.003115	0.002553	0.0004726	0.009913
k = 5	alpha	0.1688	0.1694	0.01378	0.1411	0.1954
	ratio	14.1081	2.1253	41.1302	0.1476	120.5943
	sigma.u	0.01948	0.008301	0.02722	0.0007756	0.09197
	sigma.v	0.004183	0.003685	0.002675	0.0005091	0.01053

Antwerpen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1304	0.1307	0.01183	0.1067	0.1540
	ratio	10.9513	3.1400	33.5099	0.1796	71.1992
	sigma.u	0.03390	0.02363	0.03179	0.001934	0.1154
	sigma.v	0.007323	0.007266	0.003003	0.001400	0.01350
k = 3	alpha	0.1290	0.1291	0.01158	0.1062	0.1514
	ratio	6.1186	1.7782	18.2582	0.05985	41.7410
	sigma.u	0.02339	0.01457	0.02724	0.0004992	0.1023
	sigma.v	0.007960	0.007857	0.002963	0.001907	0.01394
k = 4	alpha	0.1313	0.1314	0.01137	0.1093	0.1533
	ratio	23.8027	9.1658	41.0019	0.4081	146.8673
	sigma.u	0.05873	0.04985	0.04529	0.003743	0.1659
	sigma.v	0.005788	0.005526	0.003232	0.0008086	0.01273
k = 5	alpha	0.1313	0.1314	0.01993	0.1077	0.1546
	ratio	10.2381	4.3548	16.2564	0.3732	58.9341
	sigma.u	0.04078	0.03165	0.03442	0.003512	0.1256
	sigma.v	0.006957	0.006834	0.003028	0.001656	0.01349

Vlaams-Brabant

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1025	0.1027	0.01089	0.08041	0.1228
	ratio	22.4353	10.0829	33.5021	0.9758	105.8101
	sigma.u	0.05297	0.04449	0.03660	0.007702	0.1409
	sigma.v	0.004588	0.004146	0.002642	0.0009168	0.01106
k = 3	alpha	0.1027	0.1027	0.01055	0.08150	0.1237
	ratio	33.8948	16.4016	50.9297	0.7497	176.2250
	sigma.u	0.06096	0.05635	0.03943	0.005340	0.1477
	sigma.v	0.003979	0.003522	0.002520	0.0005661	0.01004
k = 4	alpha	0.1019	0.01018	0.01105	0.08003	0.1236
	ratio	20.4884	9.0563	33.0357	0.1466	118.0613
	sigma.u	0.04615	0.04188	0.03628	0.001082	0.1289
	sigma.v	0.004935	0.004625	0.002781	0.0007652	0.01107
k = 5	alpha	0.1018	0.1022	0.01107	0.07919	0.1227
	ratio	19.1028	6.7180	30.3781	0.3192	109.1905
	sigma.u	0.04436	0.03365	0.03837	0.002392	0.1341
	sigma.v	0.004989	0.004676	0.002710	0.0008633	0.01100

Oost-Vlaanderen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1168	0.1168	0.01251	0.09296	0.1416
	ratio	3.0102	1.4506	5.5917	0.1417	16.9487
	sigma.u	0.02250	0.01435	0.02433	0.001592	0.09624
	sigma.v	0.009929	0.009743	0.002917	0.004723	0.01632
k = 3	alpha	0.1175	0.1174	0.01210	0.09436	0.1409
	ratio	1.4631	0.8721	1.7888	0.1496	6.2545
	sigma.u	0.01310	0.008515	0.01335	0.001661	0.05043
	sigma.v	0.01014	0.009891	0.002695	0.005495	0.01604
k = 4	alpha	0.1173	0.1172	0.01221	0.09304	0.1403
	ratio	3.7896	1.2807	8.5602	0.07631	27.3985
	sigma.u	0.02365	0.01254	0.03031	0.0009039	0.1158
	sigma.v	0.009819	0.009709	0.003131	0.003684	0.01617
k = 5	alpha	0.1167	0.1165	0.01253	0.09242	0.1410
	ratio	2.2158	0.8287	6.8189	0.1106	11.9155
	sigma.u	0.01534	0.008405	0.02151	0.001339	0.07709
	sigma.v	0.01016	0.009972	0.002892	0.004880	0.01633

