Contents

Symbols

$\pi(\cdot)$	Density of its arguments
σ	Standard deviation
Var	Variance
Cov	Covariance
Prec	Precision
Corr	Correlation
$\mathbb{E}$	Expected value
$\mathbb{P}$	Probability
$\int$	Integral of its arguments
$\sum$	Sum of its arguments
$\prod$	Product of its arguments
$\exp$	Exponential function
$\log$	Logarithmic function
∂	The derivative
$\propto$	Proportional to
$\mathbb{R}$	Real numbers
$\mathbb{N}$	Natural numbers
$\mathbb{N}_0$	Natural numbers including 0
$\boldsymbol{I}$	Identity matrix

1

1 Introduction

1.1 Background

Controlling or even trying to prevent an infectious disease is a challenging task, and therefore it is crucial to find ways to combat this type of disease through new and creative ways. The number of infections can vary greatly between different countries or regions, and it is therefore of great interest for local governments or health institutes to find the underlying factors for these differences. This may lead to the identification of previously unrecognised environmental factors that could be the cause of the different risk of disease in different areas. One of the earliest examples of this type of analysis was carried out in relation to a cholera outbreak in south London in 1854 by John Snow. By creating the map shown in Figure 1.1, he was able to show that cholera cases occurred mainly around a water pump in Broad Street. These findings were crucial to understanding that cholera spread through contaminated water supplies, and thus led to the modernisation of water supply and sanitation systems in London and the rest of the world (Snow, 1857).

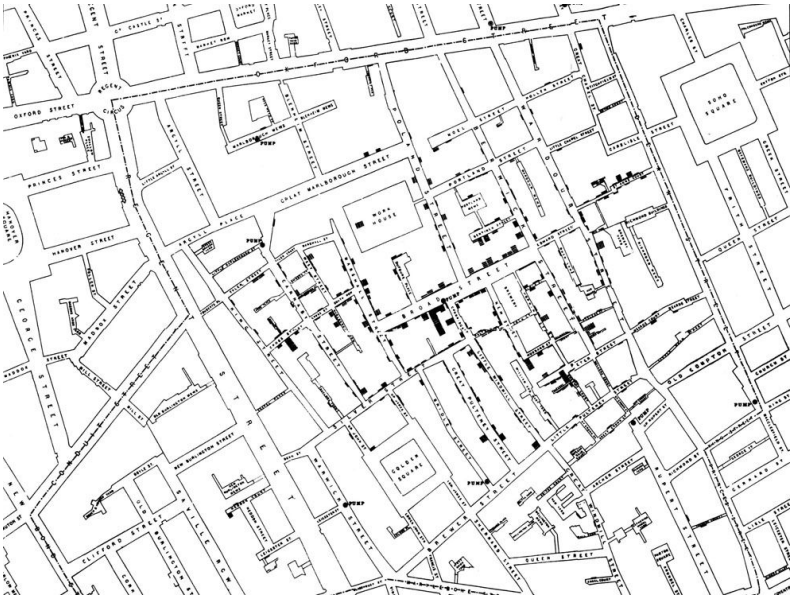

Fig. 1.1: The original map of cholera cases in southern London, created by John Snow in 1854.

2

The location itself (that is, a set of geographic coordinates) is generally unlikely to influence the risk of a certain disease; as there is no reason why one set of coordinates would inherently be at higher risk than another. Instead, the geographic location is a proxy measure, for differences in the attributes of the areas. These differences may relate to physical geography (e.g. temperature, sunlight, precipitation), environmental factors (e.g. air pollution, water quality) or population attributes(e.g. age, income, migration background). The identification of disparities in disease risk across a geographic region can lead to further investigation of the underlying reasons for the differences, which can lead to health breakthroughs such as those noted by Snow. Furthermore, by identifying areas of high risk, health authorities can focus additional resources on these areas in an attempt to influence the behaviours of the population that contribute to an increased risk of disease.

Most approaches to disease mapping are based on dividing the geographic region into spatial units, with disease risks estimated for each of these units. The reason for this is that individual-level data would violate patient confidentiality and governments are more interested in risk levels for the entire population. Each spatial unit has different demographics, so comparisons between spatial units are generally based on the standardised incidence ratio (SIR), defined as the number of observed cases in a given area divided by the number of cases expected for that area based on its population demographics. Most modern methodology for estimating disease risk relies on conditional autoregressive (CAR) models (Besag et al., 1991), which assume the existence of spatial autocorrelation between neighbouring areas, based on the notion that nearby areas are more likely to have more in common than areas that are further apart. This is due to the fact that adjacent areas are more likely to have similar socio-economic characteristics in terms of deprivation and population behaviour. It is assumed by these models that this level of spatial autocorrelation is constant across the spatial region.

1.2 Motivation

Covid-19 has had a significant impact on the lives of probably almost everyone on earth. Whether people had to work from home, children suddenly had online classes or people just stayed at home more, everyone was affected. Everyone had to get used to this new reality where suddenly you couldn't meet for a coffee or go to the cinema together because those establishments were either closed or people had no desire to risk contracting Covid-19. And that, of course, says nothing about the impact it had on the lives of people who became infected with Covid-19 or had relatives who became infected. The point is that everyone was affected by the impact of the pandemic, and still is, albeit at different levels.

Over time, different countries introduced different strategies to combat Covid-19, such as hard lockdowns where people were only allowed to leave the house if they had a legitimate reason to do so, for example to go to work or to buy groceries, while other countries did not introduce any lockdown measures. Other measures included wearing face masks in public places or limiting how many people can meet in public and private spaces. But even when the same measures were implemented in one country, there were big differences between the proportion of infected people in different parts of the country.

Finally, attitudes towards these measures have become a matter of political identity in various countries, as political parties from different spectrums have different views on how this pandemic should be handled, what measures should be implemented, or whether this pandemic even exists or if it is just much ado about nothing. This has also led to the formation of new political movements that regularly protest against the measures taken by the government and demand a return to pre-pandemic conditions.

Understanding the reason why the number of infections varies in different parts of the country can be crucial in helping local governments decide which measures to implement to limit the risk of infection and contribute to the common goal of ending the pandemic as soon as possible.

Identifying areas where people are at higher risk of becoming infected can also help governments decide on vaccination strategies, as it may make more sense to vaccinate people in high-risk areas first. By limiting the number of infections in these areas, the likelihood of the virus spreading from one of these areas to one or more neighbouring areas decreases, which in turn slows the spread of Coronavirus.

1.3 Aim and Objective

The main objective of this thesis is to analyse which factors drive the infection numbers of Covid-19 and thus increase the risk of people becoming infected and falling ill with the virus. To answer this question, a basic concept of Bayesian theory and Bayesian spatial models is developed and an introduction to geospatial data and the analysis of this specific type of data is given. For the analysis, different types of data need to be collected from different sources. The type of data includes

- Data related to the number of infections in a given municipality.
- Data related to the number of vaccinations in a given municipality.
- Demographic data related to a specific municipality.
- Data related to spatial points of interest in a given municipality.
- Shapefiles for the geographic areas of interest.

In the analysis, these data are collected for two countries, Germany and Norway. These two countries are not equally affected by the pandemic and the population is distributed differently in the two countries, which makes for an interesting comparison of what factors influence the infection numbers and what kind of models work well in each country. In order to achieve the main objective of this thesis, there are the following important goals:

1. Build a basic concept of Bayesian theory and analysis of geospatial health data.
2. Collect all the data needed for the analysis.
   - 2.1 Collect data related to Covid-19 from the National Institutes of Health.
   - 2.2 Collect demographic data and shapefiles from other official sources.
   - 2.3 Collect infrastructure data by querying OpenStreetMap.
3. Merge all data from different sources to create a clean dataset for analysis.
4. Develop and compare different types of Bayesian spatial models.
5. Critically evaluate the models.
6. Extract factors that significantly influence the risk of infection.

1.4 Corona Virus

Viral diseases continue to pose a serious public health threat. Several viral epidemics have occurred in the last 20 years, including the SARS pandemic in 2002/3, H1N1 influenza in 2009, and more recently the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), which was first detected in Saudi Arabia in 2012.
In late 2019, the first few cases of lower respiratory infections were detected in Wuhan, China. In February 2020, this viral disease was officially named "Covid-19", an acronym for "Coronavirus Disease 2019".
Due to the rapid spread of the virus, a Public Health Emergency of International Concern was declared at the end of January 2020, with 18 countries reporting cases and four countries reporting human-to-human transmission.
At the end of February 2020, the World Health Organisation (WHO) raised the risk of a Covid 19 epidemic to "very high" before declaring it a pandemic on 11 March. At that time, more than 118,000 cases in 114 countries and 4000 deaths had already been registered.
The first cases of the disease were linked to direct exposure at the Huanan Seafood Wholesale Market in Wuhan, with animal-to-human transmission suspected as the main mechanism. After subsequent cases could not be linked to this mechanism, human-to-human transmission was presumed to be the main transmission mechanism. Furthermore, sympotomatic individuals are thought to be the most common source of Covid-19 spread. However, asymptomatic individuals can also transmit the virus, therefore isolation is the best way to contain this epidemic.
Similar to other respiratory diseases, e.g. influenza, transmission is thought to occur through respiratory droplets (particles $> 5 - 10\mu m$ in diameter) when coughing and sneezing. In closed rooms, transmission by aerosol is also possible.
Based on the data from the first cases in Wuhan, the incubation period is generally between 3 and 7 days, with a median of 5.1 days. According to the data, the number of infections doubled about every seven days and the basic reproductive number R is 2.2, which means that on average each infected individual infects another 2.2 individuals.
According to a report by the Chinese Centre for Disease Control, which studied 72,314 cases, the overall mortality rate of confirmed cases was 2.3%, with most of the fatal cases affecting people over 70 years of age.
Furthermore, the clinical manifestations of the disease can be divided into three groups according to their severity:

- Mild disease: non-pneumonia and mild pneumonia; this occurred in 81% of cases.

- Severe disease: dyspnea, respiratory rate ≥ 30 min, blood oxygen level $\leq 93\%$; this occurred in 14% of cases.
- Critical disease: respiratory failure, septic shock and/or multiple organ dysfunction or failure; this occurred in 5% of cases.

Subsequent reports indicate that the disease is asymptomatic or with very mild symptoms in 70% of patients, while the remaining 30% develop a respiratory syndrome with high fever, cough and even severe respiratory failure, which may require admission to the intensive care unit.
Most countries use some kind of clinical and epidemiological information to determine who should be tested. A molecular test, for example a PCR test, can be used to detect the disease.
The WHO recommends the collection of samples from both the upper and lower respiratory tract. In the laboratory, the genetic material extracted from the saliva or mucus sample is amplified by reverse polymerase chain reaction (RT-PCR), which synthesises a double-stranded DNA molecule from an RNA form. Once the genetic material is sufficient, the parts of the genetic code of the CoV that are conserved are searched for. The probes used are based on the original gene sequence published by the Shanghai Public Health Clinical Center & School of Public Health, Fudan University, Shanghai, China on Virological.org and subsequent confirmatory evaluation by other laboratories (Cascella et al., 2021).

1.5 Related Work and Contribution

Since the start of the pandemic in 2019, numerous scientific papers have been written on Covid-19, covering a wide range of topics, such as medicine, social sciences and statistics. Naturally, there have also been papers written on the relationship between geographic regions and Covid-19. Incorporating a spatial dimension into the research process can help to better understand different phenomena and make them potentially mappable. The papers considered here can be divided into two categories. The first group consists of disease mapping, spatial analysis and spatio-temporal analysis and refers to studies that analyse the spatial and spatio-temporal patterns of Covid-19. The other group contains research that focuses on other factors that influence the dynamics of the Covid-19 pandemic, but may also include spatial and spatio-temporal analysis.

Disease Mapping, Spatial Analysis and Spatio-Temporal Analysis

Guan et al. (2020) studied cases in mainland China up to 25 February 2020 to determine the defining clinical features and severity of the disease. Among other things, they found that Covid-19 spread rapidly throughout the country and that the severity of the disease varied. Furthermore, they found that the most common symptoms experienced by patients were cough and fever. They reported a median incubation period of 4 days (Guan et al., 2020).

Chen et al. (2020) analysed how people who emigrated from Wuhan contributed to the early stages of the pandemic in China at the beginning of 2020. They found a strong correlation between the number of confirmed cases of Covid-19 in a given province and emigration from Wuhan. They also found that the lockdown of several cities in Hubei province and the implementation of nationwide control measures were effective in preventing the exponential growth of the number of cases (Z.-L. Chen et al., 2020).

Similar to this study, Gross et al. (2020) compare the infection rate in different cities in China and provinces in Italy during the early phases of the pandemic and conclude that the spread of the disease is defined by a two-stage process. The first stage, the authors say, is defined by a constant rate of infection due to a lack of means to detect infected individuals before symptoms appear. In the second stage, they observed an approximately exponential decline due to quarantine. While they found differences between China and Italy, most notably that it took longer for

outbreaks of the disease to be controlled in the Italian provinces, they found similar behaviour in terms of infection rate (Gross et al., 2020).

Chen et al. (2021) analysed the spatio-temporal distribution characteristics and influencing factors of the virus in mainland China using statistical methods, correlation analyses and geographic information system (GIS) mapping. They concluded that the outbreak in non-Hubei provinces can be divided into five phases. The initial outbreak phase, the peak phase where the highest number of new infections is observed, the containment phase where the number of new infections decreases, the rebound phase and a final phase where the number of new infections flattens out. They also observed that cities with large population flows from Wuhan were more affected by Covid-19 (Y. Chen et al., 2021).

Saha et al. (2020) provide an overview of how GIS, e.g. mapping dashboards and applications, can be used to monitor the pandemic and related activities. They conclude that the pandemic requires massive data generation and GIS to enable rapid response and analysis to help prevent and guide decisions and movements (Saha et al., 2020).

Gianquintieri et al. (2020) used geo-referenced calls to the emergency number relevant to respiratory problems and subsequent emergency medical service interventions to derive an unbiased representation of Covid-19 diffusion. This study was conducted for the Lombardy region of Italy, which was particularly hard hit by the pandemic in early 2020. The authors reported a strong correlation between Covid-19-related deaths at the provincial level and emergency calls and age- and sex-weighted ambulance dispatches (Gianquintieri et al., 2020).

Lastly, Petrov et al. (2020) examined the spatio-temporal dynamics of the pandemic in the Arctic up to July 2020. They found that the number of infections and morbidity were highly variable, but generally below national levels. They classified the Arctic regions into four groups: Iceland, the Faroe Islands, northern Norway and northern Finland, which were characterised by increased early infection rates but containment of the pandemic through quarantine and other measures; Northern Sweden and Alaska, where the first wave of infection persisted despite weak (Sweden) or variable (Alaska) quarantine measures; northern Russia, where a late start led to a steep rise in infections, deaths and several outbreaks; and northern Canada and Greenland where there was no significant spread of the pandemic (Petrov et al., 2020).

Other Factors Influencing the Pandemic

Xiong et al. (2020) carried out a correlation analysis for the number of cases in the Hubei province between 30 January 2020 and 18 February 2020. They found a

significant correlation between population, regional GDP , retail sales of consumer goods and the number of confirmed cases of Covid-19 in Hubei province, among others (Xiong et al., 2020).

Ahmadi et al. (2020) analysed the influence of climatic factors on the spread of Covid-19 in Iran and found that areas with low wind speed, humidity and solar radiation support the survival of the virus. The same study also found a direct correlation between population density and movement within provinces (Ahmadi et al., 2020). Mehmood et al. (2021) analysed the relationship between air pollution, climate, socioeconomic factors and infection rates in Pakistan. They reported a significant positive correlation between particulate matter ($PM_{2.5}$), an air pollutant, and the number of infections. In contrast to Ahmadi et al. the correlation between the factors humidity and wind speed and the number of infections was positive in some regions and negative in others. They also found a small negative relationship between population density and Covid-19 cases, suggesting that areas with higher population density reported proportionally fewer cases (Mehmood et al., 2021).

Pedrosa (2020) analysed the relationship between the number of cases in the US and weather, demographic variables and the infection timeline. He found that only population density and a time series variable, defined as the number of days between the first and the 100th case, showed statistical significance, while the climate in the USA had no influence on infection numbers (Pedrosa, 2020)

As the United States is one of the countries most affected by the pandemic, many studies have attempted to determine what factors are driving up the number of infections in the country. Mollalo et al. (2020) analysed the spatial variability of Covid-19 in the United States up to the 9th of April 2020. Out of 35 environmental, socio-economic, topographical and demographic variables, the four variables found to be most significant were: income inequality, median household income, the proportion of black females and the proportion of nurse practitioners at the county level (Mollalo et al., 2020).

Maiti et al. (2021) analysed infection counts up to 13 May 2020 in the USA. They observed a higher risk of Covid-19 clusters in metropolitan areas compared to rural counties, counties near central airports, more populous counties and counties with the highest proportion of racial and ethnic minorities (Maiti et al., 2021).

Wang et al. (2021) analysed the numbers up to 29 January 2021 in the US and find that factors of ethnicity, crime and income have positive correlations with the number of Covid-19 cases and explain most of the variance in the modelling estimate (Wang et al., 2021).

Allcott et al. (2020) examine partisan differences in Americans' responses to the Covid-19 pandemic, specifically how Republicans and Democrats socially distance themselves and make other efforts to reduce transmission of the disease. They model

not the risk of being infected, but how a person's political beliefs affect their beliefs about the Covid-19 pandemic. They find significant individual-level differences between Republicans and Democrats in self-reported social distancing, beliefs about their personal risk of being infected, and beliefs about the future severity of the pandemic. According to the study, Democrats find it significantly more important to stay inside to prevent the spread of the virus than to go outside to help the economy, compared to Republicans (Allcott et al., 2020).

Bermudi et al. (2021) modelled mortality in the country using latent Gaussian-Bayesian spatial models and found significant relationships between Covid-19 mortality and socioeconomic conditions, as higher socioeconomic levels, as measured by a socioeconomic index, were shown to lead to a lower risk of mortality due to Covid-19. In addition, they showed that men and older persons had the highest risk of mortality due to Covid-19 (Bermudi et al., 2021). Castro et al (2020), on the other hand, could not find a single narrative that explains the spread of the virus across the states of Brazil, but rather found that layers of complex scenarios intertwine, resulting in a different and simultaneous Covid 19 epidemic across Brazil (Castro et al., 2021).

The situation in India was analysed in a paper by Nandy et al. (2021) and the authors found that higher investment in health and education reduced the likelihood of the spread of Covid-19. In addition, a higher cure rate was found in states with sustained investment in health and education, with mortality rates also lower in states that invest more in education (Nandy et al., 2021).

Sannigrahi et al. (2020) found a significant correlation between selected demographic and socio-economic components, including total population, poverty and income, and the number of deaths from Covid-19 in Europe, without controlling for other factors such as environmental variables, socio-ecological status or climate extremes(Sannigrahi et al., 2020).

Studies have also been conducted analysing the impact of interdiction measures on the spread of Covid-19. Kasilingam et al. (2020) attempted to predict early containment of Covid-19 using machine learning models based on infrastructural and environmental variables, as well as government-implemented policies and infection-related independent variables for 42 countries. Using logistic regression, a significant positive association was found between healthcare infrastructure and lockdown policies and signs of early containment (Kasilingam et al., 2020). Orea and Álvarez (2020) also found a significant positive relationship between interdiction in Spain and its usefulness in preventing the spread of Covid-19 between different provinces in Spain. Furthermore, the same type of relationship was found for the spread of Covid-19 within the same province (Orea & Álvarez, 2020).

Contribution

The information contained in all the papers mentioned earlier shows just how many different factors may or may not be associated with the way that Coronavirus spreads in different countries. Finding a perfect model that explains why numbers are higher in one geographical region than in another is utopian, as there are still too many unknowns even more than a year into the pandemic. Achieving a scientific breakthrough is therefore beyond the scope of this work, the aim is rather to consider a wide range of factors, including infrastructural factors, demographic and socio-economic variables, when discussing the reason for different infection figures within a country and between two different countries. The countries selected for this work, Norway and Germany, are not equally affected by Covid-19, so looking for factors that influence infection numbers in both countries may be indicative of a variable that is driving infection numbers up or down, independent of the country.

1.6 Thesis Outline

The structure of the thesis is as follows. First, an introduction to Bayesian inference is given in Chapter 2. This part includes basic concepts of Bayesian theory, e.g. Bayes' theorem, which are essential for the methodology used in this thesis. Furthermore, different types of priors are introduced, as they form an integral part in Bayesian modelling. In addition, Markov-chain-Monte-Carlo-methods (MCMC methods) and latent Gaussian models are introduced, the latter of which can overcome the shortcomings of MCMC methods and form the basis for Bayesian spatial models. The last part of this chapter includes the introduction of goodness-of-fit indicators used to evaluate model performance and addresses some problems of Bayesian spatial models.

In Chapter 3 a brief introduction to the analysis of geospatial health data is given. First, different types of geospatial data, namely vector data and raster data, are introduced before discussing different methodologies used in modelling this type of data. These methodologies include the standardised incidence ratio (SIR) and the estimation of disease risk in spatial areas.

Chapter 4 gives a brief overview of the different types of data collected in this thesis and how the different data sources were combined into a coherent dataset. Chapter 5 then shows how these data were analysed. This includes calculating the SIR for Norway and Germany and modelling the relationships between different variables of interest and infection numbers in these countries. Two classes of models were calculated for this purpose: Models without a spatial component and models with such a component. Finally, an analysis is carried out to see how the results of the analysis change when different priors are used.

2 Introduction to Bayesian Inference

Bayesian inference is a branch of statistics that uses the Bayesian concept of probability and Bayes' theorem to investigate questions of stochastics.
Characteristic for Bayesian statistics is the consistent use of probability distributions or marginal distributions, whose form conveys the accuracy of the procedures or reliability of the data and the procedure.
The Bayesian concept of probability does not presuppose infinitely repeatable random experiments, so that Bayesian methods can be used even with small data sets. A small amount of data leads to a broad probability distribution, which is not strongly localised.
In the Bayesian approach, the parameters of interest are treated as random variables that are governed by their parameters, for instance the mean and standard deviation, and distributions.
Bayesian inference is an essential technique in mathematical statistics and the polar opposite of the frequentist approach, in which a hypothesis is tested without being assigned a probability.
In the Bayesian approach a *prior* distribution $\pi(\boldsymbol{\theta})$ is introduced as part of the model. This distribution is intended to express a state of knowledge or ignorance about $\boldsymbol{\theta}$ prior to obtaining the data. Using the prior distribution, the likelihood function $\pi(\boldsymbol{y}|\boldsymbol{\theta})$, and the observed data $\boldsymbol{y}$, it is possible to calculate the probability distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$. This distribution is called the *posterior* distribution of $\boldsymbol{\theta}$ and is used to make inferences about the parameters (Box & Tiao, 2011, p. 6).

14

2.1 Preliminaries

This work follows strict notation rules to easily represent different elements such as matrices or graphs and contains frequently used abbreviations. These and some other basic concepts used in this work are introduced below. The notation follows the one used by Rue and Held (Rue & Held, 2005, pp. 14–19).

2.1.1 Matrices and Vectors

Vectors and matrices are indicated by bold notation, such as $\boldsymbol{x}$ and $\boldsymbol{A}$. The transpose of $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^T$. The element in the ith row and jth column of $\boldsymbol{A}$ is referenced by A_{ij}. This notation is also used for vectors and x_i denotes the ith element of a vector. The vector $(x_i, x_{i+1}, ..., x_j)^T$ is abbreviated to $\boldsymbol{x}_{i:j}$. If the columns $\boldsymbol{A}_1, \boldsymbol{A}_2, ..., \boldsymbol{A}_m$ of a $n \times m$ matrix $\boldsymbol{A}$ are stacked on top of each other, this is denoted by $\text{vec}(\boldsymbol{A}) = \left(\boldsymbol{A}_1^T, \boldsymbol{A}_2^T, ..., \boldsymbol{A}_m^T\right)$. Deleting rows and/or columns from $\boldsymbol{A}$ creates a *submatrix*. If a submatrix of a $n \times n$ matrix $\boldsymbol{A}$ can be obtained by removing rows and columns of the same index, it is called a *principal submatrix*. If this matrix can be obtained by deleting the last $n - r$ rows and columns, it is called a *leading principal submatrix* of $\mathbf{A}$.
A diagonal $n \times n$ matrix $\boldsymbol{A}$ is denoted by $\text{diag}(\boldsymbol{A})$ and has the following structure:

$$\text{diag}(\boldsymbol{A}) = \begin{pmatrix} A_{11} & & \\ & \ddots & \\ & & A_{nn} \end{pmatrix}.$$

The identity matrix is denoted by $\boldsymbol{I}$.
If $A_{ij} = 0$ for $i < j$ or $A_{ij} = 0$ where $i > j$, then $\boldsymbol{A}$ is called *upper triangular* and *lower triangular* respectively. The *bandwidth* of a matrix $\boldsymbol{A}$ is defined as $\max\{|i - j| : A_{ij} \neq 0\}$. The *lower bandwidth* is given by $\max\{|i - j| : A_{ij} \neq 0 \text{ and } i > j\}$. $|\boldsymbol{A}|$ denotes the *determinant* of a $n \times n$ matrix $\boldsymbol{A}$ and is equal to the product of the eigenvalues of $\boldsymbol{A}$. The *rank* of $\boldsymbol{A}$, referenced by $\text{rank}(\boldsymbol{A})$, is the number of linearly independent rows or columns of $\boldsymbol{A}$. The sum of the diagonal elements is called *trace* of $\boldsymbol{A}$, $\text{trace}(\boldsymbol{A}) = \sum_i A_{ii}$.
Finally, '$\odot$' denotes the element-wise multiplication of two matrices of size $n \times m$, '$\oslash$' denotes the element-wise division and raising each element of a matrix $\boldsymbol{A}$ to a scalar power uses the symbol '$\circledcirc$' (Rue & Held, 2005, pp. 14–15).

2.1.2 Lattice and Torus

$\mathcal{I}_{\boldsymbol{n}}$ denotes a (regular) **lattice** (or grid) of size $\boldsymbol{n} = (n_1, n_2)$ (in the two-dimensional case). $\boldsymbol{y}$ can take values on $\mathcal{I}_{\boldsymbol{n}}$ and $y_{i,j}$ denotes the value of $\boldsymbol{y}$ at location ij, for $i = 1, ..., n_1$ and $j = 1, ..., n_2$. For easier reading this will be shortened to y_{ij}. On an *infinite lattice* $\mathcal{I}_{\boldsymbol{\infty}}$, ij are numbered as $i = 0, \pm 1, \pm 2, ...$, and $j = 0, \pm 1, \pm 2, ...$
A lattice with cyclic or toroidal boundary conditions is referred to as *torus* and is denoted by $\mathcal{I}_{\boldsymbol{\infty}}$. The dimension is $\boldsymbol{n} = (n_1, n_2)$ (in the two-dimensional case) and all indices are modulus $\boldsymbol{n}$ and run from 0 to $n_1 - 1$ or $n_2 - 1$. If a GMRF $\boldsymbol{y}$ is defined on $\mathcal{I}_{\boldsymbol{n}}$, the toroidal boundary conditions imply that y_{-2,n_2} is equal to $y_{n_1-2,0}$ since $-2 \mod n_1$ is equal to $n_1 - 2$ and $n_2 \mod n_2$ is equal to 0.
An *irregular lattice* refers to a spatial configuration of regions $i = 1, ..., n$ where the regions (mostly) have common boundaries, for instance the states of a nation (Rue & Held, 2005, pp. 15–16). An example of an irregular lattice is shown in Figure 2.1.

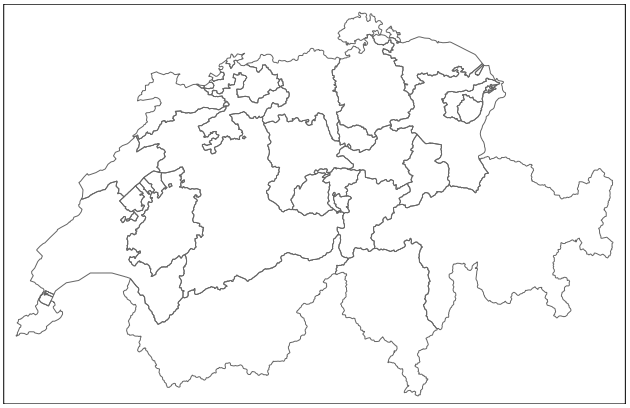

Fig. 2.1: The cantons of Switzerland, an example of an irregular lattice.

2.1.3 General Notation and Abbreviations

For $C \in \mathcal{I} = \{1, ..., n\}$ let $\boldsymbol{y}_C = \{y_i : i \in C\}$. $-C$ denotes the set $\mathcal{I} - \mathcal{C}$ such that $\boldsymbol{y}_{-C} = \{y_i : i \notin C\}$. For two sets A and B, $A \setminus B = \{i : i \in A \text{ and } i \notin B\}$.
$\pi(\cdot)$ denotes the density of its arguments, for example $\pi(\boldsymbol{y})$ for the density of $\boldsymbol{y}$ and $\pi(\boldsymbol{y}_A|\boldsymbol{y}_{-A})$ for the conditional density of $\boldsymbol{y}_A$, given $\boldsymbol{y}_{-A}$. '$\sim$' is used when a variable is 'distributed' according to the law l (Rue & Held, 2005, p. 16).

2.1.4 Symmetric Positive Definite Matrices

An $n \times n$ matrix $\boldsymbol{A}$ is *positive definite* exactly if

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0, \qquad \forall \boldsymbol{x} \neq \boldsymbol{0}.$$

If $\boldsymbol{A}$ is also symmetric, it is called a symmetric positive definite (SPD) matrix. Only SPD matrices are considered and sometimes the notation '$\boldsymbol{A} > 0$' is used for an SPD matrix $\boldsymbol{A}$.
An SPD matrix $\boldsymbol{A}$ has some of the following properties.

1. $\text{rank}\,(\boldsymbol{A}) = n$.
2. $|\boldsymbol{A}| > 0$.
3. $A_{ii} > 0$.
4. $A_{ii}A_{jj} - A_{ij}^2 > 0$, for $i \neq j$.
5. $A_{ii} + A_{jj} - 2|A_{ij}| > 0$ for $i \neq j$.
6. $\max A_{ii} > \max_{i \neq j} |A_{ij}|$.
7. $\boldsymbol{A}^{-1}$ is SPD.
8. All principal submatrices of $\boldsymbol{A}$ are SPD.

If $\boldsymbol{A}$ and $\boldsymbol{B}$ are SPD, $\boldsymbol{A} + \boldsymbol{B}$ is also SPD, but the reverse is generally not true. Additionally, if $\boldsymbol{AB} = \boldsymbol{BA}$, $\boldsymbol{AB}$ is SPD.
The following conditions are all sufficient and necessary for a symmetric matrix $\boldsymbol{A}$ to be SPD:

1. All eigenvalues $\lambda_1, ..., \lambda_n$ of $\boldsymbol{A}$ are strictly positive.
2. There exists such a matrix $\boldsymbol{C}$ that $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{C}^T$. If $\boldsymbol{C}$ is lower triangular, it is called the *Cholesky triangle* of $\boldsymbol{A}$.
3. All leading principal submatrices have strictly positive determinants.

A sufficient, but not necessary condition for a (symmetrical) matrix to be SPD is the criterion of *diagonal dominance*:

$$A_{ii} - \sum_{j:j \neq i} |A_{ij}| > 0, \qquad \forall i.$$

A $n \times n$ matrix $\boldsymbol{A}$ is called a *symmetric positive semidefinite* (SPSD) matrix. An SPSD matrix $\boldsymbol{A}$ is sometimes denoted '$\boldsymbol{A} \geq 0$' (Rue & Held, 2005, pp. 18–19).

2.2 Basic Concepts of Bayesian Theory

To understand Bayesian theory, it is helpful to first introduce a few basic concepts, first and foremost Bayes' theorem, which is introduced in Section 2.2.1, one of the most famous concepts in all of statistics. Other notions that are integral to the rest of this thesis are the concept of conditional independence, undirected graphs and the computation of summary statistics, the latter of which is an essential part of the analysis section of this thesis.

2.2.1 Bayes' Theorem

At the heart of Bayesian inference is *Bayes' theorem*, which describes the probability of an event given prior knowledge of factors that might influence the event.
Let $\boldsymbol{y}^T = (y_1, ..., y_n)$ be a vector of n observations whose probability distribution $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ depends on the values of k parameters $\boldsymbol{\theta}^T = (\theta_1, ..., \theta_k)$. Let $\pi(\boldsymbol{\theta})$ be the probability distribution of $\boldsymbol{\theta}$. Then

$$\pi(\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{y}, \boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|\boldsymbol{y})\,\pi(\boldsymbol{y}). \tag{2.1}$$

Given the observed data $\boldsymbol{y}$, the conditional distribution of $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y})}. \tag{2.2}$$

This last statement is known as Bayes' theorem (Bayes, 1763). The *prior* distribution $\pi(\boldsymbol{\theta})$ contains knowledge about $\boldsymbol{\theta}$ without knowledge of the data. $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ contains what is known about $\boldsymbol{\theta}$ given knowledge of the data and is the *posterior* distribution of $\boldsymbol{\theta}$ given $\boldsymbol{y}$.
If $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ is considered as a function of $\boldsymbol{\theta}$ instead of $\boldsymbol{y}$, it is called the *likelihood function* of $\boldsymbol{\theta}$ given $\boldsymbol{y}$ and can be written as $l(\boldsymbol{\theta}|\boldsymbol{y})$. Thus Bayes' theorem can be written as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto l(\boldsymbol{\theta}|\boldsymbol{y})\,\pi(\boldsymbol{\theta}). \tag{2.3}$$

It is evident that the posterior distribution of $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$ is proportional to the product of the distribution of $\boldsymbol{\theta}$ prior to observing the data and the likelihood function of $\boldsymbol{\theta}$ given $\boldsymbol{y}$. Therefore,

$$\text{posterior distribution} \propto \text{likelihood} \times \text{prior distribution}.$$

The data $\boldsymbol{y}$ modifies the prior knowledge of $\boldsymbol{\theta}$ through the likelihood function, and thus can be be regarded as a representation of the information about $\boldsymbol{\theta}$ derived from the data (Box & Tiao, 2011).

2.2.2 Conditional Independence

In probability theory, two random variables x and y are *independent* given a third variable z if and only if the occurrence of x and y in their conditional probability distribution given z are independent events. To calculate the conditional density of $\boldsymbol{x}_A$, given $\boldsymbol{x}_{-A}$, the following statement will repeatedly be used,

$$\pi\left(\boldsymbol{x}_A|\boldsymbol{x}_{-A}\right) = \frac{\pi\left(\boldsymbol{x}_A, \boldsymbol{x}_{-A}\right)}{\pi\left(\boldsymbol{x}_{-A}\right)} \propto \pi\left(\boldsymbol{x}\right). \tag{2.4}$$

It follows that x and y are independent precisely when $\pi\left(x, y\right) = \pi\left(x\right)\pi\left(y\right)$, which is expressed by $x \perp y$. x and y are conditionally independent for a given z if and only if $\pi\left(x, y|z\right) = \pi\left(x|z\right)\pi\left(y|z\right)$ (Dawid, 1979). The conditional independence can be easily validated with the help of the following *factorisation criterion*,

$$x \perp y|z \Longleftrightarrow \pi\left(x, y, z\right) = f\left(x, z\right) g\left(y, z\right), \tag{2.5}$$

for some functions f and g, and for all z with $\pi\left(z\right) > 0$ (Rue & Held, 2005, pp. 16–17).

2.2.3 Undirected Graphs

Undirected graphs are used to represent the conditional independence structure in a Gaussian Markov random field. An *undirected graph* $\mathcal{G}$ is defined as a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ contains all nodes in the graph and $\mathcal{E}$ is the set of edges $\{i, j\}$, with $i, j \in \mathcal{V}$ and $i \neq j$. For $\{i, j\} \in \mathcal{E}$ there exists an undirected edge from node i to node j in the other case such an edge does not exist. If $\{i, j\} \in \mathcal{E}\ \forall i, j \in \mathcal{V}$ with $i \neq j$ a graph is *fully connected*. Most often $\mathcal{V} = \{1, 2, ..., n\}$ will be assumed, which is referred to as *labelled*. A simple example of an undirected graph is shown in Figure 2.2.

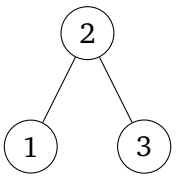

Fig. 2.2: An undirected labelled graph with 3 nodes, $\mathcal{V} = \{1, 2, 3\}$ and $\mathcal{E} = \{\{1, 2\}\{2, 3\}\}$.

The *neighbours* of node i are defined as all nodes in $\mathcal{G}$ with an edge to node i,

$$\text{ne}(i) = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}.$$

This definition can be extended to a set $A \subset \mathcal{V}$, where the neighbours of A are defined as

$$\text{ne}(A) = \bigcup_{i \in A} \text{ne}(i) \setminus A.$$

A *path* from i_1 to i_m is defined as a sequence of certain nodes in $\mathcal{V}, i_1, i_2, ..., i_m$, for which $(i_j, i_{j+1}) \in \mathcal{E}$ for $j = 1, ..., m-1$. Two nodes $i \notin C$ and $j \notin C$ are *separated* by a subset $C \subset \mathcal{V}$, if every path from i to j contains at least one node from C. Two disjoint sets $A \subset \mathcal{V} \notin C$ and $B \subset \mathcal{V} \notin C$ are separated by C, if all $i \in A$ and $j \in B$ are separated by C, that is, it is not possible to "wander" on the graph from somewhere in A and end somewhere in B without crossing C.
If i and j are neighbours in $\mathcal{G}$, this can be expressed by $i \overset{\mathcal{G}}{\sim} j$ or $i \sim j$ for the case where the graph is implicit. It follows that $i \sim j \iff j \sim i$.
Let A be a subset of $\mathcal{V}$. A *subgraph* $\mathcal{G}^A$ is a graph restricted to A, i.e., the graph obtained after removing all nodes that do not belong to A and all edges where at least one node does not belong to A. $\mathcal{G}^A = \left\{\mathcal{V}^A, \mathcal{E}^A\right\}$, where $\mathcal{V}^A = A$ and

$$\mathcal{E}^A = \{\{i, j\} \in \mathcal{A} \text{ and } \{i, j\} \in A \times A\}.$$

Let $\mathcal{G}$ be the graph in Figure 2.2 and $\mathcal{A} = \{2, 3\}$, then $\mathcal{V}^A = \{2, 3\}$ and $\mathcal{E}^A = \{\{2, 3\}\}$ (Rue & Held, 2005, pp. 17–18).

2.2.4 Calculation of Summary Statistics

As the posterior mean and the credibility intervals of coefficient are of interest, calculation of these is performed later on. This allows a better interpretation of the results.

To get the expected value given a marginal function $\pi(x)$, the expected value of a function $f(x)$ is calculated, i.e.

$$\int f(x)\,\pi(x)\,dx \tag{2.6}$$

(Aitkin, 1991).
If necessary, e.g. if the target variable follows a (negative) binomial distribution, the value must be transformed to its original scale, as in these cases the log-likelihood is modelled. Therefore, in these cases, the expected value would have to be exponentiated to allow a clear interpretation.
In practice, to obtain the credibility interval of a variable, the marginal values are first transformed to their original scale, if necessary, and then the 2.5% quantile and the 97.5% quantiles are calculated.

2.3 Prior Selection

A key question in Bayesian analysis is the effect of the prior on the posterior, and how that effect can be measured. Do posterior distributions derived with different priors become very similar as more and more data is collected? It has been formally proven that under certain regularity conditions, the impact of the prior decreases with increasing sample size. From a practical point of view, it is more important to know what happens when the sample size n is finite. In this section, different types of priors are introduced.

2.3.1 Conjugate Priors

One property of exponential families is that they have conjugate priors (Diaconis & Ylvisaker, 1979), which is an important property in Bayesian statistics. If the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and the prior distribution $\pi(\boldsymbol{\theta})$ belong to the same probability distribution family, the prior and posterior distributions are called *conjugate* distributions. Furthermore, the prior for the likelihood function $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ is called the *conjugate prior*. These priors were first discussed and formalised by Raifa and Schlaifer in 1961 (Raiffa & Schlaifer, 1961).
The construction of a conjugate prior is done by factorising the likelihood function into two parts. One part must be independent of the parameter(s) of interest but can be dependent on the data, while the other factor is a function that depends on the parameter(s) of interest and is dependent on the data only through the sufficient statistics. The family of conjugate priors is by definition proportional to the second factor. The posterior distribution resulting from the conjugate prior is itself a member of the same family as the conjugate prior (Raiffa & Schlaifer, 1961). In cases where the prior and posterior distributions are part of the same family, the prior is said to be closed under sampling. Furthermore, since the data are only incorporated into the posterior distribution through the sufficient statistics, there exist relatively simple formulas for updating the prior into the posterior.
For an example of the construction of a conjugated prior, see Fink 1997 (Fink, 1997).
A drawback of conjugated priors is that the a priori known information about μ may be insufficient for determining both parameters or may be inconsistent with the structure imposed by conjugacy (C. P. Robert et al., 2010). Moreover, these priors can be too restrictive and not every belief about the prior can be described (Irwin, 2005).
Thus, although conjugate priors are easy to handle both mathematically and com-

putationally (Irwin, 2005), they are not often used in practice because of these drawbacks.

2.3.2 Penalised Complexity Priors

One issue when selecting the prior distribution of a particular parameter is that it is not always intuitive when it comes to understanding and interpreting this distribution, something that is essential to ensure that it behaves as intended by the user. This problem can be addressed by using *penalised complexity priors*, which is a methodology that penalises the complexity of model components in relation to deviation from simple base model formulations.
PC priors provide a systematic and unified approach to calculating prior distributions for parameters of model components by using an inherited nested structure. This structure contains two models, the base model and a flexible version of the model. The first of the two is generally characterised by a fixed value of the relevant parameter, while the second version is considered a function of the random parameter. By penalising the deviation from the flexible model to the fixed base model, the PC prior is calculated (Martins et al., 2014).

2.3.2.1 The Principles Behind PC Priors

Four main principles should be followed to calculate priors in a consistent way and to understand their properties.

Support to Occam's Razor

Let $\pi(x|\xi)$ denote the density of a model component x and ξ the parameter to which a prior distribution is to be assigned. The base model is characterised by a density $\pi(x|\xi = \xi_0)$, where ξ_0 is a fixed value. The prior for ξ should be such that proper shrinkage is given to ξ_0. The simplicity of the model is therefore prioritised over the complexity of the model, preventing overfitting (Martins et al., 2014).

Penalisation of Model Complexity

Let $f_1 = \pi(x|\xi)$ and $f_0(x|\xi = \xi_0)$ denote the flexible model and the base model respectively. The complexity of f_1 compared to f_0 is characterised using the Kullback-Leibler divergence (Kullback & Leibler, 1951) to calculate a measure of complexity between the two models,

$$\text{KLD}(f_1||f_2) = \int f_1(x) \log\left(\frac{f_1(x)}{f_0(x)}\right) dx. \tag{2.7}$$

This can be used to measure the information that is lost when f_1 is approximated by the simpler model f_0. For multinormal densities with zero mean, the calculation simplifies to

$$\text{KLD}(f_1||f_0) = \frac{1}{2}\left(\text{trace}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1\right) - n - \log\left(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|}\right)\right), \tag{2.8}$$

where $f_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i), i = 0, 1$, while n represents the dimension. For easier interpretation, the Kullback-Leibler divergence is transformed into a unidirectional distance measure

$$d(\xi) = d(f_1||f_0) = \sqrt{2\text{KLD}(f_1||f_0)} \tag{2.9}$$

which can be interpreted as a measure of distance from f_1 to f_0 (Martins et al., 2014).

Constant Rate Penalisation

The derivation of the PC prior is based on a system of constant rate penalisation, given by

$$\frac{\pi_d(d(\xi) + \delta)}{\pi_d(d(\xi))} = r^\delta, \qquad d(\xi), \delta \geq 0. \tag{2.10}$$

$r \in (0, 1)$ represents the constant decay rate and thus implies that the relative change in the prior distribution for $d(\xi)$ is independent of the actual distance. Therefore, $d(\xi)$ is exponentially distributed with density $\pi(d(\xi)) = \lambda \exp(-\lambda d(\xi))$ and rate $\lambda = -\ln(r)$. By a standard variable change transformation, the corresponding PC prior for ξ is given (Martins et al., 2014).

User-Defined Scaling

Since λ characterises the shrinkage properties of the prior, it is important that the rate can be chosen in an intuitive and interpretable way. One possibility is to determine λ by including a probability statement of tail events, for example

$$\mathbb{P}\left(Q\left(\xi\right) > U\right) = \alpha, \tag{2.11}$$

where U represents an assumed upper bound for an interpretable transformation $Q\left(\xi\right)$ and α denotes a small probability (Martins et al., 2014).

2.3.2.2 Example: PC Prior for the Precision

A PC prior can be used to adjust the smoothness of a spatial field in an intuitive way by specifying such a prior for the precision τ. This makes it possible to adjust the smoothness of the spatial field in an intuitive way. In this case, the penalised complexity prior is defined by the parameter σ_0. Equation 2.11 therefore looks like this,

$$\mathbb{P}\left(\sigma > \sigma_0\right) = \alpha. \tag{2.12}$$

The actual expression of the prior is given by

$$\pi\left(\tau\right) = \frac{\lambda}{2}\tau^{-3/2}\exp\left(-\lambda\tau^{-1/2}\right), \qquad \tau > 0 \tag{2.13}$$

and is a type-2-Gumbel distribution.
The prior for τ corresponds to an exponential distribution with rate λ for the standard deviation.
λ quantifies the size of the penalty for deviation from the base model and is increased with higher values for it.
Here, $\lambda = \frac{-\log(\alpha)}{\sigma_0}$ (Martins et al., 2014).

2.4 Markov-Chain-Monte-Carlo-Methods

Markov chain Monte Carlo methods, also referred to as MCMC methods, are a set of algorithms that enable sampling from probability distributions based on the construction of Markov chains. After a sufficient number of iterations, the stationary distribution of a Markov chain can be taken as the desired distribution, with the quality of this distribution improving as the number of iterations increases. Most of the time, the construction of such a chain is relatively simple; the real challenge is to determine how many steps are needed before convergence towards the stationary distribution is achieved. MCMC methods are mostly used to compute numerical approximations of multidimensional integrals, for instance in Bayesian statistics or computational biology. The two main concepts used in MCMC methods are Monte Carlo integration and the aforementioned Markov chains, hence the name Markov Chain Monte Carlo.

2.4.1 Monte Carlo Integration

Monte Carlo integration is a technique that uses the generation of random numbers for numerical computation of definite integrals and is especially useful for higher-dimensional integrals. The problem the method addresses is the computation of the integral

$$\mathbb{E}_f\left[h\left(X\right)\right] = \int_{\chi} h(x) f(x) dx. \tag{2.14}$$

The integral can be approximated by using a sample $(X_1, ..., X_m)$ generated from f and calculating the arithmetic mean

$$\overline{h}_m = \frac{1}{m} \sum_{j=1}^{m} h\left(x_j\right). \tag{2.15}$$

According to the Strong Law of Large Numbers, $\overline{h}_m$ is likely to converge to $\mathbb{E}_f\left[h\left(X\right)\right]$. When the expectation of h^2 under f is finite, the convergence speed of $\overline{h}_m$ can be assessed. The variance too can be estimated from the sample $(X_1, ..., X_n)$ through

$$v_m = \frac{1}{m^2} \sum_{j=1}^{m} \left[h\left(x_j\right) - \overline{h}_m\right]^2. \tag{2.16}$$

For m large,

$$\frac{\overline{h}_m - \mathbb{E}_f\left[h\left(X\right)\right]}{\sqrt{v_m}} \tag{2.17}$$

is approximately distributed as a $\mathcal{N}(0,1)$ variable. This can be used for constructing a convergence test and to calculate confidence bounds for the approximation of $\mathbb{E}_f[h(X)]$ (C. Robert & Casella, 2013, pp. 83–84). The term Monte Carlo was first used in 1949 by Metropolis and Ulam to describe a method dealing with problems related to "integro-differential equations that occur in various branches of the natural sciences" (Metropolis & Ulam, 1949).

2.4.2 Markov Chains

Markov chains are stochastic processes that aim to provide the probability of the occurrence of future events. A Markov chain is defined by the fact that even if only a limited history is known, predictions about future developments can be made just as reliably as if the entire history of a process were known. Thus, the probability of moving from the current state to any state depends only on the current state of the chain. These probabilities are defined by a *transition kernel*, which is a function K on $\mathcal{X} \times \mathcal{B}(\mathcal{X})$, such that

i. $\forall x \in \mathcal{X}, K(x,\cdot)$ is a probability measure;

ii. $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ is measureable.

In the discrete case, the transition kernel is a matrix $\boldsymbol{K}$ with elements

$$\mathbb{P}_{xy} = \mathbb{P}(X_n = y | X_{n-1} = x), \qquad x, y \in \mathcal{X}.$$

If $\mathcal{X}$ is continuous, the kernel denotes the conditional density $K\left(x, x^T\right)$ of the transition $K(x,\cdot)$,

$$\mathbb{P}(X \in A|x) = \int_A K\left(x, x^T\right) dx^T.$$

Given a transition kernel K, a sequence $X_0, X_1, ..., X_t$ of random variables is a *Markov chain* (X_n), if, for any t, the conditional distribution of X_t given the previous states is the same as the distribution of X_t given the last state, x_{t-1},

$$\begin{aligned}\mathbb{P}(X_{t+1} \in A | x_0, x_1, x_2, ..., x_t) &= \mathbb{P}(X_{t+1} \in A | x_t) \\ &= \int_A K(x_t, dx). \end{aligned} \tag{2.18}$$

These chains were first introduced by Markov in 1906 (Markov, 1906). Markov chains can have certain properties that affect their long-term behaviour and are of particular importance for MCMC algorithms. In sections 2.4.2.1– 2.4.2.5, some of them will be introduced.

2.4.2.1 Irreducibility

Irreducibility is critical to the construction of Markov chain Monte Carlo algorithms, as it ensures the convergence of such an algorithm. A Markov chain is *irreducible* if all states communicate, that is, for all states i and j the probability of getting from i to j in finite time is true positive.
Formally speaking, given a measure φ, a Markov chain (X_n) with transition kernel $K(x, y)$ is *φ-irreducible*, if, for every $A \in B(\mathcal{X})$ with $\varphi(A) > 0$, there exists n such that $K^n(x, A) \forall x \in \mathcal{X}$. The chain is *strongly φ-irreducible* if $n = 1 \forall$ measurable A (C. Robert & Casella, 2013, pp. 213–214).

2.4.2.2 Periodicity

The behaviour of a Markov chain can sometimes be limited by deterministic constraints on the transitions from X_n to X_{n+1}. For discrete chains, the *period* of a state $w \in \mathcal{X}$ is defined as.

$$d(w) = \text{g.c.d. } \{m \geq 1; K^m(w, w) > 0\},$$

with g.c.d the greatest common denominator. If a Markov chain is irreducible, the transition matrix can be written as a block matrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & \boldsymbol{D}_1 & 0 & \dots & 0 \\ 0 & 0 & \boldsymbol{D}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & & \\ \boldsymbol{D}_d & 0 & 0 & & 0 \end{pmatrix}, \tag{2.19}$$

It is evident that at every d-th step there is a return to the initial group. There exists only one value for the period when a chain is irreducible. If this value is 1, the irreducible chain is *aperiodic* (C. Robert & Casella, 2013, pp. 217–218).

2.4.2.3 Transience and Recurrence

To guarantee an acceptable approximation of a simulated model, a Markov chain needs to have good stability properties. Irreducibility is not strong enough to ensure that the trajectory of (X_n) enters A often enough. This leads to the formalisation of *recurrence* and *transience*.
In a finite space $\mathcal{X}$, a state $w \in \mathcal{X}$ is *transient* if it is finitely often visited and *recurrent* if it is almost certainly infinitely often visited.

For irreducible chains, these two properties are properties of the chain, not of a particular state (C. Robert & Casella, 2013, pp. 218–219).

2.4.2.4 Ergodicity

When looking at a Markov chain (X_n) from a temporal point of view, it is essential to establish to what the chain is converging. A natural candidate for the limiting distribution is the stationary distribution π which leads to the need to define sufficient conditions on (X_n) for X_n to be asymptotically distributed according to π. There are several conditions that can be imposed on the convergence of P^n, the distribution of X_n to π. The most fundamental and important is that of *ergodicity*, that is, independence of initial conditions.
If a Markov chain (X_n) is both aperiodic and positive recurrent, it is called an *ergodic* Markov chain (C. Robert & Casella, 2013, pp. 231–234).

2.4.2.5 Stationary distribution

A chain (X_n) is more stable if the marginal distribution of X_n is independent of n. This is a requirement for the existence of a probability distribution π such that $X_{n+1} \sim \pi$ if $X_n \sim \pi$. Markov chain Monte Carlo methods rely on the fact that this condition can be satisfied.
A σ-finite measure π is *invariant* for the transition kernel $K(\cdot, \cdot)$ if

$$\pi(B) = \int_{\mathcal{X}} K(x, B)\, \pi(dx), \qquad \forall B \in \mathcal{B}(\mathcal{X}).$$

This distribution is referred to as *stationary* if π is a probability measure, as $X_0 \sim \pi$ implies that $X_n \sim \pi$ is $\forall n$. An irreducible Markov chain has a stationary distribution precisely if it is positively recurrent. The distribution is then given by

$$\pi_x = \left(\mathbb{E}_x\left[\tau_x\right]\right)^{-1}, \qquad x \in \mathcal{X}, \tag{2.20}$$

where $\mathbb{E}_x\left[\tau_x\right]$ can be interpreted as the average number of transitions between two passages in x.
In practice, the stationary distributions are often of special interest. If these distributions are defined as the starting distribution of X_0, then all following distributions of the states X_n for any n are equal to the starting distribution. The interesting question here is when such distributions exist and when any distribution converges against a stationary distribution of this kind (C. Robert & Casella, 2013, pp. 223–224).

2.4.3 The Metropolis-Hastings Algorithm

Having established the basics of MCMC methods, one of the best known MCMC algorithms, the Metropolis-Hastings algorithm, is introduced next. The algorithm is based on the Metropolis algorithm, which was developed to simulate the states of a system according to the Boltzmann distribution, with the newest state always depending on the previous state (Metropolis et al., 1953).
The Metropolis-Hastings algorithm is a procedure for drawing random samples from a probability distribution from which direct sampling is difficult if a function proportional to the *target density* f is known. This function $q(\boldsymbol{y}|\boldsymbol{x})$ is called the *proposal density* and must be easy to simulate in order for the Metropolis-Hastings algorithm to be implementable. Moreover, it must be either explicitly present or *symmetric*, meaning $q(\boldsymbol{x}|\boldsymbol{y}) = q(\boldsymbol{y}|\boldsymbol{x})$.
The Metropolis-Hastings algorithm of a target density f and proposal density q produces a Markov chain $\left(X^{(t)}\right)$ by the following transition.

Algorithm 1 The Metropolis-Hastings Algorithm

Given $f(\boldsymbol{x})$ and $q(\boldsymbol{y}|\boldsymbol{x})$
1: Initialisation: Choose arbitrary x_t as the first sample
2: **for** each iteration t **do**
3: Generate $Y_t \sim q\left(\boldsymbol{y}|x^{(t)}\right)$
4: Take

$$X^{(t+1)} = \begin{cases} Y_t & \text{with probability } \mathbb{P}\left(x^{(t)}, Y_t\right) \\ x^{(t)} & \text{with probability } 1 - \mathbb{P}\left(x^{(t)}, Y_t\right) \end{cases}$$

where

$$\mathbb{P}(x, y) = \min\left\{\frac{f(\boldsymbol{y})}{f(\boldsymbol{x})}\frac{q(\boldsymbol{x}|\boldsymbol{y})}{q(\boldsymbol{y}|\boldsymbol{x})}, 1\right\}. \tag{2.21}$$

$\mathbb{P}(x, y)$ is the *Metropolis-Hastings acceptance probability*.
The algorithm always accepts values y_t that lead to an increase in the ratio $f(y_t)/q\left(y_t|x^{(t)}\right)$ compared to the previous value $f\left(x^{(t)}\right)/q\left(x^{(t)}|y_t\right)$. In the symmetric case, the acceptance probability simplifies to

$$\mathbb{P}(x, y) = \min\left\{\frac{f(\boldsymbol{y})}{f(\boldsymbol{x})}, 1\right\}$$

(Hastings, 1970).
If the Markov chain starts with a value $x^{(0)} > 0$, then $f\left(x^{(t)}\right) > 0\ \forall t \in \mathbb{N}$ since the

values of y such that $f\left(y_t\right) = 0$ will all be rejected by the algorithm. As the number of iterations t increases, the distribution of saved states $x_0, ..., x_t$ will converge towards the target density $f(\boldsymbol{x})$ (C. Robert & Casella, 2013, pp. 270–275).

2.5 Latent Gaussian Models and INLA

In recent years, a growing amount of georeferenced data has become available, leading to an increased need for appropriate statistical modeling to handle large and complex datasets. The usual approach to inference for latent Gaussian models involves the previously introduced Markov chain Monte Carlo methods. Due to several factors, these methods may perform poorly when applied to such models. One factor is the interdependence of the components of the latent field $\boldsymbol{x}$ while another is that $\boldsymbol{\theta}$ and $\boldsymbol{x}$ are highly dependent on each other, especially for large n. The first of these problems can potentially be overcome by constructing a joint proposal based on a Gaussian approximation of the full conditional of $\boldsymbol{x}$ (Gamerman, 1997), while the second problem requires, at least in part, a joint update of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. There are several proposals to solve these shortcomings, but MCMC sampling continues to show poor computational speed (Rue et al., 2009, p. 322).
Bayesian hierarchical models have proven to be effective in capturing complex stochastic structures in spatial processes. A large proportion of these models are based on latent Gaussian models, a subclass of structured additive regression models. These models include Integrated Nested Laplace Approximations (INLA), which are a class of models used to approximate the posterior marginals of a latent Gaussian field. These approximations work by reformulating the regression model as a three-part hierarchical model. The parts are as follows:

1. Approximation of the posterior marginal of $\boldsymbol{\theta}$ by using the Laplace approximation.

2. Computation of the (simplified) Laplace approximation for selected values of $\boldsymbol{\theta}$ to improve the Gaussian approximation.