West-Vlaanderen

		Mean	Median	Std.Dev.	95% CI L	95% CI U
k = 2	alpha	0.1432	0.1434	0.01316	0.1158	0.1677
	ratio	10.2162	3.2949	18.9943	0.1812	62.6662
	sigma.u	0.04706	0.03347	0.04910	0.04910	0.1680
	sigma.v	0.009152	0.009090	0.003997	0.003997	0.01735
k = 3	alpha	0.1439	0.1438	0.01303	0.1182	0.1685
	ratio	8.8221	2.7801	19.5270	0.2345	65.3709
	sigma.u	0.04088	0.02744	0.04161	0.002897	0.1594
	sigma.v	0.009410	0.009228	0.004011	0.002072	0.01786
k = 4	alpha	0.1443	0.1448	0.01290	0.1176	0.1693
	ratio	25.3886	8.5963	43.1791	0.6420	147.9284
	sigma.u	0.08001	0.06196	0.06028	0.007492	0.2191
	sigma.v	0.007489	0.007218	0.004206	0.001175	0.01653
k = 5	alpha	0.1435	0.1434	0.01322	0.1175	0.1706
	ratio	21.2589	1.4842	63.1187	0.03552	203.0983
	sigma.u	0.04538	0.01537	0.05912	0.0003811	0.2088
	sigma.v	0.009272	0.009403	0.004523	0.0007921	0.01802

Appendix C

R Code

```

# Thesis
library("sp")
library("tmap")
library("sf")
library("readxl")
library("tidyverse")
library("dplyr")
library("raster")
library("rgdal")
library("rgeos")
library("spdep")
library("coda")
library("nimble")
library("ggmcmc")
library("INLA")
library("writexl")
library("knitr")
library("argosfilter")
library("magrittr")
library("ggplot2")
library("mapproj")
library("PrevMap")
library("spData")
library("surveillance")
library("corpcor")
library("MASS")
setwd("")

data<-readRDS("data_combined_x4.rds")
province<-readRDS('shape_province_label.rds')
province <- st_zm(province)
data$geometry <- st_transform(data$geometry, crs = 3857)
province$geometry <- st_transform(province$geometry, crs = 3857)
total <- st_join(data, province, join = st_nearest_feature, left = T)
Total <- st_sf(total)
tm_shape(TotalLB) + tm_polygons("dth2020", palette = "Blues",
breaks=c(0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100)) + tm_layout(legend.title.size = 1)
tm_shape(TotalLB) + tm_polygons("SMR2020", palette = "Reds",

```

```

breaks=c(0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 3))
+ tm_layout(legend.title.size = 1)

# Centroids of approximately 2000 subregions
coords <- as.data.frame(st_coordinates(st_centroid(st_geometry(Total))))
#Merge dataset with centroid coordinates
Total <- bind_cols(Total, coords)
Total$logE <-log(Total$E2020)
Total <- Total[Total$logE >= 0, ]
Total$geometry <- st_transform(Total$geometry, crs = 3857)
Total$X <- Total$X / 1000
Total$Y <- Total$Y / 1000
TotalLB <- Total[Total$province=="Provincie Limburg",]
TotalA <- Total[Total$province=="Provincie Antwerpen",]
TotalVB <- Total[Total$province=="Provincie Vlaams-Brabant",]
TotalOV <- Total[Total$province=="Provincie Oost-Vlaanderen",]
TotalWV <- Total[Total$province=="Provincie West-Vlaanderen",]

### Spatial autocorrelation
## Define neighborhood
nb <- poly2nb(TotalLB, queen=TRUE)
add_nb <- function(x){
  queen_nb <- poly2nb(x, queen = TRUE)
  count = card(queen_nb)
  if(!any(count==0)){
    return(queen_nb)
  }
  ## get nearest neighbour index, use centroids:
  nnbs = knearneigh(st_coordinates(st_centroid(x)))$nn
  no_edges_from = which(count==0)
  for(i in no_edges_from){
    queen_nb[[i]] = nnbs[i]
    queen_nb[[nnbs[i]]] = c(queen_nb[[nnbs[i]]],i)
  }
  return(queen_nb)
}
nb.new <- add_nb(TotalLB)
plot(st_geometry(TotalLB))
plot(nb.new,st_coordinates(st_centroid(st_geometry(TotalLB))),add=TRUE, col="red")
plot(nb,st_coordinates(st_centroid(st_geometry(TotalLB))),add=TRUE)
col.W <- nb2listw(nb, style="W", zero.policy = TRUE)

### BYM model
CONVmodel <- nimbleCode({
  for (i in 1 :N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + u[i] + v[i]
    RR[i] <- exp(alpha + u[i] + v[i])
    v[i] ~ dnorm(0,tau.v)
  }
})