3. Combination of the first two parts by numerical integration.

$\boldsymbol{\theta}$ is a vector of hyperparameters. The hyperparameters $\boldsymbol{\theta}$ can be, for example, the variance in the Gaussian likelihood or the shape parameter in the likelihood of the gamma distribution. In the case of latent fields, they can be, for instance, dispersion parameters or spatial correlation parameters (Rue et al., 2009).

2.5.1 Notation and Basic Properties

For structured additive regression models, the distribution of the response variable y_i is assumed to be a member of the exponential family, with the mean μ_i linked to

a structured additive predictor η_i by a link function $g(\cdot)$ such that $g(\mu_i) = \eta_i$. The predictor η_i takes into account the effect of multiple covariates in an additive way,

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i, \qquad i = 1, ..., n \tag{2.22}$$

(Stone, 1985).
The $\left\{f^{(j)}(\cdot)\right\}$s are unknown functions of the covariates u, while the $\{\beta_k\}$s represent the linear effect of the covariates z and the ϵ_is are unstructured terms. Latent Gaussian models assign a Gaussian prior to α, $\left\{f^{(j)}(\cdot)\right\}$ and $\{\epsilon_i\}$. In the following $\boldsymbol{x}$ shall denote the vector of all latent Gaussian variables ($\{\eta_i\}$, α, $\left\{f^{(j)}\right\}$ and $\{\beta_k\}$) and $\boldsymbol{\theta}$ the vector of hyperparameters.
The conditional density $\pi(\boldsymbol{x}|\theta_1)$ is Gaussian with an assumed zero mean and precision matrix $\boldsymbol{Q}(\theta_1)$. The Gaussian density $\mathcal{N}(\mu, \boldsymbol{\Sigma})$ with mean μ and covariance $\boldsymbol{\Sigma}$ at configuration $\boldsymbol{x}$ is denoted by $\mathcal{N}(\boldsymbol{x}; \mu, \boldsymbol{\Sigma})$. For simplicity, $\{\eta_i\}$ has been included instead of $\{\epsilon_i\}$.
The distribution for the n_d observational variables $y = \{y_i : i \in \mathcal{I}\}$ is denoted by $\pi(\boldsymbol{y}|\boldsymbol{x}, \theta_2)$ and is assumed conditionally independent given $\boldsymbol{x}$ and θ_2. Let $\boldsymbol{\theta} = \left(\theta_1^T, \theta_2^T\right)^T$ with $\dim(\boldsymbol{\theta}) = m$. For non-singular $\boldsymbol{Q}(\boldsymbol{\theta})$ the posterior is given by

$$\begin{aligned}\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_{i \in I} \pi(y_i|x_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) |\boldsymbol{Q}(\boldsymbol{\theta})|^{1/2} \exp\left[-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i \in I} \log\{\pi(y_i|x_i, \boldsymbol{\theta})\}\right].\end{aligned} \tag{2.23}$$

Most latent Gaussian models satisfy two basic properties:

1. The latent field $\boldsymbol{x}$ is of large dimension, $n \approx 10^2 - 10^5$. Therefore, the latent field is a Gaussian Markov random field with sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$.

2. The number of hyperparameters, m, is small, $m \leq 6$.

In most cases, both properties are required to produce fast inference, and thus these will be assumed to be true for the remainder of this work (Rue et al., 2009).

2.5.2 Applications for Latent Gaussian Models

Latent Gaussian models can be employed in a vast range of different domains, in fact most structured Bayesian models are of this particular form. Some of these domains are presented below.

2.5.2.1 Regression Models

Bayesian generalised linear models correspond to the linear relationship $\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki}$ (Dey et al., 2000). Either the linear relationship of the covariates can be relaxed through the $f(\cdot)$ terms (Fahrmeir & Tutz, 2013), random effects can be introduced through them or both. Smooth covariate effects are frequently modeled using penalised spline models (Lang & Brezger, 2004) or random walk models (Fahrmeir & Tutz, 2013), continuous indexed spline models (Rue & Held, 2005) or Gaussian processes (Chu et al., 2005). The incorporation of random effects allows for the consideration of overdispersion caused by unobserved heterogeneity or correlation in longitudinal data and can be introduced by defining $f(u_i) = f_i$ and $\{f_1\}$ to be independent, zero mean and Gaussian (Fahrmeir & Lang, 2001).

2.5.2.2 Dynamic Models

Temporal dependence can be introduced by using i in (2.22) as temporal index t and defining $f(\cdot)$ and $\boldsymbol{u}$ such that $f(u_t) = f_t$. Both a discrete-time and a continuous-time autoregressive model can be modeled by $\{f_t\}$. Furthermore, a seasonal effect or the latent process of a structured time series model can be modeled (Kitagawa & Gersch, 1996). Alternatively, a smooth temporal function in the same sense as for regression models can be represented by $\{f_t\}$.

2.5.2.3 Spatial and Spatio-Temporal Models

Similar to the previous type of model, spatial dependence can be modeled by a spatial covariate $\boldsymbol{u}$ such that $f(u_s) = f_s$, where s denotes the spatial location or region s. The stochastic model for f_s is constructed to promote spacial smooth realisations of some sort. Popular models of this type include the Besag-York-Mollié (Besag et al., 1991) model with extensions for regional data, continuous indexed Gaussian models (Banerjee et al., 2014) and texture models (Marroquin et al., 2001). The dependence between spatial and temporal covariates can be achieved either by using a spatio-temporal covariate (s, t) or a corresponding spatio-temporal Gaussian field (Kammann & Wand, 2003).
Often the final model consists of a sum of several components, e.g. a spatial component, random effects and both linear and smooth effects of some covariates. In order to separate the effects of the different components in (2.22), sometimes linear or sum-to-zero constraints can be imposed (Rue et al., 2009, pp. 319–321).

2.5.3 Gaussian Random Fields

Let $\boldsymbol{s} = (s_1, ..., s_n)^T$ be a vector of locations. A *Gaussian random field* (GRF)

$$\left\{ Z(s) : s \in D \subset \mathbb{R}^2 \right\} \tag{2.24}$$

is a set of random variables where the observations occur in a continuous domain and where each finite set of random variables follows a multivariate normal distribution. A random process $Z(\cdot)$ is strictly stationary if it is invariant to shifts, i.e., if for each set of locations and each $h \in \mathbb{R}^2$ the distribution of $\boldsymbol{Z(s)} = (Z(s_1), ..., Z(s_n))$ is equal to that of $\boldsymbol{Z(s+h)} = (Z(s_1 + h), ..., Z(s_n + h))$. A less constraining requirement is given by second-order stationarity. Under this condition, the process has a constant mean value

$$\mathbb{E}[\boldsymbol{Z(s)}] = \mu, \qquad \forall s \in D, \tag{2.25}$$

and the covariances depend only on the differences between locations

$$\text{Cov}(\boldsymbol{Z(s)}, \boldsymbol{Z(s+h)}) = C(\boldsymbol{h}), \qquad \forall \boldsymbol{s} \in D, \forall \boldsymbol{h} \in \mathbb{R}^2. \tag{2.26}$$

Furthermore, if the covariances depend only on the distances between the locations and not on the directions, the process is called isotropic. Else, the process is anisotropic. An intrinsically stationary process has a constant mean value and satisfies

$$\text{Var}(Z(s_i) - Z(s_j)) = 2\gamma(s_i - s_j), \qquad \forall s_i, s_j. \tag{2.27}$$

$2\gamma(\cdot)$ is the variogram and $\gamma(\cdot)$ is called the semivariogram (Cressie, 2015). Under the assumption of intrinsic stationarity, the constant-mean assumption implies

$$2\gamma(\boldsymbol{h}) = \text{Var}(\boldsymbol{Z(s+h)} - \boldsymbol{Z(s)}) = \mathbb{E}\left[(\boldsymbol{Z(s+h)} - \boldsymbol{Z(s)})^2 \right],$$

and the estimation of the semivariogram can be obtained using the empirical semivariogram as follows:

$$2\widehat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} (Z(s_i) - Z(s_j))^2, \tag{2.28}$$

where $N(\boldsymbol{h}) = \{(s_1, s_j) : s_i - s_j = \boldsymbol{h}, i, j = 1, ..., n\}$ denotes the number of pairs and $|N(\boldsymbol{h})|$ the number of distinct pairs. For isotropic processes, the semivariogram is a function of distance $h = ||\boldsymbol{h}||$.
Plotting the empirical semivariogram against the separation distance conveys essential information regarding the continuity and spatial variability of the process. Given

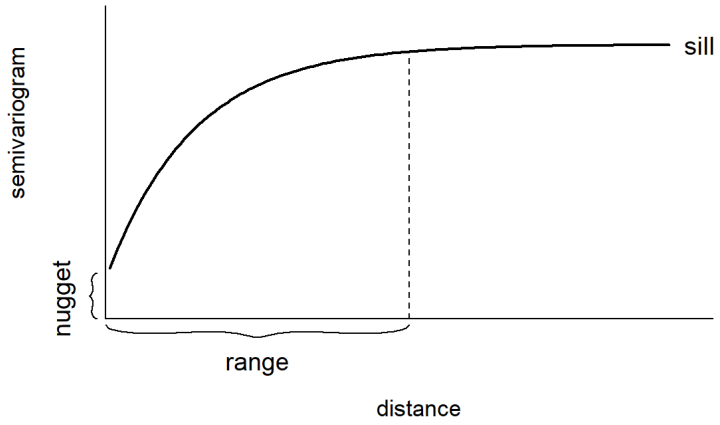

Fig. 2.3: A typical semivariogram

relatively short distances, the semivariogram tends to be small but increases with distance, indicating the similarity of observations in close proximity. The semivariogram levels off to a nearly constant value, also called the sill, as the separation distance increases, indicating a decrease in spatial dependence with distance within the range and no spatial correlation outside the range, which is reflected in a nearly constant variance. If there is a discontinuity or a vertical jump at the origin, the process has a nugget effect, which is often due to a measurement error, but may also be indicative of a spatially discontinuous process. The empirical semivariogram is an exploratory tool useful for assessing whether data exhibit spatial correlation. Furthermore, it can be compared to a Monte Carlo envelope of empirical semivariograms calculated from random permutations of the data while keeping the locations fixed (Diggle et al., 2003). If the empirical semivariogram lies outside the Monte Carlo envelope with increasing distance, this is an indication of spatial correlation.
An example of a semivariogram is shown in Figure 2.3. This graph is taken from "Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny" by Paula Moraga (Moraga, 2019).

2.5.4 Gaussian Markov Random Fields

2.5.4.1 Definition of GMRFs

Let $\boldsymbol{x} = (x_1, ..., x_n)^T$ be normally distributed with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, ..., \}$ and $\mathcal{E}$ be such that there is no edge between nodes i and j exactly when $x_i \perp x_j | \boldsymbol{x}_{ij}$. Then $\boldsymbol{x}$ is a *Gaussian Markov random field* (GMRF) with respect to $\mathcal{G}$.
Since $\boldsymbol{\mu}$ does not affect the pairwise conditional independence properties of $\boldsymbol{x}$, this information is 'hidden' in $\boldsymbol{\Sigma}$. Hence,

$$x_i \perp x_j | x_{ij} \iff Q_{ij} = 0.$$

Therefore, the non-zero pattern of $\boldsymbol{Q}$ determines $\mathcal{G}$, i.e. whether x_i and x_j are conditionally independent, and can be derived from $\boldsymbol{Q}$. If $\boldsymbol{Q}$ is a fully dense matrix, then $\mathcal{G}$ is fully connected, implying that any normal distribution with SPD covariance matrix is a GMRF and vice versa.
The elements of $\boldsymbol{Q}$ are used for conditional interpretations. For any GMRF with respect to $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$,

$$\mathbb{E}\left[x_i | \boldsymbol{x}_{-i}\right] = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j\sim i} Q_{ij} \left(x_j - \mu_j\right), \tag{2.29}$$

$$\text{Prec}\left(x_i | \boldsymbol{x}_{-i}\right) = Q_{ii} \qquad \text{and} \tag{2.30}$$

$$\text{Corr}\left(x_i, x_j | \boldsymbol{x}_{ij}\right) = -\frac{Q_{ij}}{\sqrt{Q_{ii} Q_{jj}}}, \qquad i \neq j \tag{2.31}$$

(Rue & Held, 2005, p. 21).
On the main diagonal of $\boldsymbol{Q}$ are the conditional precisions of x_i given $\boldsymbol{x}_{-i}$ are placed, while the other elements, when scaled appropriately, provide information about the conditional correlation between x_i and x_j given $\boldsymbol{x}_{ij}$. Since $\text{Var}\left(x_i\right) = \Sigma_{ii}$ and $\text{Corr}\left(x_i, x_j\right) = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$, the information about the marginal variance of x_i and the marginal correlation between x_i and x_j is given by $\boldsymbol{\Sigma}$. The marginal interpretation provided by the correlation matrix is intuitive and informative, as the scope of the interpretation is reduced from an n-dimensional distribution to a one- or two-dimensional distribution. $\boldsymbol{Q}$ is difficult to interpret marginally because either $\boldsymbol{x}_{-i}$ or $\boldsymbol{x}_{ij}$ would have to be integrated out of the joint distribution parameterized with respect to $\boldsymbol{Q}$. $\boldsymbol{Q}^{-1} = \boldsymbol{\Sigma}$ by definition, and in general Σ_{ii} depends on each element in $\boldsymbol{Q}$ and vice versa (Rue & Held, 2005, pp. 20–23).

2.5.4.2 Markov Properties of GMRFs

One property of GMRFs is that more information regarding conditional independence can be extracted from $\mathcal{G}$. The following three properties are equivalent.
The *pairwise Markov property*:

$$x_i \perp x_j | \boldsymbol{x}_{ij} \qquad \text{if } \{i,j\} \notin \mathcal{E} \text{ and } i \neq j.$$

The *local Markov property*:

$$x_i \perp \boldsymbol{x}_{-\{i,\text{ne}(i)\}} | \boldsymbol{x}_{\text{ne}(i)} \qquad \forall i \in \mathcal{V}.$$

The *global Markov property*:

$$\boldsymbol{x}_A \perp \boldsymbol{x}_B | \boldsymbol{x}_C$$

for all disjoint sets A, B and C where A and B are non-empty and separated by C. Illustrations for these properties are shown in Figure 2.4, Figure 2.5 and Figure 2.6. These illustrations are taken from Rue and Held (Rue & Held, 2005, pp. 23–24).

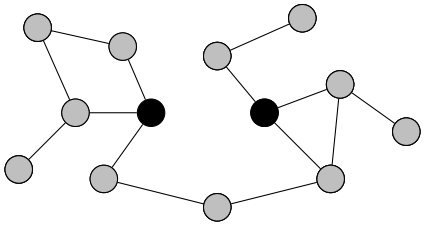

Fig. 2.4: The pairwise Markov property; the black nodes are conditionally independent given the light gray nodes.

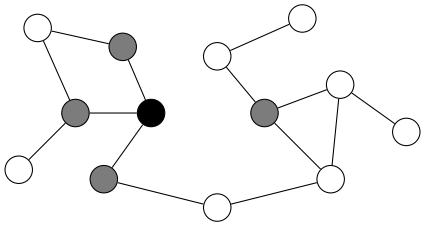

Fig. 2.5: The local Markov property; the black nodes and white nodes are conditionally independent given the dark gray nodes.

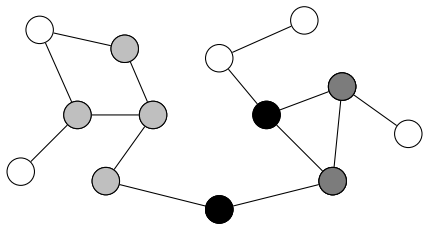

Fig. 2.6: The global Markov property; the dark gray and light gray nodes are globally independent given the black nodes.

2.5.4.3 Conditional Properties of GMRFs

An essential result of GMRFs is the conditional distribution for a subset $\boldsymbol{x}_a$ given $\boldsymbol{x}_{-A}$. Here the canonical parameterisation proves useful, since by definition it can be easily updated by successive conditioning.
By splitting the indices into the non-empty sets A and B, of which the latter is equal to -A,

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_A \\ \boldsymbol{x}_B \end{pmatrix}. \tag{2.32}$$

The mean and the precision are divided accordingly,

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \qquad \text{and} \qquad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AB} \\ \boldsymbol{Q}_{BA} & \boldsymbol{Q}_{BB} \end{pmatrix}. \tag{2.33}$$

The conditional distribution of $\boldsymbol{x}_A|\boldsymbol{x}_B$ is then a GMRF with respect to the subgraph $\mathcal{G}^A$ with mean $\boldsymbol{\mu}_{A|B}$ and precision matrix $\boldsymbol{Q}_{A|B} > 0$, where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AB}\left(\boldsymbol{x}_B - \boldsymbol{\mu}_B\right) \tag{2.34}$$

and

$$\boldsymbol{Q}_{A|B} = \boldsymbol{Q}_{AA}.$$

Thus, the explicit knowledge of $\boldsymbol{Q}_{A|B}$ is available through $\boldsymbol{Q}_{AA}$, i.e. no calculation is required to obtain the conditional precision matrix. Moreover, the conditional mean depends only on the values of $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ in $A \cup \text{ne}\left(A\right)$, since $Q_{ij} = 0 \, \forall j \not\in \text{ne}\left(i\right)$.
For successive conditioning, the canonical parameterisation for GMRF is useful.
A GMRF $\boldsymbol{x}$ with respect to $\mathcal{G}$ and canonical parameters $\boldsymbol{b}$ and $\boldsymbol{Q} > 0$ has the density

$$\pi\left(\boldsymbol{x}\right) \propto \exp\left(-1\frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x}\right).$$

The precision matrix is $\boldsymbol{Q}$ and the mean is $\boldsymbol{\mu} = \boldsymbol{Q}^{-1}\boldsymbol{b}$. The canonical parameterisation is written as

$$\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \boldsymbol{Q}).$$

Furthermore,

$$\mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{Q}^{-1}\right) \Longleftrightarrow \mathcal{N}_C(\boldsymbol{Q}\boldsymbol{\mu}, \boldsymbol{Q}).$$

If the indices are partitioned into two non-empty sets A and B and $\boldsymbol{x}$, $\boldsymbol{b}$ and $\boldsymbol{Q}$ are partitioned as in (2.32) and (2.33), then

$$\boldsymbol{x}_A|\boldsymbol{x}_B \sim \mathcal{N}_C(\boldsymbol{b}_A - \boldsymbol{Q}_{AB}\boldsymbol{x}_B, \boldsymbol{Q}_{AA}). \tag{2.35}$$

Let $\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{P}^{-1})$ and $\boldsymbol{x} \sim \mathcal{N}_C(\boldsymbol{b}, \boldsymbol{Q})$, then

$$\boldsymbol{x}|\boldsymbol{y} \sim \mathcal{N}_C(\boldsymbol{b} + \boldsymbol{P}\boldsymbol{y}, \boldsymbol{Q} + \boldsymbol{P}). \tag{2.36}$$

This allows the calculation of conditional densities with multiple sources of conditioning, e.g. conditioning on observed data and a subset of variables. Therefore, the canonical parameterisation can be repeatedly updated without explicitly calculating the mean until it is actually needed. The computation of the mean requires the solution of $\boldsymbol{Q}\boldsymbol{\mu} = \boldsymbol{b}$, but only matrix-vector products are needed for updating the canonical parameterisation (Rue & Held, 2005, pp. 25–27).

2.5.4.4 Specification Through Full Conditionals

Alternatively, a GMRF can be specified by the full conditionals $\{\pi(x_i|\boldsymbol{x}_{-i})\}$ in place of $\boldsymbol{\mu}$ and $\boldsymbol{Q}$. Suppose the full conditionals are given as normals with

$$\mathbb{E}[x_i|\boldsymbol{x}_{-i}] = \mu_i - \sum_{j:j\sim i} \beta_{ij}(x_j - \mu_j) \qquad \text{and} \tag{2.37}$$

$$\text{Prec}(x_i|\boldsymbol{x}_{-i}) = \kappa_i > 0 \tag{2.38}$$

for $i = 1, ..., n$, for $\boldsymbol{\mu}$, $\boldsymbol{\kappa}$ and some $\{\eta_{ij}, i \neq j\}$. Evidently, $\sim$ is implicitly defined by the non-zero terms of $\{\beta_{ij}\}$. For there to exist a joint density $\pi(\boldsymbol{x})$ leading to these full conditional distributions, these full conditionals must be consistent. Since $\sim$ is symmetric, it follows that if $\beta_{ij} \neq 0$, then $\beta_{ji} \neq 0$. If the entries of the precision matrix are chosen such that

$$Q_{ii} = \kappa_i, \qquad \text{and} \qquad Q_{ij} = \kappa_i\beta_{ij}$$

and $\boldsymbol{Q}$ must be symmetrical, i.e.,

$$\kappa_i \beta_{ij} = \kappa_j \beta_{ji},$$

then $\boldsymbol{x}$ is a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} = (Q_{ij})$ (Rue & Held, 2005, p. 27).

2.5.4.5 Multivariate GMRFs

A *multivariate GMRF* (MGMRF) is a multivariate extension of a GMRF that has proven useful in applications. Let $\boldsymbol{x}$ be a GMRF with respect to $\mathcal{G}$, then the Markov property implies that

$$\pi\left(x_i | \boldsymbol{x}_{-i}\right) = \pi\left(x_i | \left\{x_j : j \sim i\right\}\right).$$

x_i is the value related to node i. Often the nodes have physical interpretations such as an administrative region of a country, which can be used to define the neighbours of node i. Let each of the n nodes have an associated vector $\boldsymbol{x}_i$ of dimension p, resulting in a GMRF of size np. Such a GMRF is denoted by $\boldsymbol{x} = \left(\boldsymbol{x}_1^T, ..., \boldsymbol{x}_n^T\right)^T$. The Markov property with respect to the nodes is preserved, i.e.,

$$\pi\left(\boldsymbol{x}_i | \boldsymbol{x}_{-i}\right) = \pi\left(\boldsymbol{x}_i | \left\{\boldsymbol{x}_j : j \sim i\right\}\right),$$

where $\sim$ is with respect to *the same graph* $\mathcal{G}$. Let $\boldsymbol{\mu} = \left(\boldsymbol{\mu}_1^T, ..., \boldsymbol{\mu}_n^T\right)^T$ be the mean of $\boldsymbol{x}$, where $\mathbb{E}\left[\boldsymbol{x}_i\right] = \boldsymbol{\mu}_i$, and $\widetilde{\boldsymbol{Q}} = \left(\widetilde{\boldsymbol{Q}}_{ij}\right)$ its precision matrix, where each element of the matrix is a $p \times p$ matrix.
It follows that

$$\boldsymbol{x}_i \perp \boldsymbol{x}_j | \boldsymbol{x}_{-ij} \Longleftrightarrow \widetilde{\boldsymbol{Q}}_{ij} = \boldsymbol{0}.$$

Formally, a random vector $\boldsymbol{x} = \left(\boldsymbol{x}_1^T, ..., \boldsymbol{x}_n^T\right)^T$ with $\dim\left(\boldsymbol{x}_i\right) = p$, is called a MGMRF$_p$ with respect to $\mathcal{G} = \left(\mathcal{V} = \{1, ..., n\}, \mathcal{E}\right)$ with mean $\boldsymbol{\mu}$ and precision matrix $\widetilde{\boldsymbol{Q}} > 0$, exactly when its density has the form

$$\begin{aligned}
\pi\left(\boldsymbol{x}\right) &= \left(\frac{1}{2\pi}\right)^{np/2} \left|\widetilde{\boldsymbol{Q}}\right|^{1/2} \exp\left(-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)^T \widetilde{\boldsymbol{Q}}\left(\boldsymbol{x} - \boldsymbol{\mu}\right)\right) \\
&= \left(\frac{1}{2}\right)^{np/2} \left|\widetilde{\boldsymbol{Q}}\right|^{1/2} \exp\left(-\frac{1}{2}\sum_{ij}\left(\boldsymbol{x}_i - \boldsymbol{\mu}_i\right)^T \widetilde{\boldsymbol{Q}}_{ij}\left(\boldsymbol{x}_j - \boldsymbol{\mu}_j\right)\right)
\end{aligned}$$

and

$$\widetilde{\boldsymbol{Q}}_{ij} \neq 0 \Longleftrightarrow \{i, j\} \in \mathcal{E} \, \forall \, i \neq j.$$

A MGMRF$_p$ is equivalent to a GMRF of dimension np with identical mean vector and precision matrix. Therefore, all results valid for a GMRF are also valid for a MGMRF$_p$, with modifications, since the graph for a MGMRF$_p$ has size n and is defined with respect to $\{\boldsymbol{x}_i\}$, while for a GMRF it has size np and is defined with respect to $\{x_i\}$.
The interpretation of $\widetilde{\boldsymbol{Q}}_{ii}$ and $\widetilde{\boldsymbol{Q}}_{ij}$ can be derived from the full conditional $\pi(\boldsymbol{x}_i|\boldsymbol{x}_{-i})$. The extensions of (2.29) and (2.30) are

$$\mathbb{E}[\boldsymbol{x}_i|\boldsymbol{x}_{-i}] = \boldsymbol{\mu}_i - \widetilde{\boldsymbol{Q}}_{ii}^{-1} \sum_{j:j\sim i} \widetilde{\boldsymbol{Q}}_{ij}(\boldsymbol{x}_j - \boldsymbol{\mu}_j) \tag{2.39}$$

$$\text{Prec}(\boldsymbol{x}_i|\boldsymbol{x}_{-i}) = \widetilde{\boldsymbol{Q}}_{ii}. \tag{2.40}$$

In some applications, the full conditionals

$$\mathbb{E}[\boldsymbol{x}_i|\boldsymbol{x}_{-i}] = \boldsymbol{\mu}_i - \sum_{j:j\sim i} \boldsymbol{\beta}_{ij}(\boldsymbol{x}_j - \boldsymbol{\mu}_j) \tag{2.41}$$

$$\text{Prec}(\boldsymbol{x}_i|\boldsymbol{x}_{-i}) = \boldsymbol{\kappa}_i > 0, \tag{2.42}$$

In some applications, the full conditionals are used to define the MGMRF$_p$, for given $p \times p$-matrices $\{\boldsymbol{\beta}_{ij}, i \neq j\}$, $\{\boldsymbol{\kappa}_i\}$, and vectors $\boldsymbol{\mu}_i$. Again, $\sim$ is implicitly defined by the non-zero matrices $\{\boldsymbol{\kappa}_i\}$. Similar requirements as for $p = 1$ apply to the existence of the joint density: $\boldsymbol{\kappa}_i\boldsymbol{\beta}_{ij} = \boldsymbol{\beta}_{ij}^T\boldsymbol{\kappa}_j$ for $i \neq j$ and $\widetilde{\boldsymbol{Q}} > 0$. The $p \times p$ elements of $\widetilde{\boldsymbol{Q}}$ are

$$\widetilde{\boldsymbol{Q}}_{ij} = \begin{cases} \boldsymbol{\kappa}_i\boldsymbol{\beta}_{ij} & i \neq j \\ \boldsymbol{\kappa}_i & i = j \end{cases};$$

therefore $\widetilde{\boldsymbol{Q}} > 0 \iff (\boldsymbol{I} + (\boldsymbol{\beta}_{ij})) > 0$ (Rue & Held, 2005, pp. 29–30).

2.5.5 Integrated Nested Laplace Approximation

An alternative to MCMC methods that is both less computationally intensive and suitable for performing approximate Bayesian inference in latent Gaussian models is *Integrated nested Laplace Approximation* (INLA). The basis of INLA is the use of a combination of analytical approximations and numerical algorithms for sparse matrices to approximate the posterior distribution using closed-form expressions. This speeds up inference and circumvents problems of sample convergence and mixing, making it suitable for fitting large data sets or exploring other models (Rue

et al., 2009).
INLA can be used for all models of the following form,

$$\begin{aligned} y_i|\boldsymbol{x},\boldsymbol{\theta} &\sim \pi\left(y_i|x_i,\boldsymbol{\theta}\right), \qquad i=1,...,n, \\ \boldsymbol{x}|\boldsymbol{\theta} &\sim \mathcal{N}\left(\mu\left(\boldsymbol{\theta}\right),\boldsymbol{Q}\left(\boldsymbol{\theta}\right)^{-1}\right), \\ \boldsymbol{\theta} &\sim \pi\left(\boldsymbol{\theta}\right). \end{aligned}$$

As introduced in subsection 2.5.1, $\boldsymbol{y}$ are the observed data, $\boldsymbol{x}$ is a Gaussian field, $\boldsymbol{\theta}$ represents the hyperparameters, while $\mu\left(\boldsymbol{\theta}\right)$ and $\boldsymbol{Q}\left(\boldsymbol{\theta}\right)$ denote the mean and precision matrix respectively. To ensure fast inference, the dimension of the hyperparameter vector $\boldsymbol{\theta}$ should be small, since the approximations are computed by numerical integration over the hyperparameter space.
In most cases, the observations y_i are assumed to belong to the exponential family with mean $\mu_i = g^{-1}\left(\eta_i\right)$. As shown in equation (2.22), η_i accounts for the effects of several covariates in an additive way, which makes it suitable for a wide range of models, including spatial and spatio-temporal models, since $\left\{f^{(j)}\right\}$ can take very different forms.
Let $\boldsymbol{x} = \left(\alpha, \left\{\beta_k\right\}|\theta \sim \mathcal{N}\left(\mu\left(\theta\right),\boldsymbol{Q}\left(\theta\right)^{-1}\right)\right)$ be the vector of latent Gaussian variables, and let $\boldsymbol{\theta}$ be the vector of hyperparameters, which are not required to be Gaussian. INLA calculates accurate and fast approximations for the posterior marginals of the components of the latent Gaussian variables

$$\pi\left(x_i|\boldsymbol{y}\right), \qquad i=1,...,n,$$

as well as the posterior marginals for the hyperparameters of the latent Gaussian model

$$\pi\left(\theta_j|\boldsymbol{y}\right), \qquad j=1,...,\dim\left(\boldsymbol{\theta}\right).$$

For each element x_i of $\boldsymbol{x}$ the posterior marginals are given by

$$\pi\left(x_i|\boldsymbol{y}\right) = \int \pi\left(x_i|\boldsymbol{\theta},\boldsymbol{y}\right)\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)d\boldsymbol{\theta}, \tag{2.43}$$

and the posterior marginal for the hyperparameters can be expressed by

$$\pi\left(\theta_j|\boldsymbol{y}\right) = \int \pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)d\boldsymbol{\theta}_{-j}. \tag{2.44}$$

$\pi\left(x_i|\boldsymbol{y}\right)$ is approximated by combining analytical approximations to the full conditionals $\pi\left(x_i|\boldsymbol{\theta},\boldsymbol{y}\right)$ and $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ and numerical integration routines to integrate out $\boldsymbol{\theta}$. Similarly, $\pi\left(\theta_j|\boldsymbol{y}\right)$ is approximated by approximating $\pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ and integrating out $\boldsymbol{\theta}_{-j}$. In particular, the posterior density of $\boldsymbol{\theta}$ is obtained through Gaussian approximation

for the posterior of the latent field, $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$, evaluated at the posterior mode, $x^*(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{x}} \pi_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$,

$$\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left.\frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})}\right|_{\boldsymbol{x}=x^*(\boldsymbol{\theta})}. \tag{2.45}$$

Next, the following nested approximations are constructed,

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \int \widetilde{\pi}(x_i|\boldsymbol{\theta},\boldsymbol{y})\,\widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\,d\boldsymbol{\theta}, \qquad \widetilde{\pi}(\theta_j|\boldsymbol{y}) = \int \widetilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\,d\boldsymbol{\theta}_{-j}. \tag{2.46}$$

Finally, these approximations are numerically integrated with respect to $\boldsymbol{\theta}$

$$\widetilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \widetilde{\pi}(x_i|\theta_k,\boldsymbol{y})\,\widetilde{\pi}(\theta_k|\boldsymbol{y}) \times \Delta_k, \tag{2.47}$$

$$\widetilde{\pi}(\theta_j|\boldsymbol{y}) = \sum_l \widetilde{\pi}(\theta_l^*|\boldsymbol{y}) \times \Delta_l^*, \tag{2.48}$$

with Δ_k and Δ_l^* representing the area weights corresponding to θ_k and θ_l^*.
To obtain the approximations for the posterior marginals for the x_i's conditioned on selected values of θ_k and $\widetilde{\pi}(x_i|\theta_k,\boldsymbol{y})$, a Gaussian, Laplace or simplified Laplace approximation can be used. Using a Gaussian approximation derived from $\widetilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ is the simplest and fastest solution, but in some situations it produces errors in the location and is unable to capture skewness behaviour. Therefore, the Laplace approximation is favoured over the Gaussian approximation, although it is relatively expensive. The simplified Laplace approximation is associated with lower costs and addresses inaccuracies of the Gaussian approximation in terms of location and skewness in a satisfactory manner (Moraga, 2019).

2.6 Bayesian Spatial Models

Bayesian spatial models are often used in the field of disease mapping. Bayesian hierarchical models improve estimates of log risk by providing information about neighbouring regions in the spatially structured component as well as regional variation in the unstructured component (Blangiardo & Cameletti, 2015). One of the most well-known spatial models is Besags' spatial model, which is presented in Section 2.6.1. Several models have been developed based on the Besag model, including the Besag-York-Mollié (BYM) model, introduced in Section 2.6.2, the Leroux model, introduced in Section 2.6.3, and more recently the BYM2 model, introduced in Section 2.6.4.
In general, it can be assumed that areas in close proximity to each other have a more frequent burden of disease than areas that are further away from each other. By setting up a neighbourhood structure, this "proximity" can be defined. It is assumed that i and j are neighbours if they share a common boundary, denoted $i \sim j$. The set of neighbours of region i is denoted by δ_i and its size is given by n_{δ_i}.

2.6.1 Besag Spatial Models

2.6.1.1 Besags' Improper Spatial Model

A commonly used approach to modelling spatial correlation is the Besag model, also known as an intrinsic GMRF model. The conditional distribution for a random vector $\boldsymbol{x} = (x_1, ..., x_n)^T$ is given by

$$x_i|\boldsymbol{x}_{-i}, \tau_x \sim \mathcal{N}\left(\frac{1}{n_{\delta_i}} \sum_{j \in \delta_i} x_j, \frac{1}{n_{\delta_i}\tau_x}\right), \tag{2.49}$$

with τ_x as a precision parameter. The mean of the effects over all neighbours is given by the mean of x_i, while the precision is proportional to the number of neighbours. The joint distribution for $\boldsymbol{x}$ is given by

$$\pi\left(\boldsymbol{x}|\tau_x\right) \propto \exp\left(-\frac{\tau_x}{2} \sum_{i \sim j} \left(x_i - x_j\right)^2\right) \propto \exp\left(-\frac{\tau_x}{2} \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x}\right). \tag{2.50}$$

The precision matrix $\boldsymbol{Q}$ is given by

$$Q_{ij} = \begin{cases} n_{\delta_i} & i = j, \\ -1 & i \sim j, \\ 0 & \text{else.} \end{cases} \tag{2.51}$$

Q is a singular matrix, i.e. it has a non-empty null space $\boldsymbol{V}$, hence the model is called intrinsic or Besags' improper spatial model.
The Besag model for spatial effects has one hyperparameter, the precision τ_x, which is represented as

$$\theta_1 = \log \tau_x. \tag{2.52}$$

The prior is defined on θ_1 (Besag, 1974; Riebler et al., 2016).

2.6.1.2 Besags' Proper Spatial Model

To overcome this impropriety, the precision matrix has to be redefined as follows,

$$Q_{ij} = \begin{cases} \tau_x \left(n_{\delta_i} + d\right) & i = j, \\ -\tau & \text{else.} \end{cases} \tag{2.53}$$

$d > 0$ is an additional term added to the diagonal to control the "properness". The conditional distribution for the proper version of the Besag model is then given by

$$x_i | \boldsymbol{x}_{-i}, \tau_x, d \sim \mathcal{N}\left(\frac{1}{d + n_{\delta_i}} \sum_{i \sim j} x_j \frac{1}{\tau_x \left(d + n_{\delta_i}\right)}\right). \tag{2.54}$$

The proper version of the Besag model for spatial effects has two hyperparameters, the precision τ_x, which is represented as

$$\theta_1 = \log \tau_x \tag{2.55}$$

and the diagonal parameter d, which is represented as

$$\theta_2 = \log d. \tag{2.56}$$

The priors are defined on θ_1 and θ_2, respectively (Besag, 1974; Riebler et al., 2016).

2.6.2 The Besag-York-Mollié Model

The Besag-York-Mollié (BYM) model is a lognormal Poisson model that is a combination of a Besag model u and an ordinary random effect component v for non-spatial heterogeneity. It combines the regional spatial effect $\boldsymbol{x}$ into the sum of an unstructured and a structured spatial component, so that $\boldsymbol{x} = \boldsymbol{v} + \boldsymbol{u}$.
$\boldsymbol{v} \sim \mathcal{N}\left(0, \tau_v^{-1}\boldsymbol{I}\right)$ accounts for pure overdispersion, while $\boldsymbol{u} \sim \mathcal{N}\left(\boldsymbol{0}, \tau_u^{-1}\boldsymbol{Q}^{-}\right)$ is the Besag model.
By using a spatial and a non-spatial error term, the overdispersion that is not modelled by the Poisson variables is taken into account. Thus, if the observed variance is not fully explained by the spatial structure of the data, the error terms explain the rest of the variance.
The resulting covariance matrix of $\boldsymbol{x}$ is given by

$$\text{Var}\left(\boldsymbol{x}|\tau_u, \tau_v\right) = \tau_v^{-1}\boldsymbol{I} + \tau_u^{-1}\boldsymbol{Q}^{-}, \tag{2.57}$$

where $\boldsymbol{Q}^{-}$ denotes the generalised inverse of $\mathbf{Q}$.
The hyperparameters of the model are the precision τ_u of the Besag model u and the precision τ_v of the iid model v. They are represented as

$$\boldsymbol{\theta} = (\theta_1, \theta_2) = (\log \tau_v, \log \tau_u) \tag{2.58}$$

and the prior is defined on $\boldsymbol{\theta}$ (Besag et al., 1991; Riebler et al., 2016).

2.6.3 The Leroux Model

One problem with the BYM model is that the structured and unstructured components are not identifiable because they cannot be considered independently. Moreover, τ_v and τ_u do not represent variability at the same level, which makes the choice of hyperpriors difficult. The Leroux model is formulated in such a way that the compromise between the two variations is made more explicit. It is assumed that $\boldsymbol{x}$ follows a normal distribution with zero mean and covariance matrix

$$\text{Var}\left(\boldsymbol{x}|\tau_x, \phi\right) = \tau_x^{-1}\left((1-\phi)\,\boldsymbol{I} + \phi\boldsymbol{Q}\right)^{-1}, \tag{2.59}$$

with $\phi \in [0, 1]$ as mixing parameter. For $\phi = 0$ the model reduces to pure overdispersion and for $\phi = 1$ to the Besag model. The conditional expected value of x_i for all other random effects is the weighted mean of the unstructured model with

zero mean and the mean of the Besag model, while the conditional variance is the weighted mean of τ_x^{-1} and $(\tau_x \cdot n_{\delta_i})^{-1}$ (Leroux et al., 2000; Riebler et al., 2016).

2.6.4 The BYM2 Model

One problem that all the aforementioned models have is the lack of scaling of the spatially structured component. Scaling facilitates the assignment of hyperpriors and ensures that the interpretation of hyperpriors remains the same across different areas.
Another problem is that the marginal standard deviations of the commonly used IGMRF priors can vary greatly, a fact that should be taken into account by assigning hyperpriors to the precision parameters of these models.
Since the Besag model penalises a local deviation from its null space, the hyperprior will control this local deviation and thus affect the smoothness of the estimated spatial effects. If the estimate of the field is too smooth, the precision will be large and the spatial variation may be blurred. On the other hand, if the precision is too small, the model could overfit due to the large local variability.
The marginal variances $\tau_x^{-1} [\boldsymbol{Q}^-]_{ii}$ depend on the structure of the graph, which is reflected in the structure matrix $\boldsymbol{Q}$. A generalised variance can be calculated as the geometric mean of the marginal variance as follows

$$\sigma^2_{\text{GV}}(\boldsymbol{u}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{1}{\tau_x}[\boldsymbol{Q}^-]_{ii}\right)\right) = \frac{1}{\tau_x}\exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\left([\boldsymbol{Q}^-]_{ii}\right)\right). \quad (2.60)$$

In order to unify the interpretation of a chosen prior for τ_x and make it transferable across domains, the structured effect must be scaled such that $\sigma^2_{\text{GV}}(\boldsymbol{x}) = \tau_x^{-1}$. This implies that τ_x denotes the accuracy of the (marginal) deviation from a constant level, independent of the underlying graph.
A modification of the BYM model that addresses this scaling problem is the BYM2 model. It uses a scaled structured component $\boldsymbol{u}_*$, where $\boldsymbol{Q}_*$ denotes the precision matrix of the Besag model, scaled with the marginal variance σ^2_{GV} as a factor. The random effect is then given by

$$\boldsymbol{x} = \frac{1}{\tau_x}\left(\sqrt{1-\phi}\boldsymbol{v} + \sqrt{\phi}\boldsymbol{u}_*\right), \quad (2.61)$$

with covariance matrix

$$\text{Var}(\boldsymbol{x}|\tau_x, \phi) = \frac{1}{\tau_x}\left((1-\phi)\boldsymbol{I} + \phi\boldsymbol{Q}_*^-\right). \quad (2.62)$$

Equation 2.61 emphasises the trade-off between pure overdispersion and spatially structured correlation, where $0 \leq \phi \leq 1$ measures the fraction of the marginal variance explained by the structured effect. For $\phi = 0$ the model reduces to pure overdispersion, while for $\phi = 1$ it becomes a Besag model (Martins et al., 2014; Riebler et al., 2016).

2.7 Goodness-of-Fit indicators

The goodness of fit indicates "how well" an estimated model can explain a set of observations. Measures of goodness of fit allow a statement to be made about the discrepancy between the theoretical values of the random variables under investigation, which are expected or predicted on the basis of the model, and the values actually measured.
The goodness of fit of a model to available data can be assessed with the help of statistical tests or suitable ratios.

2.7.1 The Akaike Information Criterion

The historically oldest criterion was proposed in 1973 by Hirotsugu Akaike (1927-2009) as an information criterion and is known today as the Akaike information criterion (AIC). The AIC is one of the most frequently used criteria for model selection in the context of likelihood-based inference.
Let the population contain the distribution of a variable with unknown density function p. The maximum likelihood estimation assumes a known distribution with an unknown parameter θ, hence the density function can be written as $q(\theta)$. The Kullback-Leibler divergence is used as a distance measure between p and $q\left(\widehat{\theta}\right)$ with $\widehat{\theta}$ the estimated parameter from the maximum likelihood estimation. The better the maximum likelihood model, the small the Kullback-Leibler divergence $D(P||Q)$.
For a maximum likelihood model with a p-dimensional parameter vector $\widehat{\boldsymbol{\theta}}$, the Akaike information criterion is defined as

$$\text{AIC} = -2l\left(\widehat{\boldsymbol{\theta}}_{ML}\right) + 2p, \tag{2.63}$$

with l the log-likelihood function (Akaike, 1974).

2.7.2 The Deviance Information Criterion

In statistics, the deviance information criterion, or DIC for short, is a measure (criterion) for the prediction error of a model. This measure is an information criterion and belongs to the environment of the Bayesian method for model comparisons. The smaller the deviance information criterion, the better the model fit. The deviance information criterion can be regarded as the Bayesian equivalent of the Akaike

information criterion.
The deviance is defined as

$$D(\boldsymbol{\theta}) = -2\log\left(l\left(\boldsymbol{y}|\boldsymbol{\theta}\right)\right) + C, \tag{2.64}$$

with $\boldsymbol{y}$ the data, $\boldsymbol{\theta}$ the unknown parameters of the model and l the likelihood function. C is a constant that cancels out in all calculations that compare different models and therefore it does not need to be known (Nelder & Wedderburn, 1972).
The DIC is given by

$$\text{DIC} = D\left(\bar{\boldsymbol{\theta}}\right) + 2p_D, \tag{2.65}$$

with

$$p_D = \overline{D\left(\boldsymbol{\theta}\right)} - D\left(\bar{\boldsymbol{\theta}}\right), \tag{2.66}$$

where $\bar{\boldsymbol{\theta}}$ is the expected value of $\boldsymbol{\theta}$ (Spiegelhalter et al., 2014).

2.7.3 The Watanabe-Akaike Information Criterion

The Watanabe-Akaike information criterion (WAIC) is the generalised AIC onto singular statistical models.
The WAIC is given by

$$\text{WAIC} = -2\text{LLPD} + 2p_{\text{WAIC}}, \tag{2.67}$$

with the log pointwise predictive density (LLPD) given by

$$\text{LLPD} = \sum_{i=1}^{n} \log\left(\int \pi\left(y_i|\boldsymbol{\theta}\right)\pi_{\text{post}}\left(\boldsymbol{\theta}\right)\right) d\boldsymbol{\theta}. \tag{2.68}$$

LLPD can be seen as the Bayesian analogue of $l\left(\widehat{\boldsymbol{\theta}}_{ML}\right)$ in the calculation of the AIC.
The penalty term of the WAIC is also fully Bayesian and given by

$$p_{\text{WAIC}} = \sum_{i=1}^{n} \text{Var}_{\text{post}}\left(\log\left(\pi\left(y_i|\boldsymbol{\theta}\right)\right)\right), \tag{2.69}$$

where the term represents the variance of the individual terms in the LLPD over all data points (Watanabe & Opper, 2010; Yong, 2018).

2.7.4 The Conditional Predictive Ordinate

The conditional predictive ordinate (CPO) is a Bayesian diagnostic that can be used to detect surprising observations. It is often used in the context of univariate sampling, the multivariate normal distribution and regression models.
The conditional predictive ordinate is given by

$$\text{CPO} = \pi\left(y_i|\boldsymbol{y}_{-i}\right) \tag{2.70}$$

with $\boldsymbol{y}$ the data, $\boldsymbol{y}_{-i}$ the data without the i-th observation, and $\pi\left(\cdot|\boldsymbol{y}_{-i}\right)$ the predictive distribution of a new observation at $\boldsymbol{y}_{-i}$. Low values of CPO are an indication that y_i is surprising given prior knowledge and the other observations (Cox, 1980; Pettit, 1990).

2.8 Model Issues

One problem that plagues these models is that they cannot be directly compared due to their different parameterisations and the fact that the precision in these models is interpreted differently. Since neither a Besag model nor a BYM model nor a Leroux model is scaled, the precision parameter is not representative of the marginal precision but is confounded with the mixing parameter. Therefore, the effect of a prior assigned to the precision parameter is dependent on the graph structure of the application. Thus, a given prior is not transferable between different applications if the underlying graph changes. Furthermore, the goal of the BYM2 model is not to optimise goodness-of-fit indicators, but to provide a meaningful model formulation where all parameters have a clear meaning. By mapping the precision parameter to the marginal standard deviation, the model parameters are flexible and the assignment of meaningful hyperpriors is made easier (Riebler et al., 2016).

Additionally, the goodness-of-fit indicators introduced in Section 2.7 have their own problems. The DIC, for example, produces unreasonable results if the posterior distribution is not well summarised by its mean, while the WAIC is based on a data partition that would create difficulties for structured models, such as for spatial or network data (Gelman et al., 2014).

Finally, the choice of the prior also affects the value of these criteria and depending on the values chosen for the PC priors used in this work, overfitting of the models may occur, which is then reflected in these criteria, but more on this later.

2.9 The Variance Inflation Factor

The Variance Inflation Factor (VIF) is a measurement that can be used to avoid multicollinearity between covariates. The VIF quantifies the severity of multicollinearity in a generalised linear model. It provides an index that measures the extent to which the variance of an estimated regression coefficient is increased due to collinearity. For $p-1$ independent variables,

$$\text{VIF}_i = \frac{1}{1-R^2}, \qquad i = 1, ..., p-1, \tag{2.71}$$

with R^2 the coefficient of determination. In most literature, a value of at least 5 is suggested as too high and is therefore used as the threshold in this work (Craney & Surles, 2002).

3 Analysis of Geospatial Health Data

Healthcare data provides information for detecting public health problems and reacting adequately when they occur. With this information, prevention and control of a multitude of health conditions including infectious diseases, non-communicable diseases, injuries and health-related behaviours can be achieved. To analyse and interpret health data, the process involves a wide variety of system designs, analytical methods, modes of presentation and interpretive uses(Teutsch, Churchill, et al., 2000). Descriptive methods generally form the basis of routine reporting of surveillance data. Rather than focusing on observed patterns in the data, these may also attempt to compare the relative occurrence of health outcomes in different subgroups. More specific hypotheses can be explored using inferential methods. The aim of these methods is to draw statistical inferences about patterns or outcomes of health.

The increasing availability of geo-referenced health data, population data, satellite imagery of environmental factors influencing levels of disease activity, and the development of geographic information systems (GIS) and address geocoding software, the rise of studies of spatial and spatio-temporal variation in disease has been facilitated.

A broad range of spatial and spatio-temporal methods exist for disease surveillance, including methods for disease mapping, clustering and geographic correlation studies. These methods can be used to identify areas of high risk, risk factors, evaluate spatial variations in temporal trends, measure excess disease risk near a suspected source and detect outbreaks at an early stage.

55

3.1 Geographic Data

In spatial statistics, two fundamental types of geographic data exist, namely *vector data* and *raster data*. In the vector data model, the world is represented by points, lines and polygons with discrete, well-defined boundaries, which tends to result in high accuracy. Raster data, on the other hand, divides the surface into cells of uniform size, and raster datasets are used as the basis for background images in web mapping.
Determining which data type to use depends on the domain of the application. Vector data dominates in the social sciences because human settlements typically have discrete boundaries, while raster data are commonly used in many environmental sciences because they are based on remote sensing data. Naturally, there is also some overlap and both types can be used together or one form can be converted into the other (Lovelace et al., 2019).

3.1.1 Vector Data

The geographic vector data model is based on points located within a *coordinate reference system* (CRS), in which points either represent self-standing features or form more complex geometric shapes, i.e. lines and polygons. Using this system, Trondheim can be represented by the coordinates $(10.4, 63.4)$, meaning 10.4 degrees east of the prime meridian and 63.4 degrees north of the equator. It could also be written as $(1157722.70, 9199010.75)$, which is the position of Trondheim using the Web Mercator projection, the de facto standard for web mapping applications. More will be said about CRS later, but for now it is sufficient to know that it is possible to display coordinates in various ways. An example of a CRS is shown in Figure 3.1.

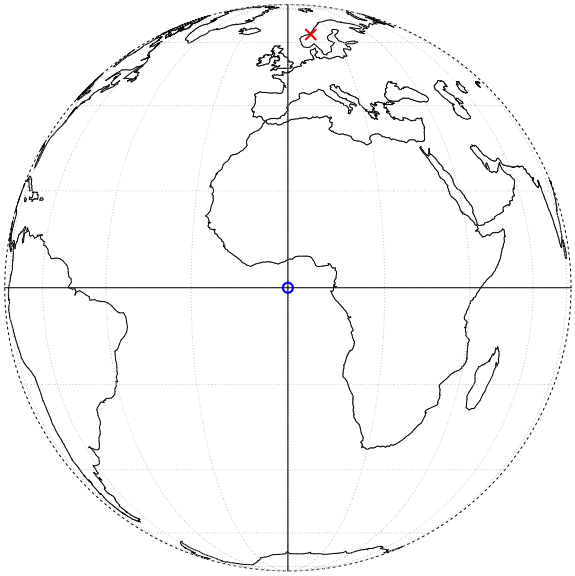

Fig. 3.1: A geographic CRS with an origin at 0° longitude and latitude. The red X denotes the location of Trondheim.

Different Types of Vector Data

As mentioned earlier, there are different types of vector data. There are 17 different geometry types in the standard *simple features*, but there are seven core types that can be used in most analysis software. These types are visualised in Figure 3.2.

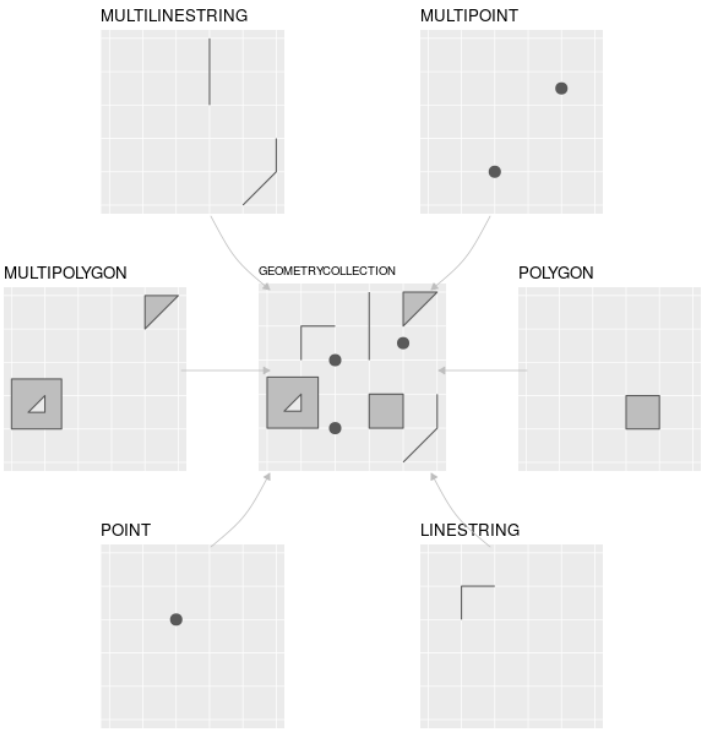

Fig. 3.2: The most commonly used simple feature types.

Simple Features was developed by the Open Geospatial Consortium and is an open, standardised, hierarchical data model that represents a wide range of geometry types. The use of this data model ensures that scientific work can be transferred to other institutions, e.g. when importing from and exporting to spatial databases (Lovelace et al., 2019).

3.1.2 Raster Data

The geographic raster data model consists in most cases of a raster header and a matrix representing uniformly distributed cells/pixels. The raster header defines the CRS, the origin (starting point) and the extent. Since the number of columns and rows and the resolution of the cell size are stored in the extent, starting from the origin, it is easy to access and change each cell by its ID or by specifying the row and column number. In this type of representation, the coordinates of the four vertices of each cell are not explicitly stored, instead only the origin is stored. This speeds up data processing and makes it more efficient, but each raster layer can only contain a single value, which can be either numeric or categorical. Typically, raster maps are

used to depict continuous features such as elevation or temperature, but categorical variables, for example soil or land cover, as shown in Figure 3.3 (Lovelace et al., 2019).

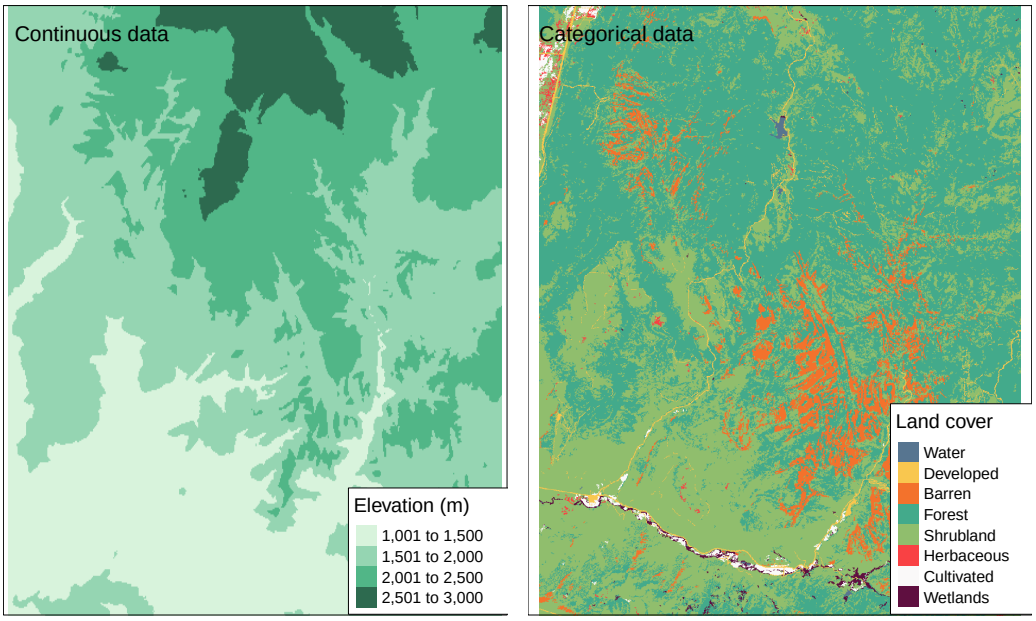

Fig. 3.3: An example of continuous and categorical raster data

Coordinate Reference Systems

A common denominator of vector and raster data are that both use the coordinate reference system (CRS), which defines how spatial elements relate to the surface of the Earth. The CRS can be either geographic or projected.

Geographic Coordinate Systems

Geographic coordinate systems use two values, *longitude* and *latitude*, to identify any location on Earth. Longitude is defined as the east-west location at an angular distance from the prime meridian plane, while latitude is the angular distance north or south of the equator. Consequently, distances in geographic CRS are not measured in metres.

The Earth's surface is typically represented in geographical coordinate systems by a spherical or ellipsoidal surface. The former assumes that the Earth is a perfect sphere of a certain radius, which has the advantage of being a simplistic model, but is associated with inaccuracies owing to the fact that the Earth is not a sphere. Ellipsoidal models are defined by the equatorial radius and the polar radius, providing a better model since the equatorial radius is approximately 11.5 km longer than the polar radius.
The *datum* is a broader component of CRS that contains information about which ellipsoid to use and the exact relationship between Cartesian coordinates and the location on the Earth's surface. The notation *proj4string* is used to store these additional details. It allows for local variations of the Earth's surface, such as large mountain ranges, to be taken into account in local CRS. Datum can again be divided into two categories, *local* and *geocentric*, the difference being that in the local datum the ellipsoidal surface is shifted to match the surface at a particular location, whereas in the geocentric datum the centre of gravity of the Earth is the centre and the accuracy of the projections is not optimised for any particular location (Lovelace et al., 2019).