```

```

}
# CAR prior distribution for random effects:
u[1:N] ~ car.normal(adj[1:L], weights[1:L], num[1:N], tau.u, zero_mean=1)
for(k in 1:L) {
  weights[k] <- 1
}
# Other priors:
alpha ~ dnorm(0,tau.alpha)
mean <- exp(alpha)
tau.u ~ dgamma(0.001, 0.001)
tau.v ~ dgamma(0.001, 0.001)
tau.alpha ~ dgamma(0.001, 0.001)
sigma.u<-1/tau.u
sigma.v<-1/tau.v
ratio<-sigma.u/sigma.v
for (i in 1:N){
  ## 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])
  ## Residuals
  res[i] <- (O[i] - mu[i])/sqrt(mu[i])
  prexR1[i] <- step(res[i]-1)
  prexR2[i] <- step(res[i]-2)
}
## Overall MSPE and MAPE
MSPE <- mean(SPE[1:N])
MAPE <- mean(APE[1:N])
})
##### Prepare data:
N<-dim(TotalLB)[1]
my.data <- list(O=TotalLB$dth2020, E= TotalLB$E2020)
my.constants <- list(N=N,adj=unlist(nb.new), L=sum(card(nb.new)),
num=card(nb.new))
##### Initial parameters
my.inits <- list(
  list(u = rep(0,N),v = rep(0,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1),
  list(u = rep(-0.5,N),v = rep(0.5,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1)
)
##### Specify parameters to monitor
parameters <- c("RR","alpha","sigma.u","sigma.v","u","v","ratio","MSPE","MAPE")
##### Running NIMBLE:
model.sim <- nimbleMCMC(code = CONVmodel,
                        data = my.data,
                        constants = my.constants,
                        inits = my.inits,
                        monitors = parameters,
                        niter = 10000,

```

```

        nburnin = 5000,
        nchains = 2,
        thin = 10,
        summary = TRUE,
        samplesAsCodaMCMC = TRUE,
        WAIC = TRUE)

#### Convert into mcmc.list
model.mcmc <- as.mcmc.list(model.sim$samples)
#### Produce general summary of obtained MCMC sampling
model.sim$summary$all.chains
summary(model.sim)
# WAIC
model.sim$WAIC

TotalLB$theta<-model.sim$summary$all.chains[startsWith(rownames
(model.sim$summary$all.chains),"RR"),"Mean"]
TotalLB$u<-model.sim$summary$all.chains[startsWith(rownames
(model.sim$summary$all.chains),"u"),"Mean"]
TotalLB$v<-model.sim$summary$all.chains[startsWith(rownames
(model.sim$summary$all.chains),"v"),"Mean"]
dev.off()
tm_shape(TotalLB) +
  tm_polygons(c("SMR"), palette = "-RdYlGn",
              breaks=c(0,0.5,0.6,0.8,1.2,1.5,2,2.5, 3, 3.5),
              contrast=c(0,1)) +
  tm_layout(legend.title.size = 1)
tm_shape(TotalLB) +
  tm_polygons(c("theta"), palette = "-RdYlGn",
              breaks=c(0, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 2),
              contrast=c(0,1)) +
  tm_layout(legend.title.size = 1)
tm_shape(TotalLB) +
  tm_polygons(c("u"), palette = "Reds") +
  tm_layout(legend.title.size = 1)
tm_shape(TotalLB) +
  tm_polygons(c("v"), palette = "Reds") +
  tm_layout(legend.title.size = 1)
RR.CONV<-TotalLB$theta
# Exceedance probability
TotalLB$ExProb <- sapply(paste0("RR[", 1:N, "]"),
                        function(X)
                          mean(c(model.sim$samples$chain1
                                [,X],model.sim$samples$chain2[,X])>1))
tm_shape(TotalLB) +
  tm_polygons(c("ExProb"), palette = "Blues",
              breaks=c(0.6, 0.65, 0.7, 0.75, 0.8, 0.85,
                      0.9, 0.95, 1)) +
  tm_layout(legend.title.size = 1)

```

```

### Continuous modelling spatial process with Matern correlation function ###
# Limburg
TotalLB <- as.data.frame(TotalLB)
check.spat <- spat.corr.diagnostic(dth2020~1,
                                   coords = ~I(X)+I(Y),
                                   data=TotalA,likelihood = "Poisson",
                                   units.m=~logE,
                                   uvec=NULL,n.sim=1000)

#geostatistical model
fit.LA <- glm.LA(dth2020~1,
                 coords=~I(X)+I(Y),kappa=0.5,
                 start.cov.pars = c(10, 5), units.m=~logE,
                 fixed.rel.nugget = NULL,
                 data=TotalA,family="Poisson")

summary(fit.LA)
par0 <- coef(fit.LA)
c.mcmc <- control.mcmc.MCML(n.sim=20000,burnin=1000,thin=8)
fit.MCML <- poisson.log.MCML(dth2020~1,control.mcmc = c.mcmc,
                             par0=par0, units.m=~logE,
                             coords=~I(X)+I(Y), kappa=0.5,
                             start.cov.pars = c(8.45, 5) ,fixed.rel.nugget = NULL,
                             data=TotalLB, method="nlminb")

summary(fit.MCML)
variog.diag.MC.cov <- variog.diagnostic.glm(fit.MCML,n.sim=1000)

## order k = 2
##### kth order distance based neighbourhood structure #####
TotalLB <-st_as_sf(TotalLB)
coords <- as.data.frame(st_coordinates(st_centroid(st_geometry(TotalLB))))
adj<-poly2nb(TotalLB,queen = T)
nbmat <- nblag(adj, maxlag = 2)
nbmat_cumul<-nblag_cumul( nbmat )
plot(st_geometry(TotalLB))
plot(nbmat_cumul, coords=coords, add=TRUE, col="red")
summary(nbmat_cumul, coords)
col.lags <- nblag(nbmat_cumul, 2)
print(col.lags)
summary(col.lags[[2]], coords)
plot(st_geometry(TotalLB3$geometry), border="grey")
plot(nbmat_cumul, coords, add=TRUE)
title(main="k =2")
plot(col.lags[[2]], coords, add=TRUE, col="red", lty=2)
# BYM model using new neighbourhood structure
CONVmodel <- nimbleCode({
  for (i in 1 :N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + u[i] + v[i]
  }
})

```

```

RR[i] <- exp(alpha + u[i] + v[i])
v[i] ~ dnorm(0,tau.v)
}
# CAR prior distribution for random effects:
u[1:N] ~ car.normal(adj[1:L], weights[1:L], num[1:N], tau.u, zero_mean=1)
for(k in 1:L) {
  weights[k] <- 1
}
# Other priors:
alpha ~ dnorm(0,tau.alpha)
mean <- exp(alpha)
tau.u ~ dgamma(0.001, 0.001)
tau.v ~ dgamma(0.001, 0.001)
tau.alpha ~ dgamma(0.001, 0.001)
sigma.u<-1/tau.u
sigma.v<-1/tau.v
ratio<-sigma.u/sigma.v
for (i in 1:N){
  ## 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])
  ## Residuals
  res[i] <- (O[i] - mu[i])/sqrt(mu[i])
  prexR1[i] <- step(res[i]-1)
  prexR2[i] <- step(res[i]-2)
}
## Overall MSPE and MAPE
MSPE <- mean(SPE[1:N])
MAPE <- mean(APE[1:N])
})
#### Prepare data:
N<-dim(TotalLB)[1]
my.data <- list(O=TotalLB$dth2020, E= TotalLB$E2020)
my.constants <- list(N=N,adj=unlist(nbmat_cumul),
L=sum(card(nbmat_cumul)), num=card(nbmat_cumul))
#### Initial parameters
my.inits <- list(
  list(u = rep(0,N),v = rep(0,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1),
  list(u = rep(-0.5,N),v = rep(0.5,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1)
)
#### Specify parameters to monitor
parameters <- c("RR","alpha","sigma.u","sigma.v","u","v","ratio","MSPE","MAPE")
#### Running NIMBLE:
model.sim <- nimbleMCMC(code = CONVmodel,
                        data = my.data,
                        constants = my.constants,
                        inits = my.inits,