Projected Coordinate Systems

Projected CRS are based on Cartesian coordinates on an implicitly flat surface and have an origin, x and y axes, and a linear unit of measurement, metres for instance. They are based on geographic CRS and rely on map projections to convert between the three-dimensional surface of the Earth and the east/north values (x and y) in a projected CRS.
This transition always entails some distortion, skewing some of the properties of the earth's surface, such as area, direction, distance and shape. Generally, the name of a projection is based on a property it preserves, e.g. equal area projection preserves area, equidistant projection preserves distance and conformal projection preserves local shape.
Again, subgroups exist in projection coordinate systems, *conic*, *cylindrical* and *planar* projections. In a conic projection, the earth's surface is projected onto a cone along one or two tangent lines. Along these lines the distortions are minimised and increase with the distance to the lines. The projection is therefore best suited for maps of mid-latitude areas. Cylindrical projections map the surface onto a cylinder. These types of projections can be created by touching the surface of the Earth along one or two tangent lines. They are often used to map the entire Earth. A planar projection projects data onto a flat surface that touches the globe at a point or along

a tangent line, and is typically used in mapping polar projections (Lovelace et al., 2019).

3.2 Modeling and Visualising Health Data

After collecting and cleaning all the data needed to analyse a research question, the next step is the analysis itself. This analysis may involve visualising the data at hand, for example by visualising the neighbourhood structure of spatial areas or simply by plotting the locations of all points of interest on a base map. When deciding whether to model spatial dependence, Moran's I can be used to construct a test for spatial correlation between different spatial entities. Another method that can be helpful in deciding whether to use a spatial model is the standardised incidence ratio (SIR), which relates the expected number of cases to the actual number of cases. These two methods are presented in Section 3.2.1.2 and Section 3.2.1.3. If there is a spatial correlation, it is useful to model this spatial effect using models for risk assessment in spatial areas. The way this is done is introduced in Section 3.2.1.4. If data are available over a longer period of time and some variables change their values over time or a temporal trend is present, it may be useful to model the spatial and temporal effect by using spatio-temporal models, which are introduced in Section 3.2.1.5.

3.2.1 Areal Data

Areal or lattice data are the result of segmenting a fixed domain into a finite number of sub-regions where results are aggregated, e.g. the number of infections with a specific disease in districts or the number of overweight people in provinces. Often the aim of disease risk models is to assess the risk within the same areas for which data are available. This can be done with a simple measure such as the *standardised incidence ratio* (SIR) or by using a Bayesian hierarchical model, which allows information to be drawn from neighbouring areas and incorporates covariates, thereby smoothing and reducing extreme values.

3.2.1.1 Spatial Neighbourhood Matrices

Spatial or proximity matrices are useful for exploratory analysis of area data. Let w_{ij} denote the (i, j) element of a *spatial neighbourhood matrix* $\boldsymbol{W}$. w_{ij} connects the two areas in some spatial way. The neighbourhood structure over the complete study region is defined by $\boldsymbol{W}$, and the elements of the matrix can be considered as

weights. The closer j is to i, the more weight is associated with it. The simplest neighbourhood definition is given by the binary matrix

$$w_{ij} = \begin{cases} 1 & \text{if regions } i \text{ and } j \text{ share a border} \\ 0 & \text{else} \end{cases} \tag{3.1}$$

Since a region cannot share a boundary with itself, $w_{ii} = 0$ (Moraga, 2019).
In Figure 3.4, the number of shared borders of each canton in Switzerland are mapped.

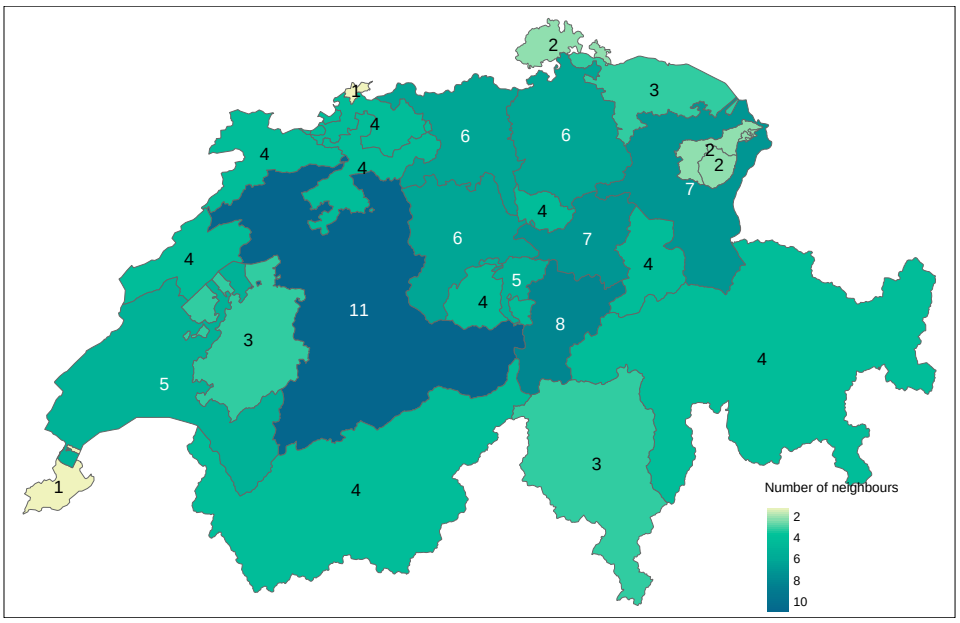

Fig. 3.4: The number of shared borders of cantons in Switzerland

3.2.1.2 Moran's I

Moran's I is a measure of spatial autocorrelation developed by Patrick Moran. Spatial autocorrelation is characterised by a correlation in a signal between close locations in space. Spatial autocorrelation is inherently more complex than one-dimensional autocorrelation due to the fact that spatial correlation is multidimensional (i.e. 2 or 3 spatial dimensions) and multi-directional. The formula for Moran's I is given by

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \overline{x}\right) \left(x_j - \overline{x}\right)}{\sum_{i=1}^{n}}, \tag{3.2}$$

with n denoting the number of spatial units indexed by i and j, x the parameter of interest, $\boldsymbol{w}$ a spatial neighbourhood matrix and W the sum of all w_{ij}.

Using Moran's I, a test for spatial autocorrelation can be constructed with the following hypotheses:

$$H_0: \text{ No spatial autocorrelation vs. } H_1: \text{ Spatial autocorrelation.} \tag{3.3}$$

Under H_0 the expected value is given by

$$\mathbb{E}[I] = \frac{-1}{n-1}. \tag{3.4}$$

As n approaches infinity, the expected value therefore approaches 0 (Moran, 1950).

3.2.1.3 Standardised Incidence Ratio

A basic measure of disease risk is the *standardised incidence ratio*, which yields an estimate in each of the areas that form a partition of the study region. It is defined as the ratio of observed counts to expected counts

$$\text{SIR}_i = \frac{Y_i}{E_i}. \tag{3.5}$$

E_i represents the sum of the expected number of cases of a given area i that behave according to the way the standard population behaves. It is calculated using indirect standardisation as

$$E_i = \sum_{j=1}^{m} r_j^{(s)} n_j^{(i)}, \tag{3.6}$$

with $r_j^{(s)}$ the rate in stratum j in the standard population and $n_j^{(i)}$ the population in stratum j of area i. If the stratum information is unavailable, the expected counts can be calculated as follows

$$E_i = r^{(s)} n^{(i)},$$

where $r^{(s)}$ denotes the rate in the standard population and $n^{(i)}$ is the population of area i. If the standardised incidence rate is greater than 1, area i has a higher risk than expected from the standard population, while for $\text{SIR}_i = 1$ the risk is the same and for $\text{SIR}_i < 1$ it is lower than expected. The ratio is also called the standardised mortality ratio when applied to mortality data (Moraga, 2019).

3.2.1.4 Spatial Small Area Disease Risk Estimation

While SIRs may prove useful in some situations, in areas with low population sizes or rare diseases, expected counts may be low, making SIRs insufficiently reliable for

reporting. It is therefore preferable to assess disease risk using models that allow information to be borrowed from neighbouring areas and incorporate information from covariates, thus smoothing or shrinking extreme values due to small sample sizes (Gelfand et al., 2010).
The observed counts Y_i in area i are typically modeled with a Poisson distribution with mean $E_i\theta_i$, where E_i is the expected counts and θ_i denotes the relative risk in area i. To account for extra Poisson reliability, the logarithm of the relative risk is expressed as the total of the intercept and the random effects. θ_i quantifies whether area i has a higher $(\theta_i > 1)$ or lower $(\theta_i < 1)$ risk than the average risk in the standard population. If the risk of an area i is half the average risk, then $\theta_i = 0.5$. The general model for spatial data is formulated as follows:

$$Y_i \sim \text{Po}\left(E_i\theta_i\right), \qquad i = 1, ..., n, \tag{3.7}$$

$$\log\left(\theta_i\right) = \alpha + u_i + v_i. \tag{3.8}$$

The overall risk in the region of study is represented by α, u_i is a random effect specific to each area to model the spatial dependence between relative risks, and v_i is an unstructured exchangeable component that models uncorrelated noise, $v_i \sim \mathcal{N}\left(0, \sigma_v^2\right)$. Covariates are often included to measure risk factors and other random effects to deal with different sources of variability. For example,

$$\log\left(\theta_i\right) = \boldsymbol{d}_i\boldsymbol{\beta} + u_i + v_i,$$

with $\boldsymbol{d}_i = (1, d_{i1}, ..., d_{ip})$ a vector of the intercept and p covariates corresponding to the area i and $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)^T$ the vector of coefficients. An increase in d_j $(j = 1, ..., p)$ by one unit, leads to an increase in the relative risk by a factor of $\exp\left(\beta_j\right)$, provided that all other covariates remain constant (Moraga, 2019).

3.2.1.5 Spatio-Temporal Small Area Disease Risk Estimation

When disease counts are monitored over time, spatio-temporal models are useful as they take into account not only the spatial structure but also temporal correlations and spatio-temporal interactions (Martínez-Beneito et al., 2008). Let Y_{ij} be the counts observed in area i and at time j, θ_{ij} be the relative risk, E_{ij} be the expected number of cases in area i and at time j, then

$$Y_{ij} \sim \text{Po}\left(E_{ij}\theta_{ij}\right), \qquad i = 1, ..., I,\ j = 1, ..., J. \tag{3.9}$$

$\log(\theta_{ij})$ is written as the sum of several components, including spatial and temporal structures, to consider that neighbouring areas and successive times may have similar risk. Spatio-temporal interactions can be included to account for the fact that temporal trends may differ from area to area but may be more alike in neighbouring areas.

Bernardinelli et al. (Bernardinelli et al., 1995), for example, propose a spatio-temporal model with parametric time trends that expresses the logarithm of relative risks as

$$\log(\theta_{ij}) = \alpha + u_i + v_i + (\beta + \delta_i) \times t_j. \tag{3.10}$$

The intercept is denoted by α, $u_i + v_i$ is a random area effect, β represents a global linear trend effect and δ_i is an interaction between space and time which is the difference between β and the area-specific trend. For modeling u_i and δ_i, a CAR distribution is used and v_i is i.i.d.. This specification allows each of the areas to have its individual time trend, where the spatial intercept is given by $\alpha + u_i + v_i$ and the slope by $\beta + \delta_i$. δ_i is referred to as the differential trend of the i-th area and represents the amount by which the time trend of area i deviates from the overall time trend β. If $\delta_i \neq 0$, then area i has a time trend with a slope that is either steeper or less steep than the overall time trend β.

For models that do not demand linearity of the time trend, non-parametric models such as the one proposed by Knorr-Held (Knorr-Held, 2000) can be used. This specific model incorporates spatial effects, temporal random effects and an interaction between space and time as follows:

$$\log(\theta_{ij}) = \alpha + u_i + v_i + \gamma_j + \phi_j + \delta_{ij}. \tag{3.11}$$

The intercept is again denoted by α, $u_i + v_i$ is a spatial random effect defined as before, i.e. u_i follows a CAR distribution and v_i is i.i.d.. $\gamma_j + \phi_j$ represents a temporal random effect and γ_j follows either a first order random walk in time (RW1)

$$\gamma_j | \gamma_{j-1} \sim \mathcal{N}\left(\gamma_{j-1}, \sigma_\gamma^2\right), \tag{3.12}$$

or second order random walk in time (RW2)

$$\gamma_j | \gamma_{j-1}, \gamma_{j-2} \sim \mathcal{N}\left(2\gamma_{j-1} - \gamma_{j-2}, \sigma_\gamma^2\right). \tag{3.13}$$

The unstructured temporal effect is given by $\phi_j \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_\phi^2\right)$. The interaction between space and time, δ_{ij}, can be specified in a number of ways by combining the structure of the random effects that interact. The interactions proposed by Knorr-Held are those between the effects (u_i, γ_j), (u_i, ϕ_j), (v_i, γ_j) and (v_i, ϕ_j) (Knorr-Held,

2000).
Using the last of these interactions leads to the assumption that there is no spatial or temporal structure on δ_{ij}. Thus, the interaction term can be modeled as $\delta_{ij} \sim \mathcal{N}\left(0, \sigma_\delta^2\right)$ (Moraga, 2019).

3.2.1.6 Issues With Areal Data

The analysis of spatially aggregated data is subject to the "misaligned data problem" (MIDP), which arises when the data to be analysed is at a different scale from that at which it was collected (Banerjee et al., 2014). This may be solely due to the fact that the aim is to obtain the spatial distribution of a variable at a new spatial level of aggregation, e.g. if predictions are to be made at the county level with data that was originally collected at the postcode level. Another objective may be to try to find an association between variables available at different spatial scales, e.g. determining whether the risk of an unfavourable outcome provided at the country level correlates with exposure to an environmental pollutant measured at different stations, taking into account the population at risk and other demographic information available at the postcode level.
The Modifiable Area Unit Problem (MAUP) (Openshaw, 1984) describes a problem where the inference may differ when the same underlying data are grouped at a new spatial level of aggregation. It consists of two interrelated effects, the first of which is the scale/aggregation effect. It relates to the different conclusions obtained when the same data are grouped into larger and larger areas. The other effect is the grouping/zoning effect, which accounts for the variability in results due to alternative formations of the areas, resulting in differences in area shape given the same or similar scales.
Ecological studies are defined by their reliance on aggregated data (Robinson, 2009) and the inherent potential for ecological fallacies. This phenomenon occurs when estimated associations obtained from the analysis of variables measured at the aggregate level lead to conclusions that differ from analyses based on the same variables measured at the individual level. This can be considered a special case of MAUP and the resulting so-called ecological bias is composed of two effects similar to the aggregation and zoning effects in MAUP. Namely, the aggregation bias caused by the aggregation of individuals and the specification bias due to the different distribution of confounding variables that results from the aggregation (Gotway & Young, 2002; Moraga, 2019).

4 Dataset Collection

As is often the case with statisticians, the construction of the dataset used to analyse a research question is an essential task and frequently involves the merging of multiple data sources to create a dataset. This was the case in this thesis and in the following chapter a brief overview of the data sources used, their pre-processing and how they were combined is given. All of the data that is used in this work is openly available, either through government sources or Github.

68

4.1 Covid-19 Data

4.1.1 Covid-19 Data for Norway

The Covid-19 data for Norway comes from a dataset made available to the public via the a repository on the website Github.com, created by the user thohan88. The repository contains a daily updated dataset that is the result of combining several data sources, which include the Institute of Public Health and the Norwegian Directorate of Health. According to the author of the repository, the project is "an open-source effort to make data about the Covid-19 situation in Norway available to the public in a timely and coherent manner" (Hansen, 2020).
A few sample data points from this dataset are displayed in Table 4.1.

Tab. 4.1: An excerpt from the Covid-19 data for Norway. Does not contain all variables. The number of infections are the cumulative number of infections.

Id	Municipality	Population	Inf. 2020-03-26	Inf. 2020-03-27
1103	Stavanger	143574	87	88
1507	Ålesund	66258	20	20
4601	Bergen	283929	231	248
5001	Trondheim	205163	113	136

4.1.2 Covid-19 Data for Germany

In Germany, the Robert Koch Institute publishes daily situation reports in which the number of new cases is published at NUTS 3 level. These reports are available as pdf files via the Institute's website. They can be downloaded and grouped via the R package `covid19germany`(Schmid et al., 2021), as was done for this work.
A few sample data points from this dataset are displayed in Table 4.2.
The variable *CumNumberTestedIll* contains the cumulative number of people that have tested positive for Covid-19.

Tab. 4.2: An excerpt from the Covid-19 data for Germany. Does not contain all variables.

Municipality	Date	Cumulative number of infections	Population
SK München	2020-01-29	1	1471508
SK München	2020-02-03	2	1471508
SK München	2020-02-11	3	1471508
LK Rosenheim	2020-02-29	1	260983
LK Rosenheim	2020-03-08	2	260983
LK Rosenheim	2020-03-10	6	260983

4.2 Vaccination Data

At the end of 2020, the first countries began vaccinating against Covid-19. Vaccination leads to a milder course of the disease, but it is not yet known whether vaccinated people can continue to transmit Covid-19, as of early May 2021. By including the proportion of people vaccinated, one can give an indication of the extent to which vaccination helps prevent the spread of Covid-19. These data are available for Norway at the municipality level via the statistics database of the National Institute of Health, FHI (Folkehelseinstituttet, 2020). For Germany, these data are only available at the state level and are therefore not used.

4.3 Demographic Data

As demographics tend to differ between different geographic units, the decision was made to include demographic variables in the analysis of the research question to see if the risk for infection may be higher when a certain characteristic is present in the population.

4.3.1 Demographic Data for Norway

The demographic data collected for Norway comes from Statistisk Sentralbyrå and is made available to the public through their online database, StatBank(Sentralbyrå, 2016).
The first characteristic collected was the age of the population in a given municipality. For each age, starting at 0 and ending at 105, the number of people of that age is known.
Next, unemployment data were collected for a given municipality. For each municipality, the percentage of all people out of work is known, as well as the percentage of all immigrants out of work.
Other data collected include data related to the number of workers in a particular industry, as well as immigration data. Since there is discussion about whether workers from certain industries, in this case construction, contribute to the spread of Covid-19, the decision was made to collect this type of data. For each community, the number of workers across all industries is known, as well as the number of workers in the construction industry. Workers, in this case, are individuals employed in a given municipality who are between the ages of 20 and 66. It is also known how many people work full-time and how many work part-time.
Finally, for immigration data, it is known how many immigrants live in a given municipality and how many Norwegians were born to immigrant parents. These figures are known in terms of the percentage of the population in 2020.

4.3.2 Demographic Data for Germany

The demographic data collected for Germany comes from the federal and state statistical offices and is made available to the public through their online database, Regionaldatenbank Deutschland (des Bundes und der Länder, 2020).
The first characteristic collected was unemployment data at the NUTS 3 level. For each municipality, the number of unemployed people as well as the number of

unemployed foreigners was collected.
Next, data related to the European elections in 2019 were collected. In each municipality, it is known how many people voted in total, how many people voted for the six largest parties, and how many votes the remaining parties received combined.
Data was also collected in relation to people seeking protection, welfare recipients and in relation to asylum seeker benefits. It is known how many people sought protection in Germany, how many received social welfare and how many received asylum seeker benefits.
Finally, trade tax, income tax, and payroll tax data were collected for each municipality.

4.4 Shapefiles

In addition to numeric variables, the dataset also contains a geographic variable containing the geographic boundaries of a given municipality or city/district.

4.4.1 Shapefiles for Norway

The data for the Norwegian shapefiles comes from Geonorge (Geonorge, 2021) and is downloaded from a Github repository, as the data there was in a cleaner state (Smistad, 2020). In addition to the geographic shape, the dataset also includes a variable that contains the ID of each municipality.

4.4.2 Shapefiles for Germany

The data for the German shapefiles comes from Esri Germany (Deutschland, 2020).

4.5 OpenStreetMap Data

OpenStreetMap (OSM) is a free project that collects, structures and stores freely usable geodata in a database for use by anyone (Open Data). This data is available under a free license, the Open Database License. The core of the project is therefore an openly accessible database of all contributed geoinformation (OpenStreetMap contributors, 2017).

In R, the OpenStreetMap API can be queried using the R package `osmdata` (Padgham et al., 2017). To download all locations of a given type in a given region, a shape or bounding box must be specified along with a key and optionally a value. These key-value pairs are used to specify the type of location, for example, the "amenity" key is used for all facilities used by visitors and residents. If you use the "biergarten" value together with the "amenity" key, the locations of all beer gardens in a given geographic region will be downloaded.

OpenStreetMap users have the option to map a location as either `POINT`, `POLYGON`, `MULTIPOLYGON`, `LINESTRING`, or `MULTILINESTRING`. Conventionally, the first three are used. Therefore, only sites mapped as one of these were used for this work. If a location was mapped as either `POLYGON` or `MULTIPOLYGON`, the centroid of the location was calculated.

A complete list of all key-value pairs used for this work can be found in the Appendix.

4.6 Data Wrangling

The final step before analysing the research question at hand is to combine all of these data sources into one dataset. This section will show how this was achieved.

4.6.1 Data Wrangling for Norway

The initial step in creating the final dataset was to convert the data from a wide format, as seen in Table 4.1, to a long format. This was done using the function `melt()` from the R package `reshape2` (Wickham, 2007). The long version of the dataset is shown in Table 4.3.

Tab. 4.3: An excerpt from the long version of the Norwegian Covid-19 data. Does not contain all variables.

Id	Municipality	Population	Date	Infections
1507	Ålesund	66258	2020-03-26	20
5001	Trondheim	205163	2020-03-26	113
1507	Ålesund	66258	2020-03-27	20
5001	Trondheim	205163	2020-03-27	136

Next, the demographic data for Norway was loaded and processed. Since the age data contains the number of people of a certain age, the median age was calculated for each region based on how many people of each age group live in each region. The other demographic variables are left unchanged. To combine the demographic data with the Covid-19 data, the municipality IDs were extracted using the `str_extract()` function from the `stringr` (Wickham, 2019) R package using the regular expression `[0-9]{4}`. Next, all demographic datasets and the Covid-19 dataset were merged using the `merge()` function.
Using the `st_intersects()` function from the `sf` (Pebesma, 2018) R package, the number of points of interest downloaded via OpenStreetMap was calculated for each municipality. Since the shapefiles contain the ID for each community, these data were then merged with the data containing the demographic and Covid-19 data.
For each numeric variable, e.g. the number of schools or the number of employees, this number was scaled.
If there were missing values in the covariates, these values were imputed using the median of the respective variable.
Next, the vaccination data for Norway was loaded. As the daily number of vaccinated persons for each municipality is included in the data, these numbers only needed to

be cumulated before being merged with the rest of the data based on municipality name and date.
Finally, seven new variables were created:

1. Expected count, which is the expected number of cases in each municipality.
2. SIR, which is the standardised incidence ratio in each municipality.
3. An area ID, which is a unique ID given to each municipality.
4. Higher education, which counts the number of universities and colleges in a given area.
5. Sex, which gives the proportion of females living in a given area.
6. Population density, i.e. the number of people per square kilometre in a given area.
7. Urban density, i.e. the number of residential buildings per square kilometre in a given area.

The final dataset contains the variables shown in Table 4.4.

Tab. 4.4: The variables contained in the final dataset.

Variable Name	Explanation	Scale
Id	The municipality ID	None
Municipality	The municipality name	None
Population	Population in a municipality	None
Date	The date of the data used	None
Infections	The number of infected people	None
Median age	The median age	scaled
Total unemployment	The proportion of unemployed people	scaled
Unemployed immigrants	The proportion of unemployed immigrants	scaled
Full-time workers	The number of full-time workers	scaled
Part-time workers	The number of part-time workers	scaled
Full-time construction	The number of full-time construction workers	scaled
Part-time construction	The number of part-time construction workers	scaled
Total immigrants	The proportion of immigrants	scaled
Marketplace	The number of marketplaces	scaled
Entertainment	The number of entertainment venues	scaled
Sport	The number of sports amenities	scaled
Clinic	The number of clinics	scaled
Hairdresser	The number of hairdresser	scaled
Shops	The number of shops	scaled
Place of worship	The number of places of worship	scaled
Retail	The number of retail stores	scaled
Nursing home	The number of nursing homes	scaled
Restaurant	The number of restaurants	scaled
Aerodrome	The number of aerodromes	scaled
Office	The number of offices	scaled
Platform	The number of public transport platforms	scaled
Kindergarten	The number of kindergartens	scaled
Schools	The number of schools	scaled
Bakeries	The number of bakeries	scaled
Residential	The number of residential buildings	None
Higher education	The number of colleges and universities	scaled
Expected count	The expected number of infections	None
SIR	The standardised incidence ratio	None
Area Id	A unique ID	None
Area	The area in km^2	None
Population density	People per km^2	scaled
Urban density	Residential buildings per km^2	scaled
Sex	The proportion of females	scaled
Vaccinations	The proportion of vaccinated	scaled

4.6.2 Data Wrangling for Germany

The data processing procedure for Germany is identical to that for Norway. First, all demographic variables were loaded and left unchanged before being merged with the Covid-19 data prior to calculating the spatial intersections between the points of interest and the NUTS-3 areas. After merging all the data, the scaled numbers were calculated for the numeric variables. For the variables containing the number of people who voted for a particular political party, the relative percentage of votes the party received was calculated. Again, missing values were imputed using the median. Finally, the same seven new variables were created. The final dataset contains the variables shown in Table 4.5.

Tab. 4.5: The variables contained in the final dataset.

Variable Name	Explanation	Scale
Id	The municipality ID	None
Municipality	The municipality name	None
Population	Population in a municipality	None
Date	The date of the data used	None
Infections	The number of infected people	None
Logarithmic trade tax	The trade tax in Euros	scaled
Logarithmic income tax	The income tax in Euros	scaled
Logarithmic total income	The income and payroll tax in Euros	scaled
Asyl benefits	The number of people receiving asylum seeker benefits	scaled
Welfare recipients	The number of welfare recipients	scaled
Unemployed people	The number of unemployed people	scaled
Unemployed foreigners	The number of unemployed foreigners	scaled
Protection seekers	The number of protection seekers	scaled
Die Union	Percentage of vote for Union	scaled
SPD	Percentage of vote for SPD	scaled
Greens	Percentage of vote for the Greens	scaled
FDP	Percentage of vote for FDP	scaled
The left	Percentage of vote for the left	scaled
AfD	Percentage of vote voted for AfD	scaled
Marketplace	The number of marketplaces	scaled
Entertainment	The number of entertainment venues	scaled
Sport	The number of sports amenities	scaled
Clinic	The number of clinics	scaled
Hairdresser	The number of hairdresser	scaled
Shops	The number of shops	scaled
Place of worship	The number of places of worship	scaled
Retail	The number of retail stores	scaled
Nursing home	The number of nursing homes	scaled
Restaurant	The number of restaurants	scaled
Aerodrome	The number of aerodromes	scaled
Office	The number of offices	scaled
Platform	The number of public transport platforms	scaled
Kindergarten	The number of kindergartens	scaled
Schools	The number of schools	scaled
Bakeries	The number of bakeries	scaled
Residential	The number of residential buildings	None
Higher education	The number of colleges and universities	scaled
Expected count	The expected number of infections	None
SIR	The standardised incidence ratio	None
Area Id	A unique ID	None
Area	The area in km^2	None
Population density	People per km^2	scaled
Urban density	Residential buildings per km^2	scaled
Sex	The proportion of females	scaled

5 Data Analysis

This chapter will focus on the analysis of the research question. First, it will take a look at the standardised incidence rate in each country. Next, the relationship between the number of infections and several factors of interest will be analysed. For this analysis, a probability must be found that fits the data reasonably well. One way to do this is presented in Section 5.2.1. Thereafter, two different types of models are calculated; models without a spatial component and models with a spatial component. Section 5.3 contains the results of the models without a spatial component and Section 5.4 contains the results of the models with the spatial component. Next, it is analysed how the performance of the spatial models changes when changing the value for the standard deviation σ_0 in Equation 2.12.

81

5.1 Standardised Incidence Ratio (SIR)

This section takes a brief look at the SIR for the countries of interest. Recall from Equation 3.5, that the SIR is defined as the ratio of observed counts to expected counts.

5.1.1 SIR for Germany

When looking at the SIR for Germany in Figure 5.1, it is noticeable that the actual number of infections in the eastern parts of Germany, especially in Saxony, is considerably higher than the expected number of infections. Furthermore, parts of Bavaria have an increased SIR compared to the rest of Germany, excluding Saxony. This could be due to the fact that the regions share a border with the Czech Republic, a country that is substantially more affected by Covid-19 than Germany. The northern parts of Germany show the lowest SIR which is possibly due to the fact that this region is sparsely populated.