```

```

        monitors = parameters,
        niter = 10000,
        nburnin = 5000,
        nchains = 2,
        thin = 10,
        summary = TRUE,
        samplesAsCodaMCMC = TRUE,
        WAIC = TRUE)

#### Convert into mcmc.list
model.mcmc <- as.mcmc.list(model.sim$samples)
#### Produce general summary of obtained MCMC sampling
plot(model.mcmc)
model.sim$summary$all.chains[c(152:155,1:2),]
summary(model.sim)
# WAIC
model.sim$WAIC

### Simulations ###
##read in data
data<-readRDS("data_combined_x4.rds")
#create neighbourhood graph
start <- Sys.time()
newWmat0 <- nbOrder(poly2adjmat(data,queen=T), maxlag = 3)
end <- Sys.time()
#correlation matrix
newWmat<-1/sqrt(1+newWmat0)
#re-define correlation matrix
newWmat[newWmat==1]<-0 #truncate correlation to 0 for neighbors >9th order
newWmat[row(newWmat)==col(newWmat)]<-1 #set diagonal elements
alpha<-0.15 #intercept
#simulate spatial random effects (u)
sigma.u2<-0.5 #original
p.dist<-newWmat #correlation matrix
cov.dist<-sigma.u2*p.dist #covariance matrix
cov.dist <- make.positive.definite(cov.dist, tol=1e-6)
#Multinomial count
n<-20
for (i in 1:n){
  set.seed(123+i)
  u<-rep(NA,nrow(data))
  v<-rep(NA,nrow(data))
  u<-mvrnorm(1,rep(0, nrow(data)),cov.dist) #spatial random effects
  v<-rnorm(nrow(data), mean=0, sd=0.55)#simulate unstructured random effects (v)
  ####simulate counts
  sim <- list(coords=data$geometry, units.m = data$E2020)
  r0<-exp(alpha+u+v) #true RR#unstructured random effects
  simdf <- rmultinom(1,sum(data$dth2020), prob = data$E2020*exp(alpha+u+v)) #Multinomial
  simLB<-data #get simulated sf object
}

```



```

simLB$dthsim<-simdf # add simulated case counts
simLB$RR0<-r0 #add true RRs
saveRDS(simLB,paste0("data_mult_bym_", i, '.rds'))
}

# analysis of simulated dataset
data<-readRDS("data_mult_bym_2.rds")
province<-readRDS('shape_province_label.rds')
province <- st_zm(province)
data$geometry <- st_transform(data$geometry, crs = 3857)
province$geometry <- st_transform(province$geometry, crs = 3857)
total <- st_join(data, province, join = st_nearest_feature, left = T)
Total <- st_sf(total)
Total$dthsim <- as.numeric(Total$dthsim)
rate <- Total %>%
  summarize(Y = sum(dthsim),N = sum(pop2020)) %>%
  mutate(rate=Y/N)
Total <- Total %>%
  mutate(Ei=rate$rate*pop2020)
Total <- Total %>%
  mutate(SMR=dthsim/Ei)
# Centroids of approximately 2000 subregions
coords <- as.data.frame(st_coordinates(st_centroid(st_geometry(Total))))
#Merge dataset with centroid coordinates
Total <- bind_cols(Total, coords)
Total$logE <-log(Total$Ei)
Total <- Total[Total$logE >= 0, ]
Total$X <- Total$X / 1000
Total$Y <- Total$Y / 1000
TotalLB <- Total[Total$province=="Provincie Limburg",]
TotalOV <- Total[Total$province=="Provincie Oost-Vlaanderen",]

tm_shape(TotalLB) + tm_polygons("dthsim", palette = "Blues",
breaks=c(0, 20, 40, 60, 80, 100)) +
tm_layout(legend.title.size = 0.9, legend.text.size = 0.65)
tm_shape(TotalLB) + tm_polygons("SMR", palette = "Reds",
breaks=c(0, 0.5, 1, 1.5, 2, 2.5, 3))
+ tm_layout(legend.title.size = 0.9, legend.text.size = 0.60,)

TotalOV <- as.data.frame(TotalOV)
check.spat <- spat.corr.diagnostic(dthsim~1,
                                coords = ~I(X)+I(Y),
                                data=TotalOV,likelihood = "Poisson",
                                units.m=~logE,
                                uvec=NULL,n.sim=1000)

#geostatistical model
fit.LA <- glm.LA(dthsim~1,
                coords=~I(X)+I(Y),kappa=0.5,
                start.cov.pars = c(3, 3), units.m=~logE, fixed.rel.nugget = NULL,

```

```

                                data=TotalOV,family="Poisson")
summary(fit.LA)
par0 <- coef(fit.LA)
c.mcmc <- control.mcmc.MCML(n.sim=20000,burnin=1000,thin=8)
fit.MCML <- poisson.log.MCML(dthsim~1,control.mcmc = c.mcmc,
                             par0=par0, units.m=~logE,
                             coords=~I(X)+I(Y), kappa=0.5,
                             start.cov.pars = c(3, 1.94) ,fixed.rel.nugget = NULL,
                             data=TotalOV, method="nlminb")

summary(fit.MCML)

### Spatial autocorrelation
## Define neighborhood
nb <- poly2nb(TotalOV, queen=TRUE)
add_nb <- function(x){
  queen_nb <- poly2nb(x, queen = TRUE)
  count = card(queen_nb)
  if(!any(count==0)){
    return(queen_nb)
  }
  ## get nearest neighbour index, use centroids:
  nnbs = knearneigh(st_coordinates(st_centroid(x)))$nn
  no_edges_from = which(count==0)
  for(i in no_edges_from){
    queen_nb[[i]] = nnbs[i]
    queen_nb[[nnbs[i]]] = c(queen_nb[[nnbs[i]]],i)
  }
  return(queen_nb)
}
nb.new <- add_nb(TotalOV)
plot(st_geometry(TotalOV))
plot(nb.new,st_coordinates(st_centroid(st_geometry(TotalOV))),add=TRUE, col="red")
plot(nb,st_coordinates(st_centroid(st_geometry(TotalOV))),add=TRUE)

### BYM model
CONVmodel <- nimbleCode({
  for (i in 1 :N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + u[i] + v[i]
    RR[i] <- exp(alpha + u[i] + v[i])
    v[i] ~ dnorm(0,tau.v)
  }
  # CAR prior distribution for random effects:
  u[1:N] ~ car.normal(adj[1:L], weights[1:L], num[1:N], tau.u, zero_mean=1)
  for(k in 1:L) {
    weights[k] <- 1
  }
  # Other priors:
  alpha ~ dnorm(0,tau.alpha)
})

```

```

mean <- exp(alpha)
tau.u ~ dgamma(0.001, 0.001)
tau.v ~ dgamma(0.001, 0.001)
tau.alpha ~ dgamma(0.001, 0.001)
sigma.u<-1/tau.u
sigma.v<-1/tau.v
ratio<-sigma.u/sigma.v
for (i in 1:N){
  ## 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])
  ## Residuals
  res[i] <- (O[i] - mu[i])/sqrt(mu[i])
  prexR1[i] <- step(res[i]-1)
  prexR2[i] <- step(res[i]-2)
}
## Overall MSPE and MAPE
MSPE <- mean(SPE[1:N])
MAPE <- mean(APE[1:N])
})
##### Prepare data:
N<-dim(TotalOV)[1]
my.data <- list(O=TotalOV$dthsim, E= TotalOV$Ei)
my.constants <- list(N=N,adj=unlist(nb.new), L=sum(card(nb.new)), num=card(nb.new))
##### Initial parameters
my.inits <- list(
  list(u = rep(0,N),v = rep(0,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1),
  list(u = rep(-0.5,N),v = rep(0.5,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1)
)
##### Specify parameters to monitor
parameters <- c("RR","alpha","sigma.u","sigma.v","u","v","ratio","MSPE","MAPE")
##### Running NIMBLE:
model.sim <- nimbleMCMC(code = CONVmodel,
  data = my.data,
  constants = my.constants,
  inits = my.inits,
  monitors = parameters,
  niter = 10000,
  nburnin = 5000,
  nchains = 2,
  thin = 10,
  summary = TRUE,
  samplesAsCodaMCMC = TRUE,
  WAIC = TRUE)
##### Convert into mcmc.list
model.mcmc <- as.mcmc.list(model.sim$samples)
##### Produce general summary of obtained MCMC sampling