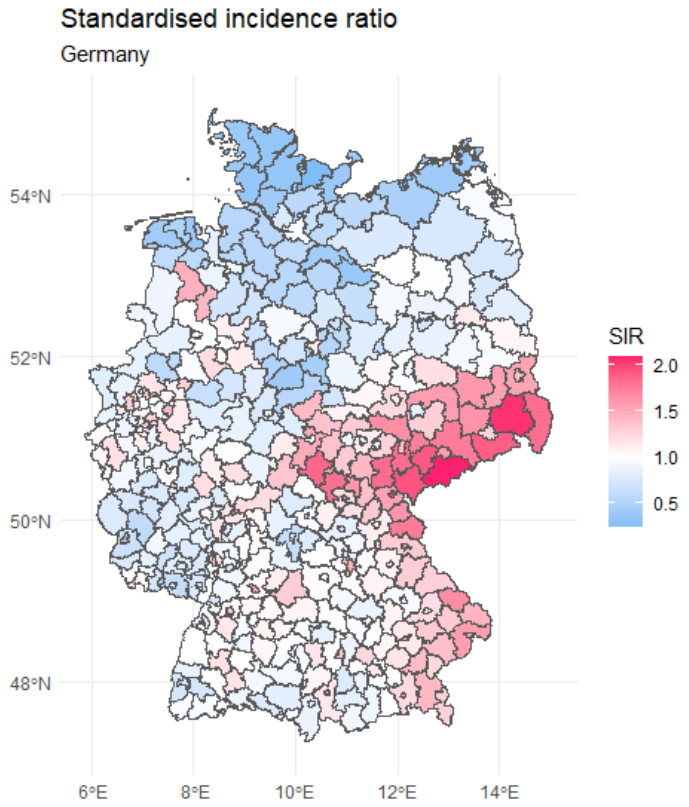

Fig. 5.1: The SIR for Germany based on the data of the 2nd of May 2021

5.1.2 SIR for Norway

Looking at the standardised incidence rate for Norway in Figure 5.2, a standardised incidence rate of less than 1 can be seen for most municipalities north of Trondheim. In the southern parts of Norway there are several municipalities with a rate above 1, for example the standardised incidence rate around the capital Oslo is around 2. However, the two small municipalities, Hyllestad and Ulvik, have the highest standardised incidence rate in Norway. In Hyllestad, 95 of 1328 people have been infected with Covid-19 so far, while in Ulvik, 134 of 1080 people have been infected so far.
The SIR in Hyllestad is around 3.4, following an outbreak in a shipyard in autumn 2020 (Korsvoll, 2020), while Ulvik has a ratio of around 5.7, following an outbreak of the UK variant of Covid-19. According to the head of the municipality, Hans Petter Thorbjørnsen, the infections are thought to have spread through children (NTB, 2021).

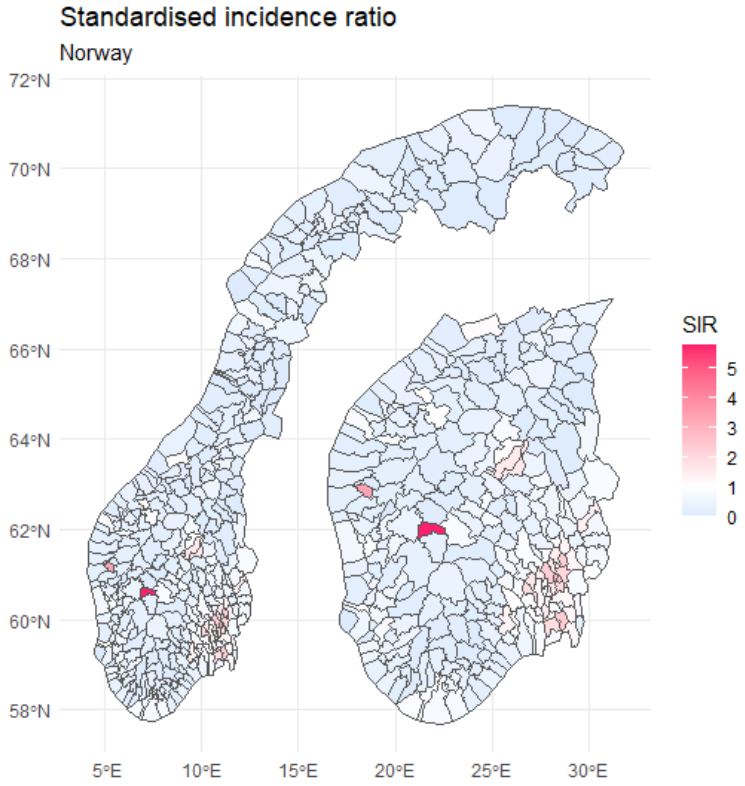

Fig. 5.2: The SIR for Norway based on the data of the 2nd of May 2021

Because the high numbers from two small municipalities complicate the interpretation of Figure 5.2, Figure 5.3 shows the SIR on a log10 scale. On this scale, a

value of 0 means that the risk of infection in a given municipality is neither lower nor higher. Values below 0 mean that the risk of infection in a municipality is lower than average, while values above 1 mean that the risk of infection in a municipality is higher than average. It is now clearer that the standardized incidence ratio is below 1 in most parts of Norway, but that there is a higher risk in the region around Oslo.

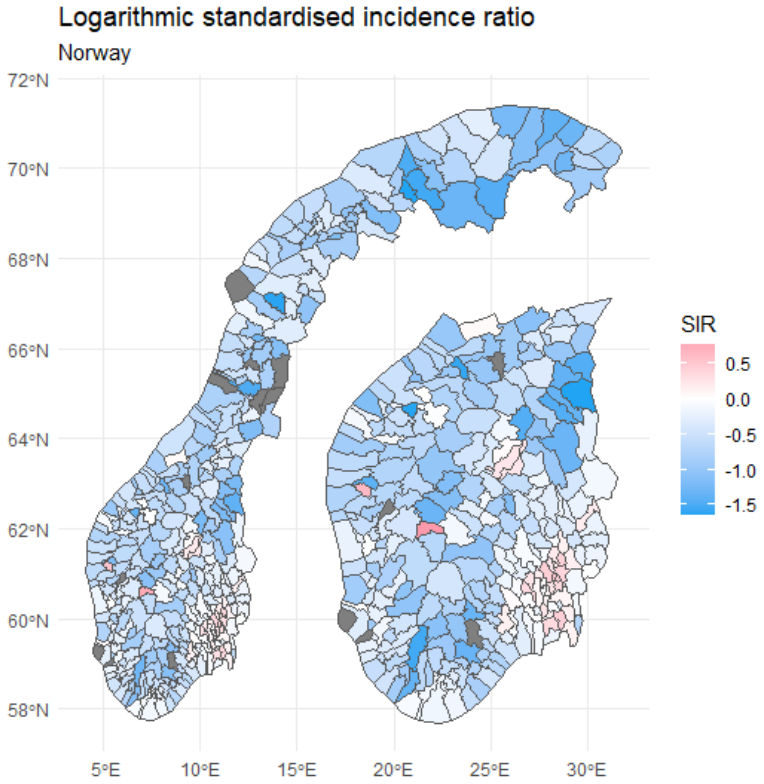

Fig. 5.3: The log10 SIR for Norway based on the data of the 2nd of May 2021

5.2 Data Modelling

After looking at the standardised incidence rates for the countries of interest, the next step is to take a closer look at the current figures for the respective countries. Spatial models are used to try to extract the factors that cause some populations to be at higher risk than populations in other geographical regions. Three different types of models are used for each country:

1. Besags Proper Spatial Model
2. A Leroux Model
3. A BYM2 Model

All of these models are computed using the INLA (Bivand et al., 2015) R package. To specify each type of model, the code shown in Listing 8.1 can be used.
The measures introduced in Section 2.7, namely the DIC, the WAIC, the CPO and the mean absolute error (MAE), are used to compare the models.
In addition to specifying what type of spatial model to use, if any, there is also the option of specifying a prior.
As can be seen in Section 2.3.2, a pc prior can be specified for the precision parameter τ, which is what is done here.
For the parameters σ_0 and α in Equation 2.13 the values 1 and 0.01 are chosen.
The models are compared using the mean absolute error. For this, 20% of the observations are removed from the training set and used for testing instead. The predicted number of infections for these municipalities is then compared to the actual numbers.

5.2.1 Choice of Likelihood

Before the models are computed, however, the distribution that fits the number of cases must first be found. One way to do this, the function `descdist()` from the `fitdistrplus` R package is used. The Cullen and Frey graph illustrates how "close" a sample is to a theoretical distribution based on the kurtosis and the square of the skewness, defined in Equation 8.18 and Equation 8.17. It can be used to get a preliminary idea of which distributions fit the data, in this case the number of infections, reasonably well.
The plots for Germany and Norway can be seen in Figure 5.4 and Figure 5.5. The blue dot represents the data, the star a theoretical normal distribution, the dashed line a theoretical Poisson distribution and the grey area a theoretical negative

binomial distribution. In both cases, the blue dot is relatively far from the star and lies in the region of a negative binomial distribution. For the Norwegian sample shown in Figure 5.5, the sample is closer to a Poisson distribution than is the case for the German sample in Figure 5.4.

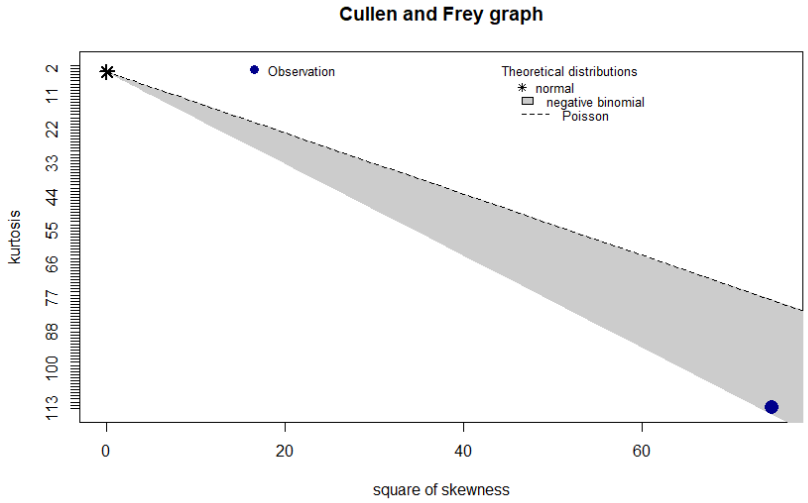

Fig. 5.4: The Cullen and Frey graph for Germany

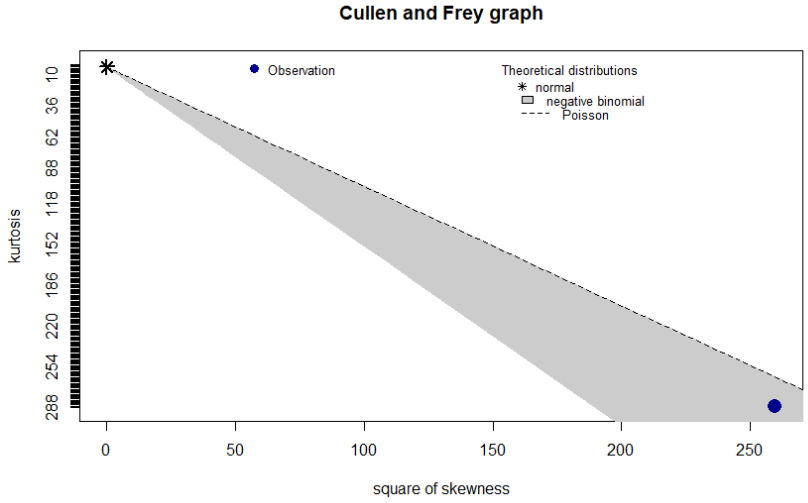

Fig. 5.5: The Cullen and Frey graph for Norway

Next, a negative binomial distribution, a normal distribution, and a Poisson distribution are fitted to the data using the maximum likelihood method. The negative binomial fits for both countries can be seen in Figure 5.6 and Figure 5.7. The fits for the normal and Poisson distribution for both countries, are shown in the Appendix

in Figure 8.1, Figure 8.2, Figure 8.4 and Figure 8.5.
The QQ-plot for Germany and Norway looks quite similar, as there appears to be a linear relationship between the theoretical quantile and the sample quantiles, up to a certain point where the sample quantiles have a higher value than the theoretical quantiles, indicating that the distribution is right skewed. Since there are many municipalities with relatively few cases and few municipalities with a large number of cases, this is to be expected. It can also be seen that the empirical cumulative density function closely follows the theoretical cumulative density function.

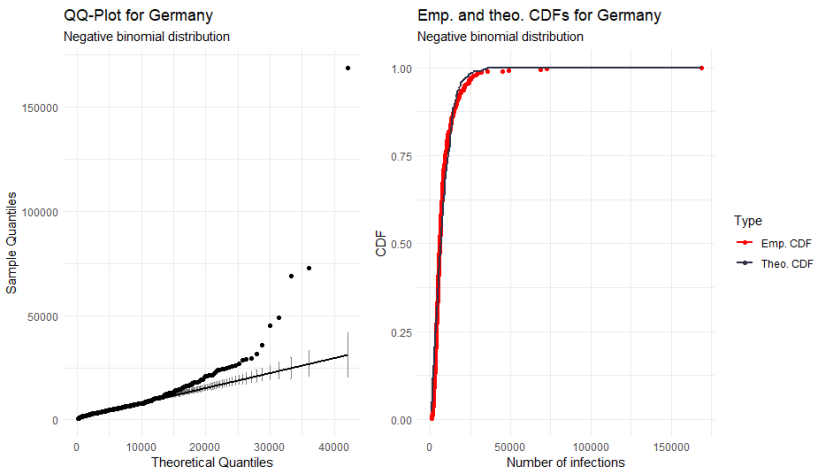

Fig. 5.6: A negative binomial fit to the number of cases in German municipalities

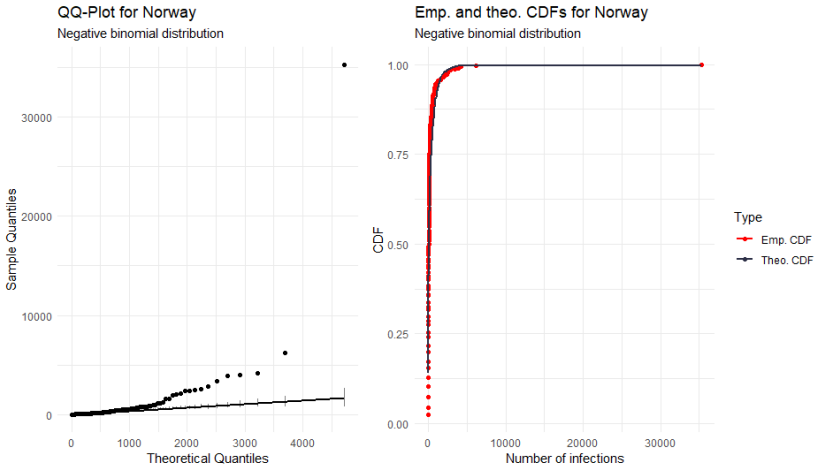

Fig. 5.7: A negative binomial fit to the number of cases in Norwegian municipalities

Lastly, the AIC is calculated for fitting a normal distribution to the data, a Poisson distribution to the data and a negative binomial distribution to the data. The values

can be seen in Table 5.1. Afterwards, the negative binomial distribution is chosen as the distribution of the target variable in both cases.

Tab. 5.1: The AIC for different distributions for Germany and Norway

Country	Distribution	AIC
Germany	Normal	8619
Germany	Poisson	2699357
Germany	Negative Binomial	8010
Norway	Normal	6400
Norway	Poisson	509139
Norway	Negative Binomial	4293

The poor fit for the Poisson distribution can be explained by looking at the range of the number of confirmed cases in a given municipality. For Germany, this number ranges from 697 to 169021 (as of May 2, 2021), while for Norway, the number ranges from 0 to 34654 (as of May 2, 2021). This results in a mean and standard deviation for Germany of 8550 and 11204, respectively. For Norway, the values for these metrics are 326 and 1930. This is problematic because, as shown in Equation 8.12 and Equation 8.13, for a Poisson distribution the expected value and the variance should be equal.
Looking at a histogram for the confirmed number of cases and overlaying the densities of a normal, Poisson and a negative binomial distribution helps to confirm the choice of a negative binomial distribution as the distribution that the data most closely resembles. Figure 5.8 and Figure 5.9 both show that a negative binomial distribution fits the data better than a normal distribution. Due to the high values for the AIC, the Poisson distribution is excluded from these graphics.

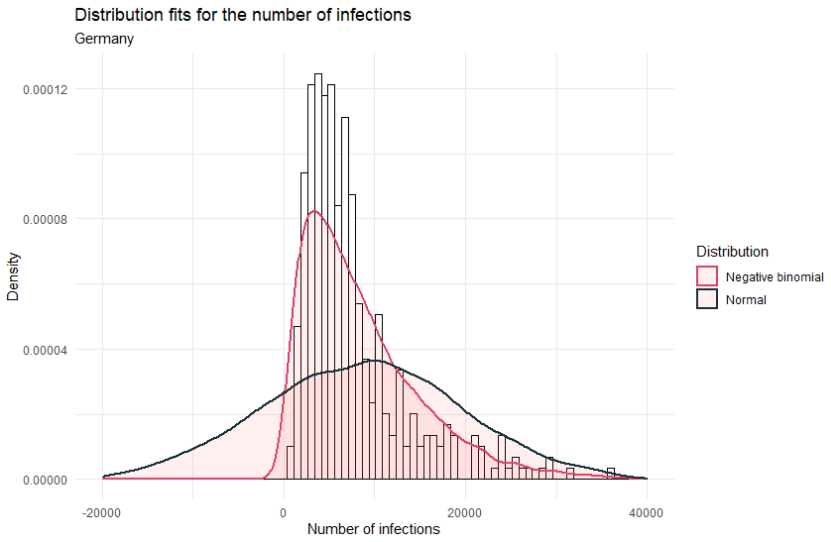

Fig. 5.8: Histogram for the number of cases in German municipalities with a normal and a negative binomial distribution overlayed.

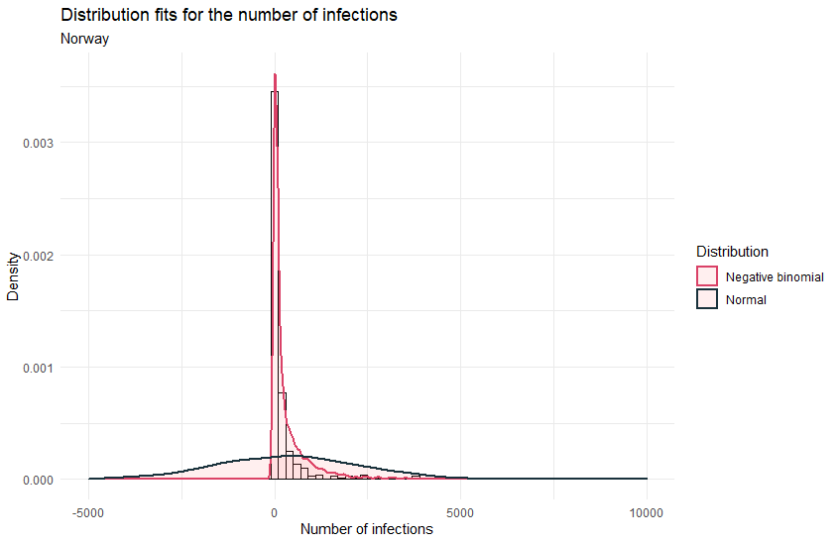

Fig. 5.9: Histogram for the number of cases in Norwegian municipalities with a normal and a negative binomial distribution overlayed.

5.3 Models without a Spatial Component

To establish a baseline, a look is first taken at models that do not include a spatial effect. This way, it can be observed how the means and credibility intervals of the covariates change when a spatial effect is added to a model and how the performance of the model changes with respect to the goodness-of-fit indicators introduced in Section 2.7.

Before computing the models, using the VIF introduced in Section 2.9, predictors are removed if their VIF is above 5. To do this, a GLM is first run on all variables, followed by the calculation of VIF. Then the variable with the highest value is removed before running the model again with all remaining variables. This process is repeated until only variables with a VIF of less than 5 remain.

In sections 5.3.1 and 5.3.2 the results of these models are presented. Their goodness-of-fit indicators as well as the coefficients together with their credibility intervals calculated as in section 2.2.4 are reported.

These models are based on data from 2 May 2021, when 3,428,487 people were infected with Covid-19 in Germany, while 87,537 people were infected in Norway. The five municipalities with the most infections in Germany are shown in Table 5.2 and in Table 5.3 for Norway.

Tab. 5.2: The German municipalities with the most infections as of May 2nd 2021.

Municipality	Population	Number of infections
SK Berlin	3644826	169021
SK Hamburg	1841179	72595
SK Munich	1471508	68762
SK Cologne	1085664	48831
Region Hannover	1157624	45043

Tab. 5.3: The Norwegian municipalities with the most infections as of May 2nd 2021.

Municipality	Population	Number of infections	% first vaccine shot
Oslo	693494	34654	32.7%
Bergen	283929	6039	30.0%
Drammen	101386	4097	34.1%
Bærum	127731	3860	31.7%
Lillestrøm	85983	3819	30.2%

5.3.1 Models without a Spatial Component for Germany

Table 5.4 contains the performance measures for the baseline model for Germany, while Table 5.5 contains the posterior mean, the exponentiated posterior mean and the credibility intervals of the coefficients. It can be seen that the intercept as well as six of the coefficients are significant.

Tab. 5.4: The performance measures for the model without a spatial component.

DIC	WAIC	CPO	MAE_{train}	MAE_{test}
5595	5598	-2816	1433	1306

Tab. 5.5: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	$mean_p$	$exp(mean_p)$	$exp(q0025_p)$	$exp(q0975_p)$	sig.
(Intercept)	-0.149	0.960	0.936	0.985	*
AfD	0.179	1.198	1.062	1.346	*
Population density	0.157	1.170	1.117	1.226	*
Logarithmic trade tax	0.084	1.088	1.041	1.136	*
Platform	0.037	1.038	0.988	1.089	
Die Union	0.032	1.035	0.897	1.188	
Higher Education	0.018	1.018	0.981	1.058	
Sex	0.016	1.016	0.984	1.049	
Urban density	0.008	1.009	0.973	1.047	
FDP	0.007	1.007	0.972	1.044	
Place of worship	-0.004	0.996	0.955	1.039	
Nursing home	-0.012	0.988	0.959	1.018	
Clinic	-0.013	0.988	0.940	1.039	
Aerodrome	-0.014	0.986	0.954	1.019	
Office	-0.029	0.971	0.930	1.015	
Marketplace	-0.044	0.957	0.905	1.011	
SPD	-0.087	0.918	0.850	0.989	*
The left	-0.117	0.890	0.823	0.961	*
Greens	-0.149	0.863	0.758	0.979	*

5.3.2 Models without a Spatial Component for Norway

Table 5.6 contains the performance measures for the baseline model for Germany, while Table 5.7 contains the posterior mean, the exponentiated posterior mean and the credibility intervals of the coefficients. It can be seen that the intercept as well as five of the coefficients are significant.

Tab. 5.6: The performance measures for the model without a spatial component.

DIC	WAIC	CPO	MAE_{train}	MAE_{test}
2830	2835	-1689	287	82

Tab. 5.7: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	$mean_p$	$exp(mean_p)$	$exp(q0025_p)$	$exp(q0975_p)$	sig.
(Intercept)	-0.831	0.436	0.399	0.476	*
Unemployed immigrants	0.229	1.262	1.077	1.475	*
Total immigrants	0.208	1.234	1.079	1.407	*
Urban density	0.188	1.211	1.049	1.413	*
Total unemployment	0.112	1.125	0.909	1.382	
Vaccinations	0.073	1.078	0.960	1.206	
Marketplace	0.044	1.046	0.953	1.159	
Platform	0.040	1.043	0.907	1.204	
Higher education	0.024	1.025	0.943	1.129	
Nursing home	0.010	1.011	0.930	1.115	
Place of_ worship	-0.024	0.979	0.846	1.136	
Median age	-0.044	0.959	0.848	1.079	
Aerodrome	-0.128	0.881	0.802	0.968	*
Office	-0.154	0.860	0.736	1.008	
Sex	-0.164	0.850	0.756	0.952	*

5.4 Spatial Models

Looking at the SIR value for Germany in Figure 5.1 and the SIR value for Norway in Figure 5.2 and Figure 5.3, it clearly looks like there is a correlation between the SIR and the spatial units. For Germany, the SIR is higher in eastern Germany than in the rest of the country and in Norway the risk seems to be mainly concentrated around the Oslo region.
To check whether there is spatial autocorrelation, Moran's I, introduced in Section 3.2.1.2, can be calculated and the Moran test can then be used to tell whether there is spatial autocorrelation. Under the null hypothesis of no spatial autocorrelation, a p-value greater than 0.05 would be expected. The results of the test are presented in Table 5.8. Looking at the p-value for both countries, it can be seen that the number of infections in a municipality and the spatial units are correlated.

Tab. 5.8: Results of the Moran test for Germany and Norway.

Country	**Moran's I**	$\mathbb{E}[I]$	**p-Value**
Germany	0.110	-0.003	< 0.01
Norway	0.110	-0.003	< 0.01

Therefore, after the models without spatial effect have been calculated and established as baseline models, a spatial term is added to the models calculated in Section 5.3 , in order to model this spatial correlation.

5.4.1 Spatial Models for Germany

Looking at the performance of the spatial models and the model with the spatial component shown in Table 5.9, it can be seen that the spatial models perform better in terms of the DIC, WAIC and MAE, while they perform equally well or better in terms of the CPO.
The best performance of all models, in terms of MAE, was observed for the Besag model, just ahead of the BYM2 model.

Tab. 5.9: The performance measures for the best performing model of each type.

Model	DIC	WAIC	CPO	MAE_{train}	MAE_{test}
No spatial	5595	5598	-2816	1433	1306
Besag	4832	4875	-2802	85	1017
BYM2	4732	4705	-2751	73	1030
Leroux	4818	4882	-2919	78	1271

Figure 5.10 shows the differences between the coefficients in the model without the spatial component and the BYM2 model. Excluding the intercept, only three effects are significant in the BYM2 model compared to six in the model without the spatial component. Moreover, the coefficients of the BYM2 model are closer to 1.

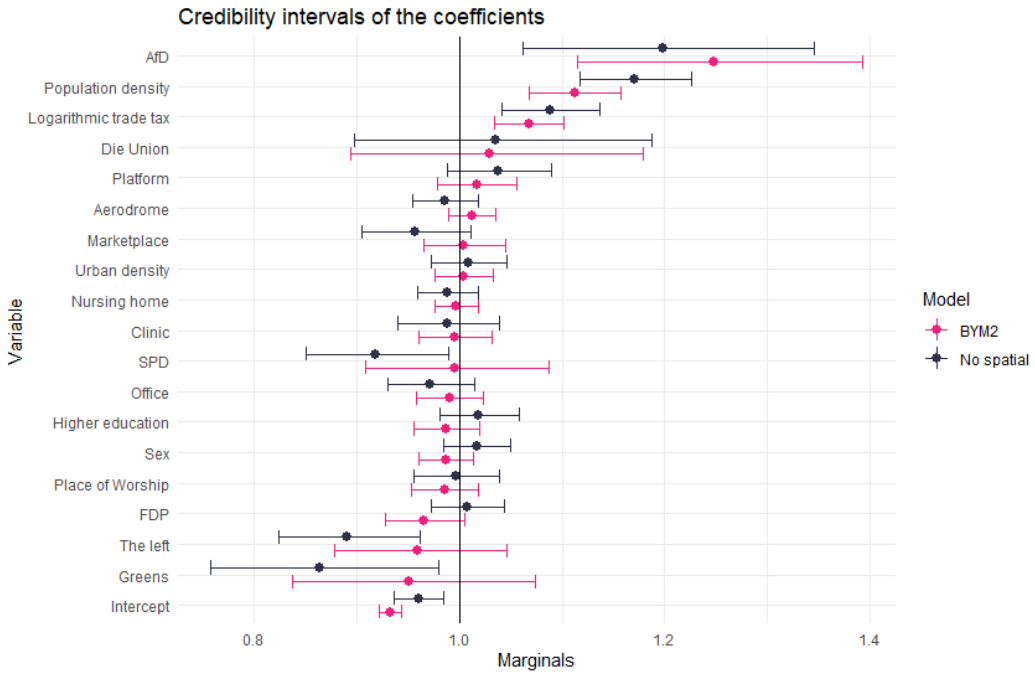

Fig. 5.10: The posterior mean and credibility intervals of the coefficients

The values of the coefficients and credibility intervals are shown in Table 5.10.

Tab. 5.10: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	mean_p	$\exp(\text{mean}_\text{p})$	$\exp(\text{q0025}_\text{p})$	$\exp(\text{q0975}_\text{p})$	sig.
(Intercept)	-0.070	0.932	0.921	0.943	*
AfD	0.220	1.250	1.114	1.393	*
Population density	0.106	1.112	1.068	1.157	*
Logarithmic trade tax	0.065	1.067	1.034	1.101	*
Die Union	0.026	1.029	0.894	1.179	
Platform	0.016	1.017	0.978	1.056	
Aerodrome	0.012	1.012	0.989	1.035	
Marketplace	0.004	1.004	0.965	1.044	
Urban density	0.004	1.004	0.976	1.032	
Nursing Home	-0.003	0.997	0.975	1.018	
Clinic	-0.005	0.996	0.960	1.032	
SPD	-0.006	0.995	0.908	1.085	
Office	-0.010	0.990	0.958	1.023	
Higher Education	-0.013	0.987	0.955	1.019	
Sex	-0.014	0.987	0.960	1.014	
Place of Worship	-0.015	0.985	0.953	1.018	
FDP	-0.036	0.965	0.927	1.005	
The left	-0.043	0.959	0.878	1.046	
Greens	-0.053	0.950	0.837	1.073	

For the hyperparameters, a value of 18.61 is reported for the precision and a value of 0.9032 for ϕ. Hence, 90.32% of the marginal variance is explained by the structured effect. Therefore, this model is far from reducing to pure overdispersion and comes close to a Besag model, which is also reflected in the similar values of the goodness-of-fit indicators in Table 5.9.

5.4.2 Spatial Models for Norway

Comparing the performance of the models in Table 5.11, the spatial models again showed better performance in terms of DIC and WAIC and this time also significantly better performance in terms of CPO. The best performance in terms of MAE was observed for the Leroux model, but the MAE was quite close for all models.

Tab. 5.11: The performance measures for the best performing model of each type.

Model	DIC	WAIC	CPO	MAE_{train}	MAE_{test}
No spatial	2830	2835	-1689	287	82
Besag	2824	2830	-2099	265	79
BYM2	2780	2787	-4387	171	78
Leroux	2359	2340	-4820	11	73

Looking at the differences between the coefficients and credibility intervals in Figure 5.11, the picture is similar to Figure 5.10. This time, however, all significant effects of the model without the spatial component are still significant. This again shows that the spatial effect is weaker in Norway than in Germany, as no new variables turn out to be significant and no variables lose their significance when the spatial term is added.

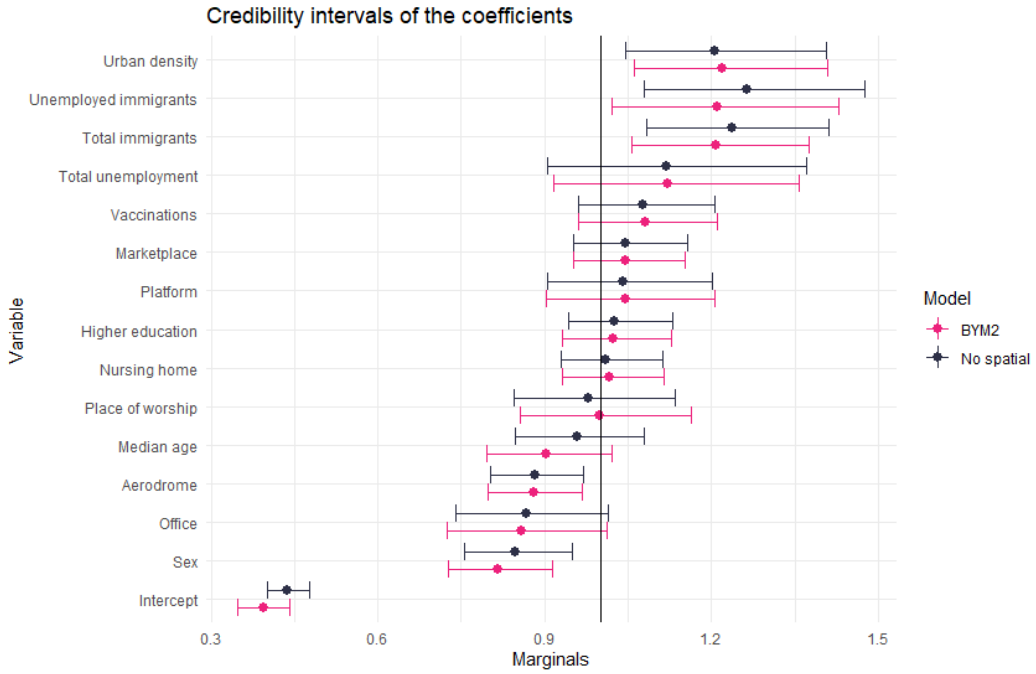

Fig. 5.11: The posterior mean and credibility intervals of the coefficients

The values of the coefficients and credibility intervals are shown in Table 5.12.

Tab. 5.12: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	mean_p	$\exp(\text{mean}_\text{p})$	$\exp(\text{q0025}_\text{p})$	$\exp(\text{q0975}_\text{p})$	sig.
(Intercept)	-0.939	0.392	0.345	0.440	*
Urban density	0.200	1.225	1.064	1.415	*
Unemployed immigrants	0.185	1.208	1.017	1.425	*
Total immigrants	0.183	1.203	1.052	1.371	*
Total unemployment	0.115	1.127	0.921	1.367	
Vaccinations	0.076	1.080	0.959	1.211	
Marketplace	0.046	1.048	0.955	1.156	
Platform	0.044	1.048	0.905	1.211	
Higher education	0.021	1.022	0.931	1.128	
Nursing home	0.016	1.017	0.933	1.117	
Place of worship	-0.003	1.001	0.857	1.166	
Median age	-0.105	0.902	0.794	1.020	
Aerodrome	-0.311	0.878	0.797	0.966	*
Office	-0.164	0.852	0.719	1.003	
Sex	-0.202	0.818	0.729	0.916	*

For the hyperparameters, a value of 5.398 is reported for the precision and a value of 0.130 for ϕ. Hence, 13.00% of the marginal variance is explained by the structured effect. Therefore, this model is close to pure overdispersion.

5.5 Choice of Hyperpriors

As can be seen in Equation 2.12, there is flexibility when it comes to choosing the values for the standard deviation σ_0 as well as the probability α. Therefore, an upper bound for the standard deviation can be chosen as well as the weight placed on this "tail event", describing how informative the resulting prior is.
Some of the issues that come with the choice of these hyperpriors were already discussed in Section 2.8.
In the following, an assessment is made of how the performance of a Besag model, a BYM2 model and a Leroux model changes when playing around with the value for the standard deviation σ_0. To create these plots, models were calculated with σ_0 values of $\boldsymbol{\sigma_0} = (0.1, 0.11, 0.12, ..., 5)$.
In Figure 5.12 it can be seen that when choosing a higher value for σ_0, the DIC and WAIC is lower in the case of the Besag model and the BYM2 model. For the Leroux model, on the other hand, the WAIC gets lower until about 2 before it rises until around $\sigma_0 = 2.5$ and then flattens out. It is a positive sign that this is not the case with the BYM2 model, as it was designed to avoid exactly this kind of thing.
For the MAE in Figure 5.13, it can be seen that for all model types, a higher value for σ_0 leads to a lower MAE, however not by much.

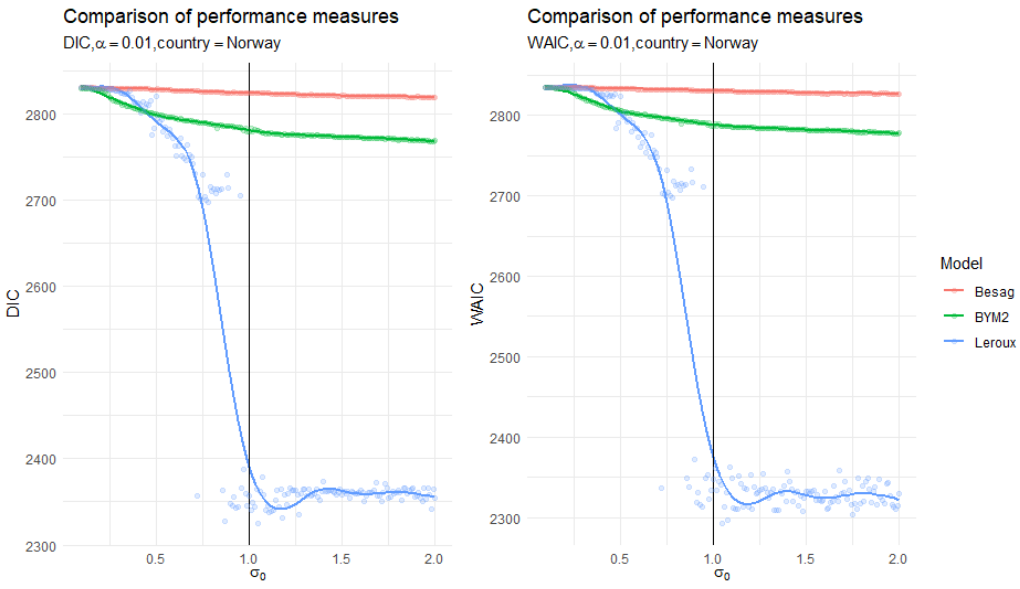

Fig. 5.12: Values of the DIC and the WAIC when changing the value for σ_0. The black line highlights the values for $\sigma_0 = 1$.

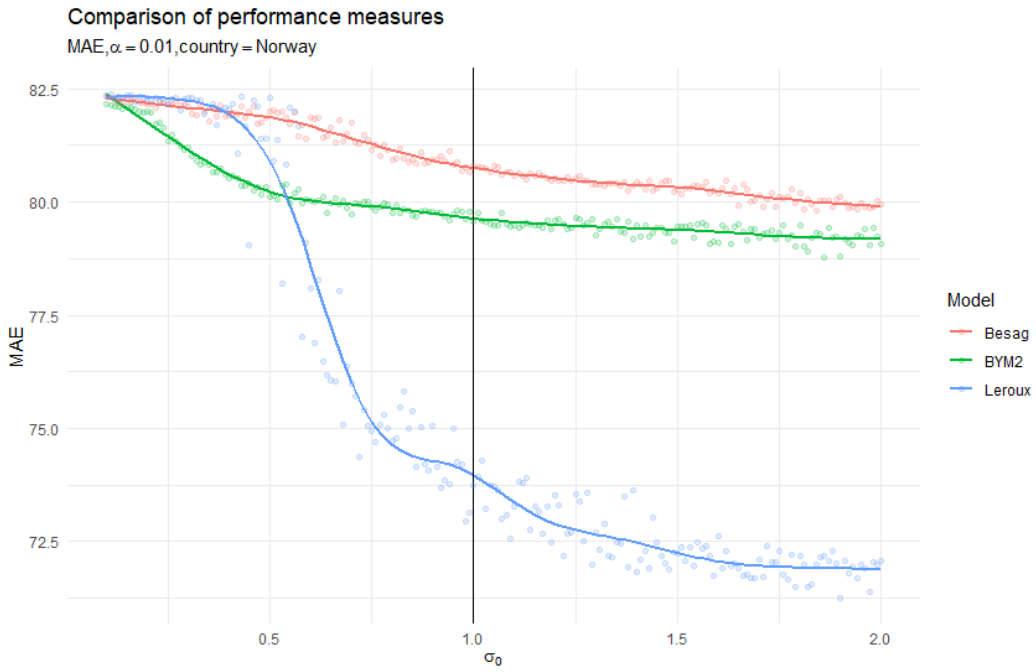

Fig. 5.13: Value of the MAE when changing the value for σ_0. The black line highlights the values for $\sigma_0 = 1$.

By allowing the precision to be greater, the variance is forced to be smaller. Hence, choosing a lower value for the precision leads to lower values for the WAIC. While this indicates a better fit to the training data, Figure 5.13 also shows that the MAE increases when a higher value for σ_0 is chosen, as the models overfit on the training data and therefore make worse predictions.
The corresponding figures for Germany are shown in Figure 8.10 and Figure 8.11 in the Appendix.
Figure 5.14 shows how the credibility intervals of the coefficients of a BYM2 model change when the value for σ_0 is increased. The values of the coefficients tend to remain relatively similar most of the time, especially when the value of the coefficient is close to 1. However, a few times, for example for the variables immigrants_total, platform and unemp_immg, the values differ. Furthermore, the coefficients for $\sigma_0 = 1$ and $\sigma_0 = 2$ are more closer to 1 than the ones for $\sigma_0 = 0.1$.

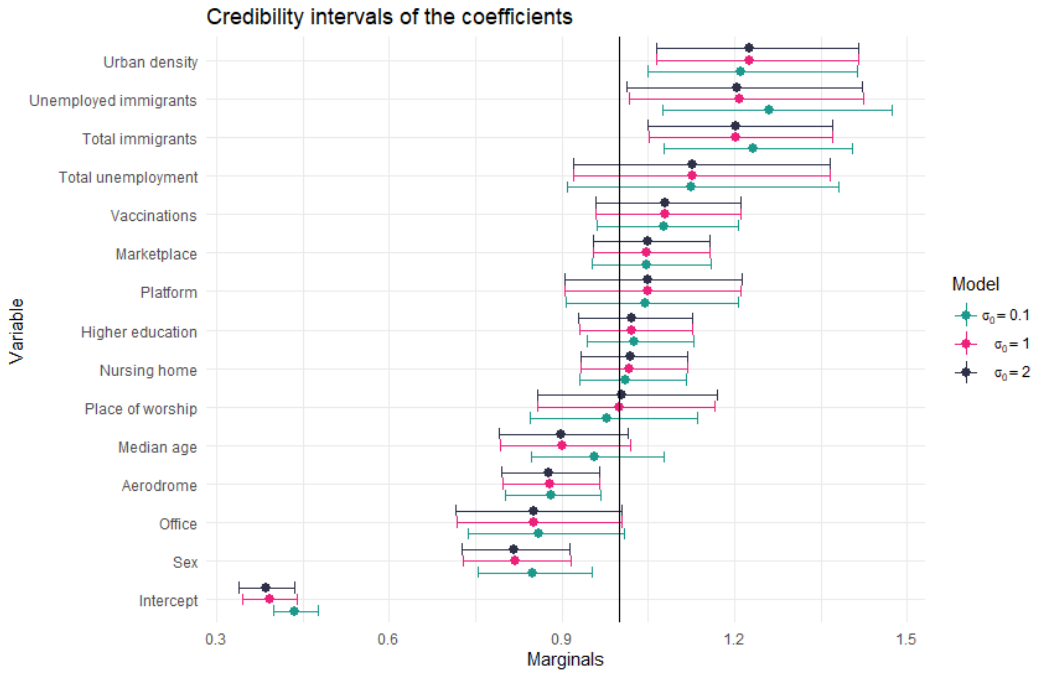

Fig. 5.14: Comparison of the credibility intervals of a BYM2 model for different values of σ_0.

Having credibility intervals and posterior means that are very similar to each other, regardless of the value chosen for σ_0, is a great sign as it means that the calculated model is robust. That is, it means that the fixed covariates are important in their own right because they provide different information than the spatial field, even if the spatial field is allowed to be very smooth or coarse. This is exactly the case for Germany, which can be seen in Figure 5.15.

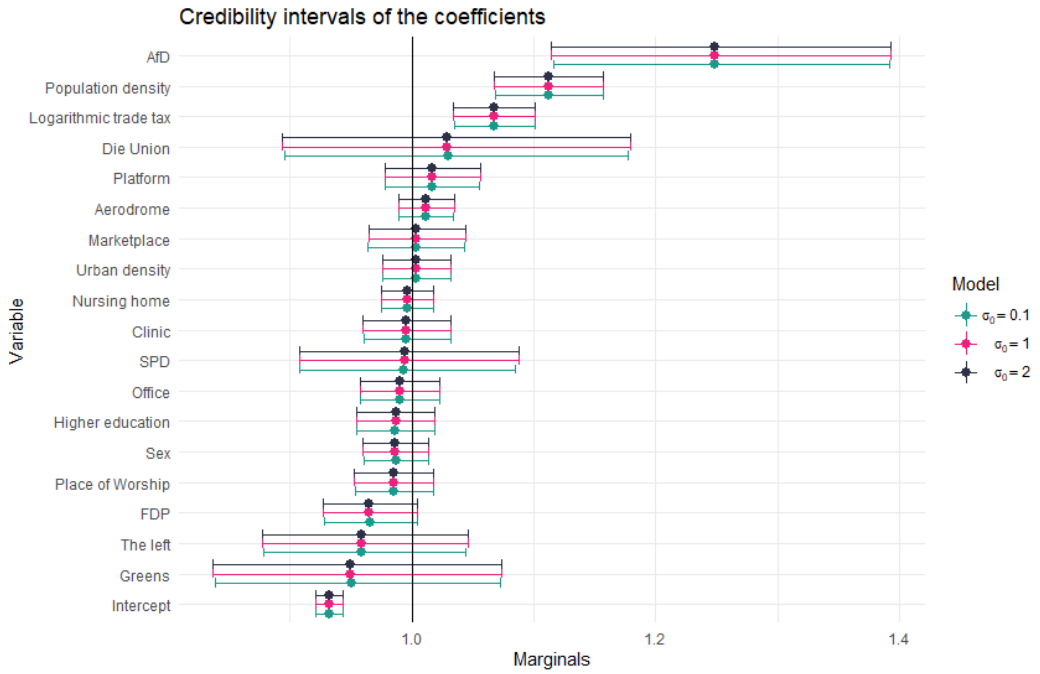

Fig. 5.15: Comparison of the credibility intervals of a BYM2 model for different values of σ_0

Figure 5.16 and Figure 5.17 underline the problem of the models not being comparable with each other. Figure 5.16 already shows huge differences in the spatial field of the Besag model and the Leroux model, with the spatial field of the Besag model looking smoother than the spatial field of the Leroux model. This is a sign that the Leroux model is quickly overfitting to the data. In the left part of Figure 5.17 the values of Equation 2.61 are plotted, while in the right part the values of u_* are plotted.

For the spatial field of the unstructured random effect, the values of the posterior mean are similar to the values of the Besag model. For the structured component, however, the absolute values are somewhat higher. Nevertheless, in both cases the values are far from the values for the Leroux model.

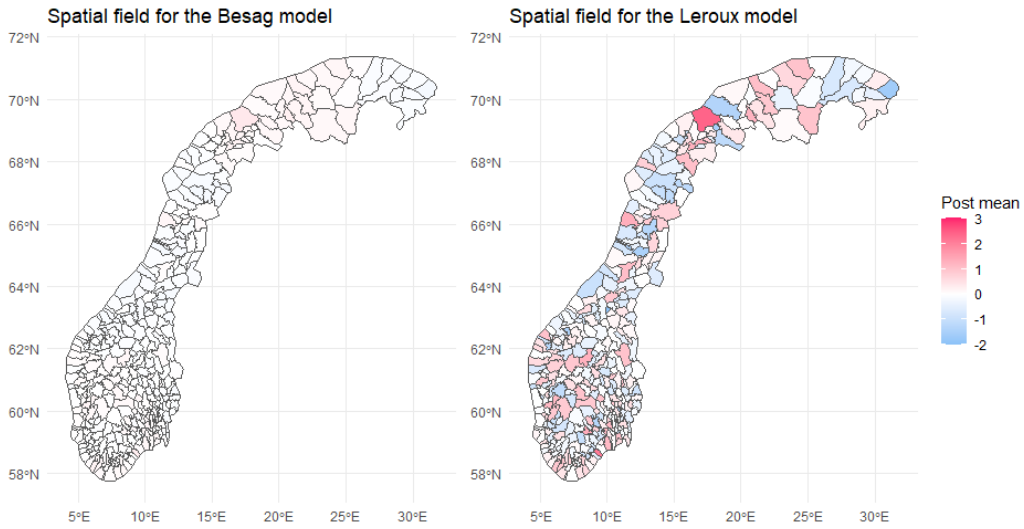

Fig. 5.16: Spatial field for a Besag model and a Leroux model.

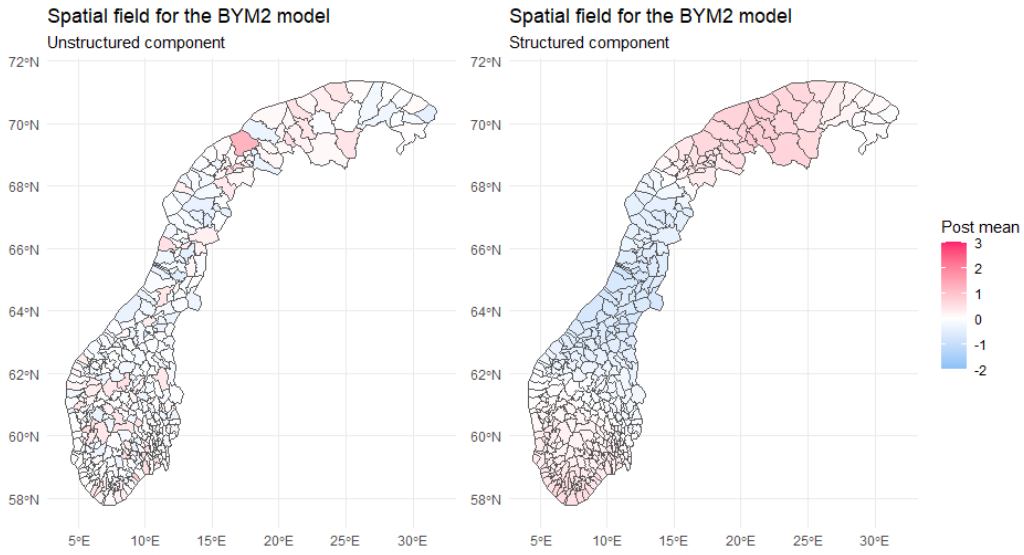

Fig. 5.17: Spatial fields for a BYM2 model.

The corresponding figures for Germany are shown in Figure 8.12 and Figure 8.13 in the Appendix.

Finally, looking at the spatial field of the structured component when changing the value for σ_0, as seen in Figure 5.18, it can be seen that for a small value like $\sigma_0 = 0.1$, the values of the posterior mean are mostly around 0, while for a higher value these

values get slightly higher. The reason for this is simply that higher values for σ_0 make the spatial field fit the data more closely, making it less smooth.

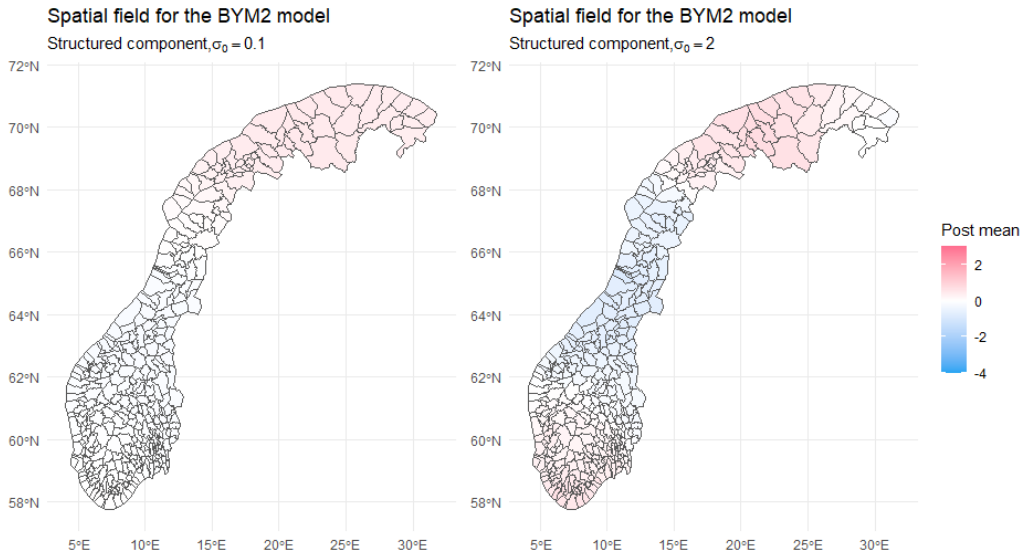

Fig. 5.18: Spatial fields for the structured component of a BYM2 model when changing the value for σ_0.

Looking at Figure 5.19, there are hardly any differences in the spatial fields for $\sigma_0 = 0, 1$ and $\sigma_0 = 2$, which again shows how robust the model is for Germany.

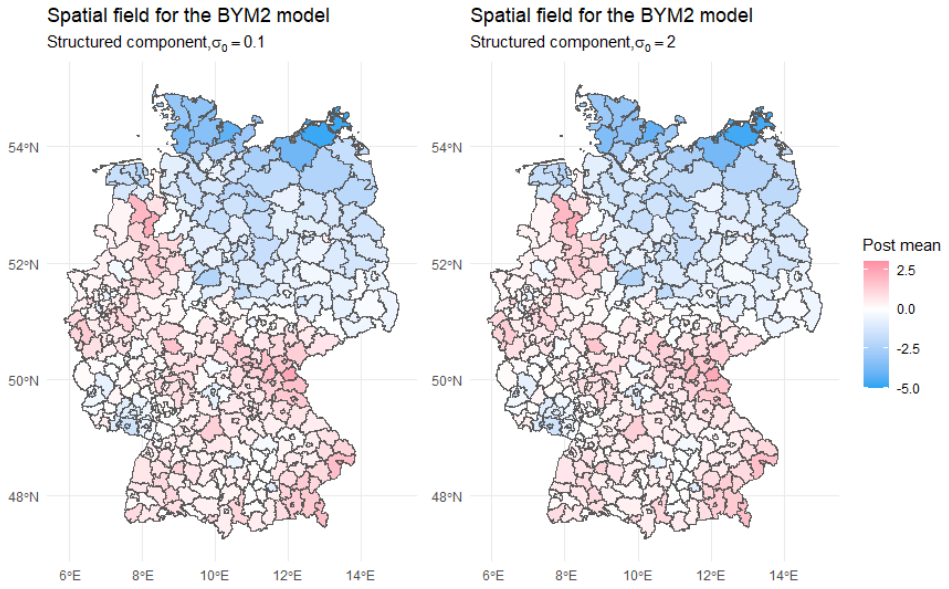

Fig. 5.19: Spatial fields for the structured component of a BYM2 model when changing the value for σ_0.

5.6 Predictive models

Another way to determine whether calculating a Bayesian spatial model is the right approach for this research question is to compare the calculated models with a completely different class of models, namely predictive models calculated using machine learning methods. The algorithms used in this chapter do not assume any kind of prior distributions or assumptions about the likelihood of the data.
The approach used to calculate these models is described in detail below.

1. Define the five base learners, which are:
   - A regression tree
   - A k-nearest neighbours (knn) algorithm
   - A neural net
   - A random forest
   - An eXtreme Gradient Boosting (xGBoost) algorithm
2. Tune each algorithm using iterated F-Racing.
3. Train each model using the parameters determined during the tuning process.
4. Make predictions on the data.

The remainder of this chapter presents the results of the models calculated for Norway and Germany and provides some insights into how different factors influence infection rates.
I will add a small theory chapter on these models as well as the theory behind the interpretable ml stuff.

5.6.1 Predictive models for Germany

A look at Table 5.13 shows that the knn algorithm, unlike the other models, overfits the training data, as evidenced by an MAE of 0. The random forest had the lowest MAE of all the predictive models for both the training and test data, but still underperformed compared to the BYM2 model from Section 5.4.1. In the remainder of this section, the random forest model is evaluated in more detail to gain an understanding of which variables influence infection rates according to the model.

Tab. 5.13: The MAE for the BYM2 model and the predictive models.

Model	MAE_{train}	MAE_{test}
BYM2	73	1030
Regression tree	2759	3249
K-nearest neighbours	0	3182
Neural net	3035	2468
Random forest	1123	1901
eXtreme Gradient Boosting	1942	3040

Looking at the feature importance of the variables in Figure 5.20, it can be seen that log trade tax is the most important feature, followed by the number of clinics, marketplaces and platforms in a municipality.
Concept behind feature importance: We measure the importance of a feature by calculating the increase in the model's prediction error after permuting the feature. A feature is "important" if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction. A feature is "unimportant" if shuffling its values leaves the model error unchanged, because in this case the model ignored the feature for the prediction.

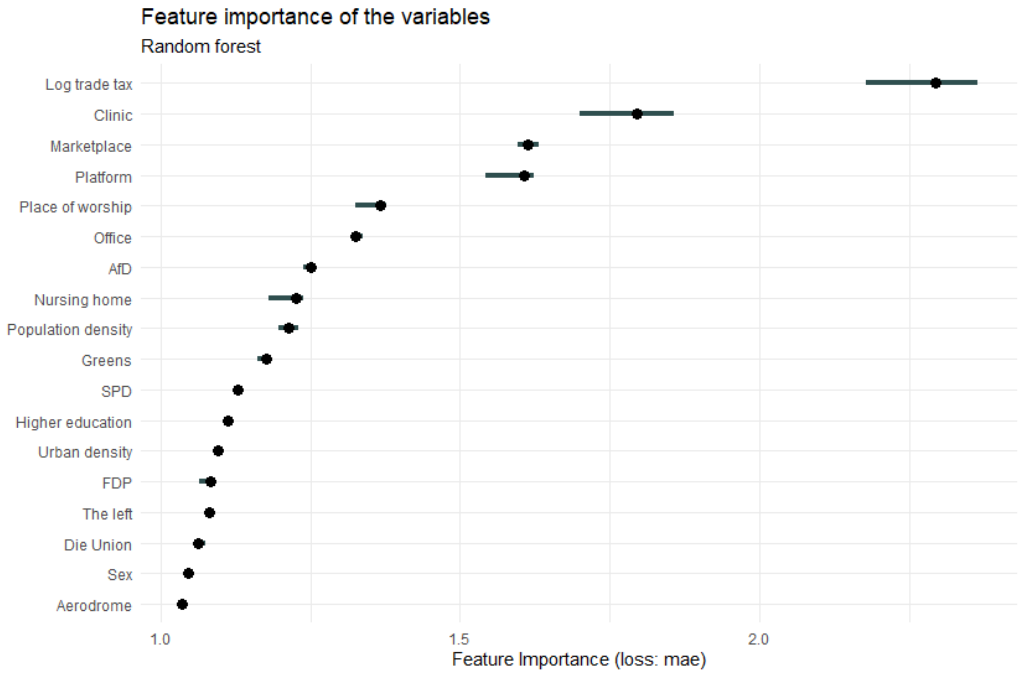

Fig. 5.20: The variable importance plots for the random forest.