```

```

model.sim$summary$all.chains[c(646:670),]
summary(model.sim)
# WAIC
model.sim$WAIC

##### kth order distance based neighbourhood structure #####
TotalOV <-st_as_sf(TotalOV)
coords <- as.data.frame(st_coordinates(st_centroid(st_geometry(TotalOV))))
adj<-poly2nb(TotalOV,queen = T)
nbmat <- nblag(adj, maxlag = 10)
nbmat_cumul<-nblag_cumul( nbmat )
summary(nbmat_cumul, coords)

# BYM model using new neighbourhood structure
CONVmodel <- nimbleCode({
  for (i in 1 :N) {
    O[i] ~ dpois(mu[i])
    log(mu[i]) <- log(E[i]) + alpha + u[i] + v[i]
    RR[i] <- exp(alpha + u[i] + v[i])
    v[i] ~ dnorm(0,tau.v)
  }
  # CAR prior distribution for random effects:
  u[1:N] ~ car.normal(adj[1:L], weights[1:L], num[1:N], tau.u, zero_mean=1)
  for(k in 1:L) {
    weights[k] <- 1
  }
  # Other priors:
  alpha ~ dnorm(0,tau.alpha)
  mean <- exp(alpha)
  tau.u ~ dgamma(0.001, 0.001)
  tau.v ~ dgamma(0.001, 0.001)
  tau.alpha ~ dgamma(0.001, 0.001)
  sigma.u<-1/tau.u
  sigma.v<-1/tau.v
  ratio<-sigma.u/sigma.v
  for (i in 1:N){
    ## 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])
    ## Residuals
    res[i] <- (O[i] - mu[i])/sqrt(mu[i])
    prexR1[i] <- step(res[i]-1)
    prexR2[i] <- step(res[i]-2)
  }
  ## Overall MSPE and MAPE
  MSPE <- mean(SPE[1:N])
  MAPE <- mean(APE[1:N])

```

```
}  
##### Prepare data:  
N<-dim(TotalOV)[1]  
my.data <- list(O=TotalOV$dth2020, E= TotalOV$E2020)  
my.constants <- list(N=N,adj=unlist(nbmat_cumul),  
L=sum(card(nbmat_cumul)), num=card(nbmat_cumul))  
##### Initial parameters  
my.inits <- list(  
  list(u = rep(0,N),v = rep(0,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1),  
  list(u = rep(-0.5,N),v = rep(0.5,N),alpha=0, tau.u=1, tau.v=1, tau.alpha=1)  
)  
##### Specify parameters to monitor  
parameters <- c("RR","alpha","sigma.u","sigma.v","u","v","ratio","MSPE","MAPE")  
##### Running NIMBLE:  
model.sim <- nimbleMCMC(code = CONVmodel,  
  data = my.data,  
  constants = my.constants,  
  inits = my.inits,  
  monitors = parameters,  
  niter = 10000,  
  nburnin = 5000,  
  nchains = 2,  
  thin = 10,  
  summary = TRUE,  
  samplesAsCodaMCMC = TRUE,  
  WAIC = TRUE)  
##### Convert into mcmc.list  
model.mcmc <- as.mcmc.list(model.sim$samples)  
##### Produce general summary of obtained MCMC sampling  
model.sim$summary$all.chains[c(661:671),]  
# WAIC  
model.sim$WAIC
```