To get a better idea of how the features influence the predicted outcome, the partial dependence plots for the two most important features are shown in Figure 5.21, while Figure 5.22 shows how the two most important parties, the AfD and the Greens, influence the predicted outcome. For the logarithmic trade tax, a slow increase up until 0.5 can be seen before a steeper incline sets in, which eventually turns into a linear slope. For the clinic, it starts with a steep slope before changing to a less steep linear relationship. In general, however, the higher either value is, the higher the predicted outcome. Looking at the plots for the political parties, it can be seen that the higher the share of the vote for the AfD, the higher the predicted number of infections, while a higher share of the vote for the Greens leads to a lower predicted number of infections. The same relationship was observed in the BYM2 model, as can be seen in Table 5.10, but the effect for the Greens was not significant in this case.

Concept behind partial dependence plots: The partial dependence plot (short PDP or PD plot) shows the marginal effect one or two features have on the predicted outcome of a machine learning model. A partial dependence plot can show whether the relationship between the target and a feature is linear, monotonic or more complex.

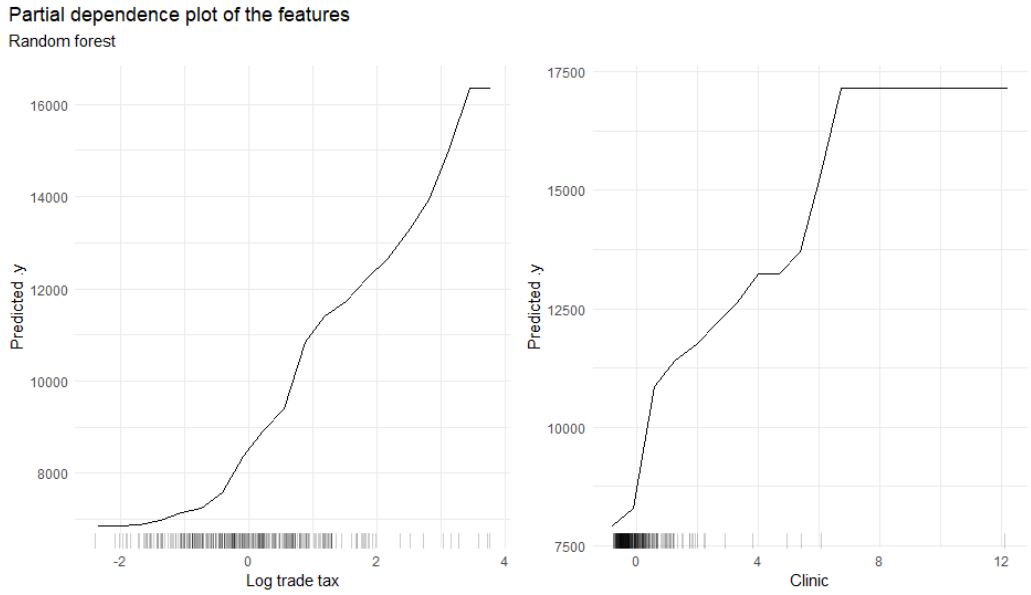

Fig. 5.21: The partial dependence plots for the logarithmic trade tax and the number of clinics.

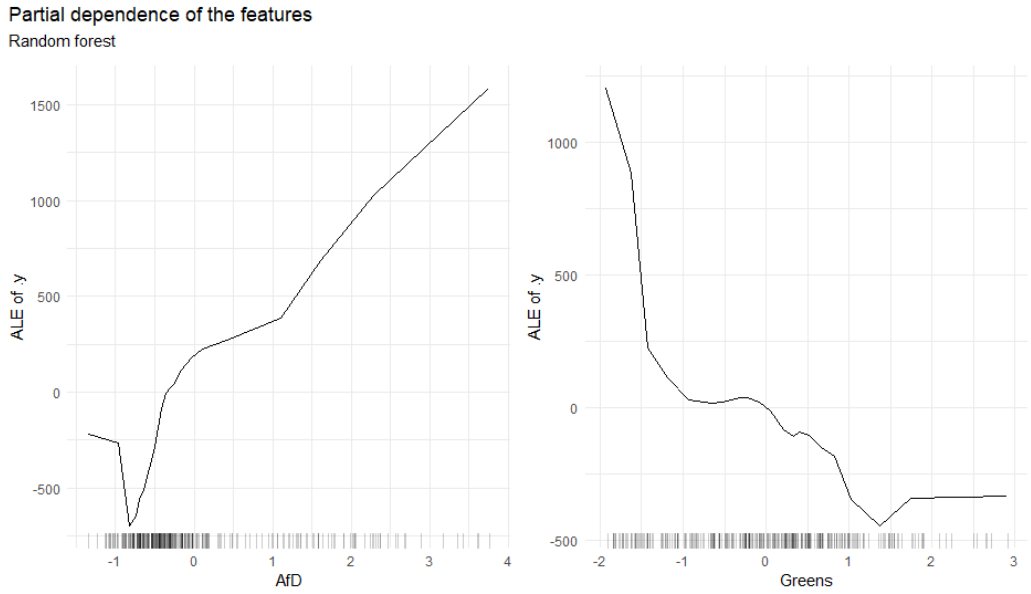

Fig. 5.22: The partial dependence plots for the share of the vote the AfD and the Greens get.

To check whether there are any interactions that could distort the relationship between the features shown in Figure 5.21 and Figure 5.22, the ICE plots of these variables in Figure 5.23 and Figure 5.24 can be looked at. All curves seem to follow the same course, so there are no obvious interactions. This means that the PDP already represents a good summary of the relationships between the displayed features and the predicted number of infections.

Concept behind ice curves: Individual Conditional Expectation (ICE) plots display one line per instance that shows how the instance's prediction changes when a feature changes. The partial dependence plot for the average effect of a feature is a global method because it does not focus on specific instances, but on an overall average. The equivalent to a PDP for individual data instances is called individual conditional expectation (ICE) plot. What is the point of looking at individual expectations instead of partial dependencies? Partial dependence plots can obscure a heterogeneous relationship created by interactions.

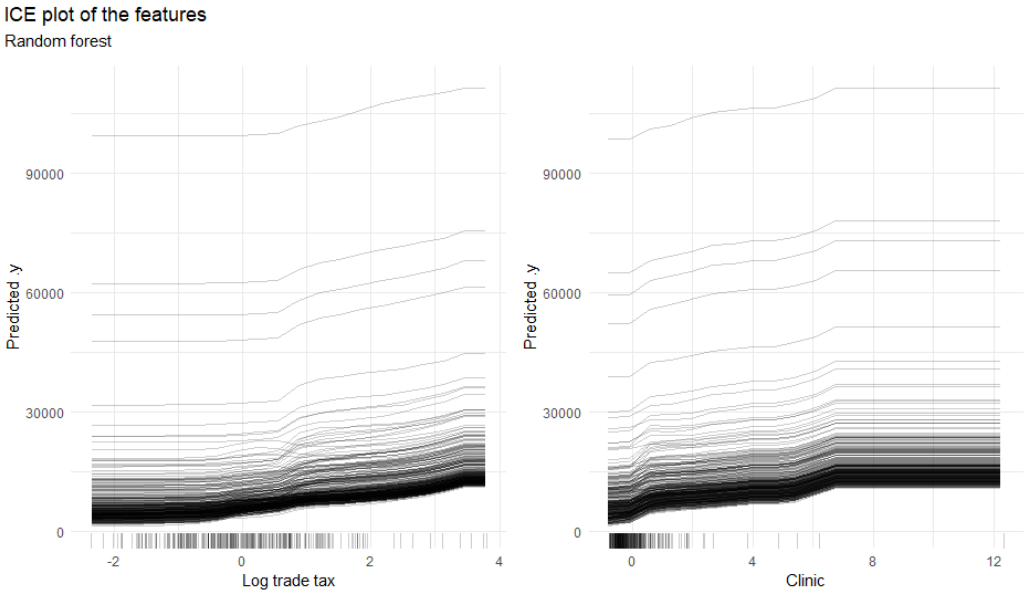

Fig. 5.23: The individual conditional expectation for the logarithmic trade tax and the number of clinics.

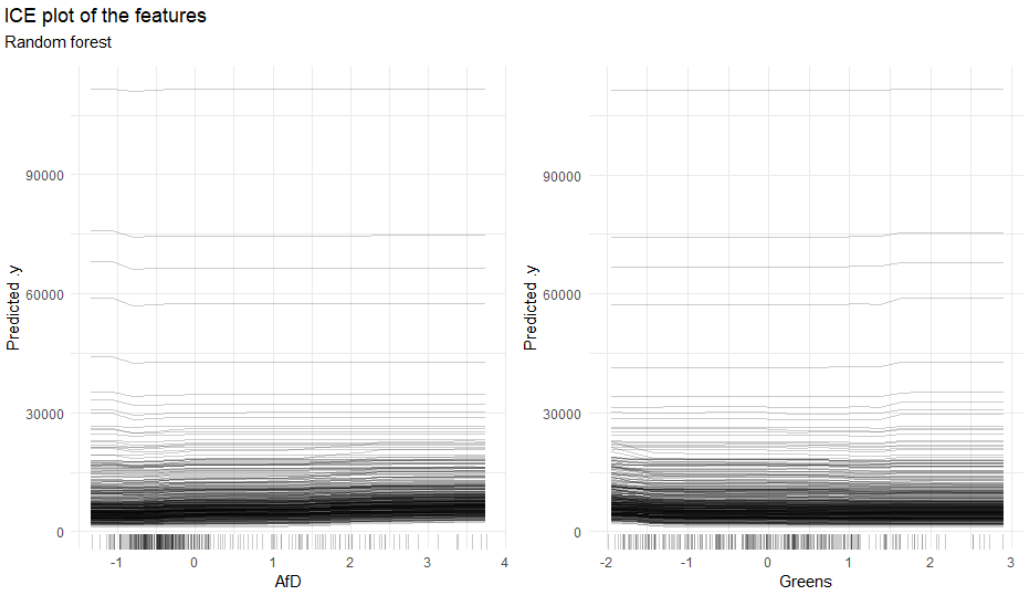

Fig. 5.24: The individual conditional expectation for the share of the vote the AfD and the Greens get.

Lastly, looking at the Shapley values for the region of Hannover in Figure 5.25, the prediction is nearly 50,000 above the average prediction. The number of marketplaces in the region increased the prediction the most, followed by the logarithmic trade tax and the number of offices in the region. For the city of Munich

it was the same three features, however this time the logarithmic trade tax increased the prediction the most, followed by the number of offices and the number of marketplaces. For Munich, the actual prediction was around 33,000 above the average prediction.
Concept behind the shapley value: The Shapley value of a feature value is not the difference of the predicted value after removing the feature from the model training. The interpretation of the Shapley value is: Given the current set of feature values, the contribution of a feature value to the difference between the actual prediction and the mean prediction is the estimated Shapley value.

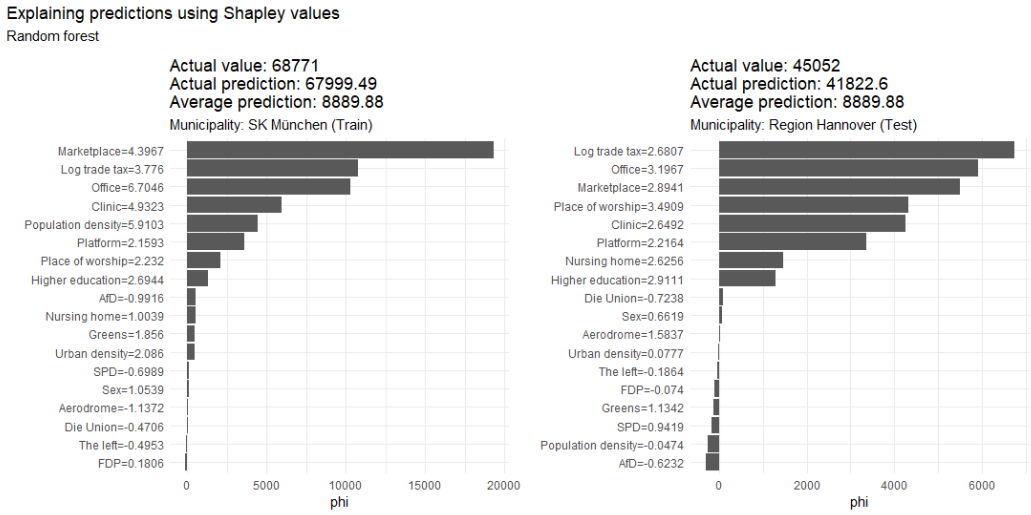

Fig. 5.25: Shapley values for the cities of Munich and Hannover.

5.6.2 Predictive models for Norway

Looking at the performance of the different algorithms in Table 5.14, the knn algorithm did not completely overfit the training data this time and proved to be the algorithm with the second best performance in terms of both MAEs, not counting the BYM2 model. Again, the random forest model had the lowest MAE of all the predictive models for the training and test data, but performed worse than the BYM2 model from Section 5.4.2. Therefore, the random forest model is again evaluated in more detail to gain an understanding of which variables influence the infection rates according to the model.

Tab. 5.14: The MAE for the BYM2 model and the predictive models.

Model	MAE_{train}	MAE_{test}
BYM2	171	78
Regression tree	353	238
K-nearest neighbours	210	155
Neural net	361	223
Random forest	162	137
eXtreme Gradient Boosting	248	188

Looking at the feature importance of the variables in Figure 5.26, the most important feature is the number of places of worship in a municipality, followed by the number of offices and urban density.

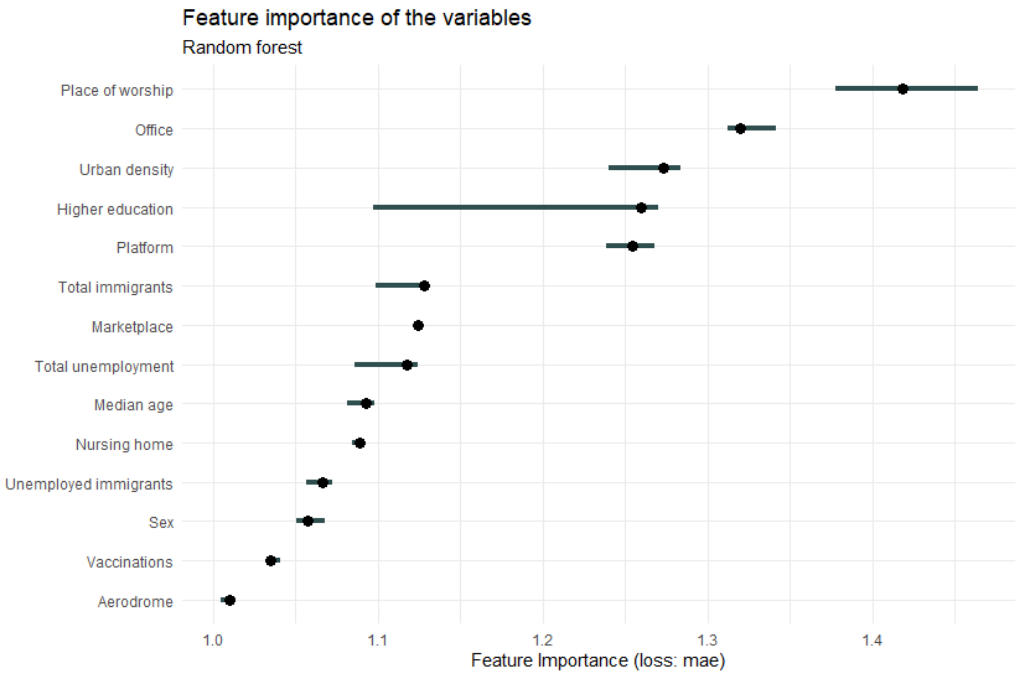

Fig. 5.26: The variable importance plots for the random forest.

Figure 5.27 shows how the predicted number of infections changes as the number of places of worship and offices in a community increases. For places of worship, a slow increase is seen until 3, followed by a steep increase that continues until 7 and then ends in a plateau. For the number of offices, it is a slow increase until about 2.5, followed by a steep increase until about 6 before ending on a plateau.

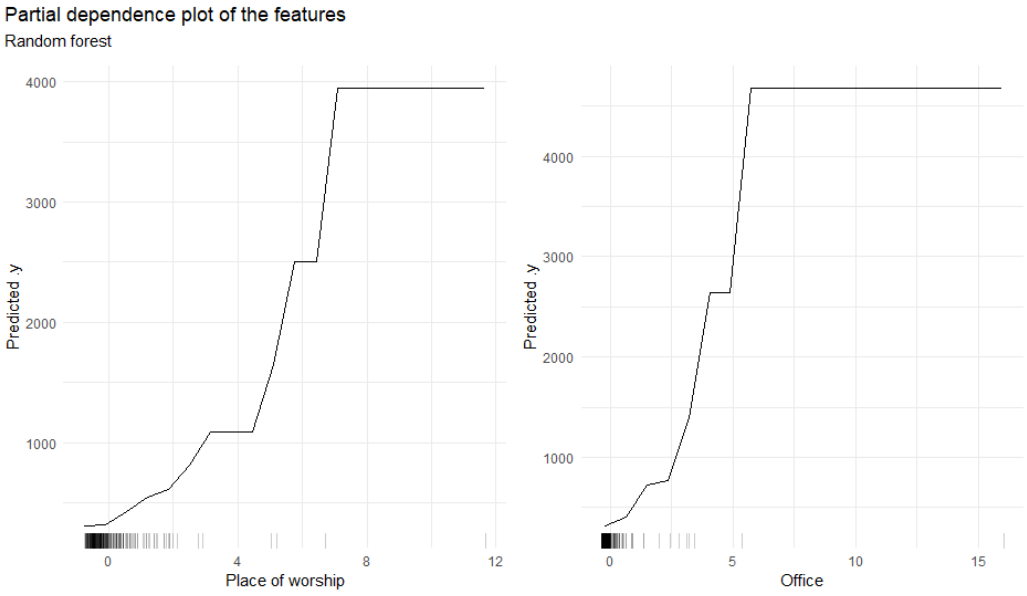

Fig. 5.27: The partial dependence plots for the number of places of worship and the number of offices.

The curves display in the ICE plot in Figure 5.27 all follow the same course, so there are no obvious interactions. Therefore, again, the PDP already represents a good summary of the relationships between the displayed features and the predicted number of infections.

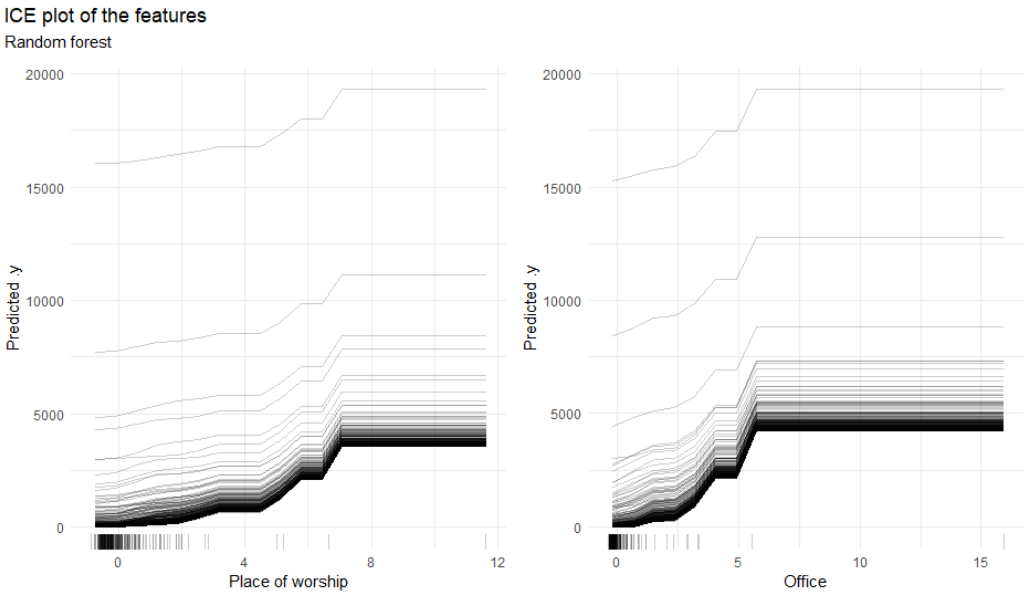

Fig. 5.28: The individual conditional expectation for the logarithmic trade tax and the number of clinics.

Finally, looking at the Shapley values for Tromsø municipality in Figure 5.29, the prediction is about 1,100 above the average prediction. The feature that increases the prediction the most is the number of nursing homes in the municipality, while the number of offices is the feature that decreases the prediction the most. For Nordre Follo, the prediction is about 1,000 above the average prediction, with urban density and the number of platforms being the two features that increase the prediction the most, while the number of buildings with higher education is the feature that decreases the prediction the most.

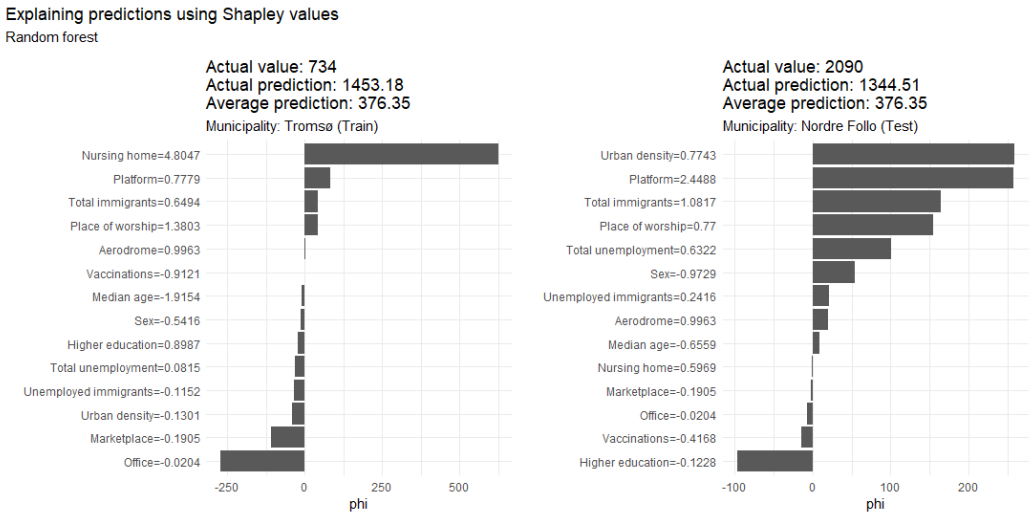

Fig. 5.29: Shapley values for the municipalties of Tromsø and Nordre Follo.

5.7 Temporal models

I will write some kind of introduction here relating to temporal effects such as government measures or waves of covid. i will also write sth about baseline temporal models used for comparison

5.7.1 Choice of Likelihood

As with the non-temporal models calculated in Section 5.3 and Section 5.4, a probability must first be found that describes the distribution of the daily number of infections in each country. Returning to the Cullen and Frey graph, Figure 5.30 and Figure 5.31 show that the observations do not actually follow a normal, negative binomial or Poisson distribution.

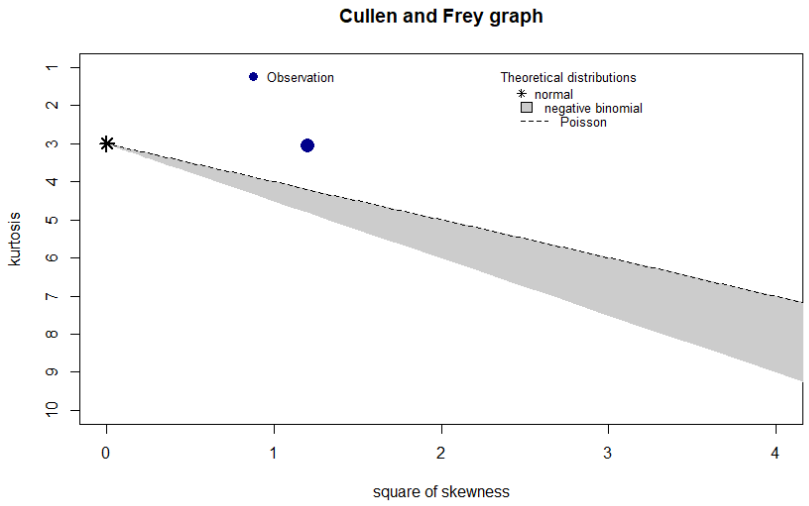

Fig. 5.30: The Cullen and Frey graph for Germany

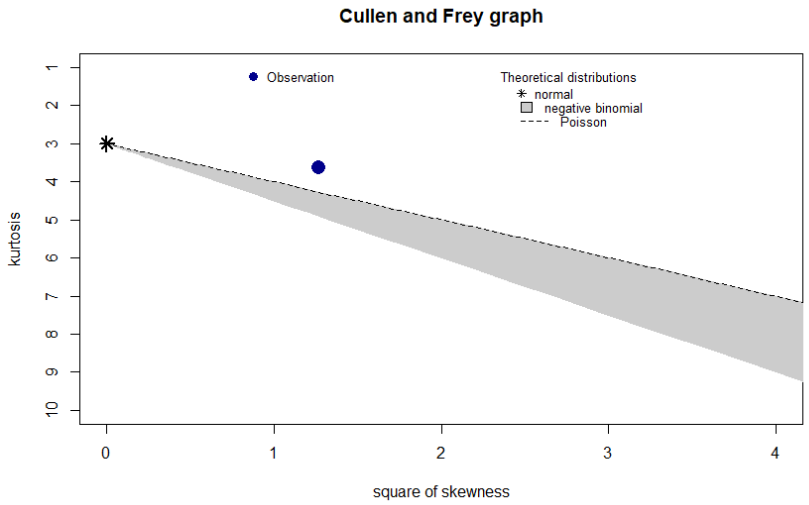

Fig. 5.31: The Cullen and Frey graph for Norway

After fitting all three distributions to the data using the maximum likelihood method, QQ-plots and the empirical and theoretical cumulative density function can be used to help decide which likelihood to choose. The QQ-plots in Figure 5.32 and Figure 5.33 clearly show that there is no linear relationship between the theoretical quantiles and the sample quantiles for both distributions, while the empirical CDF roughly follows the theoretical CDF in both cases. The plots for the Poisson distribution as well as all plots for the Norwegian municipalities are in Section 8.3.1.2 and section 8.3.2.1 in the Appendix.

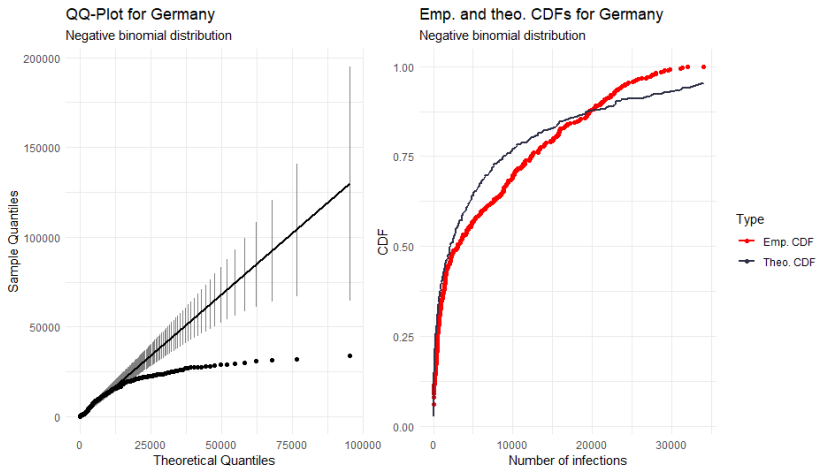

Fig. 5.32: A negative binomial fit to the number of cases in German municipalities

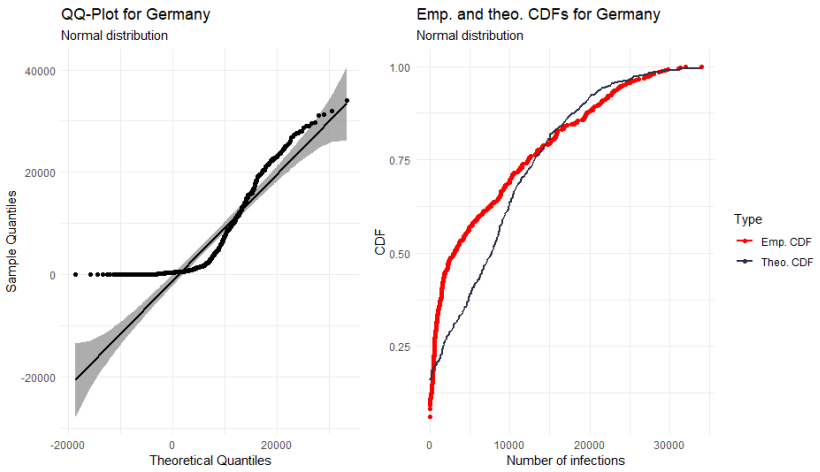

Fig. 5.33: A normal fit to the number of cases in German municipalities

Looking at the AIC for all the distribution fits shown in Table 5.15, the negative binomial distribution again seems to be the best fit for the data. For Germany the mean and standard deviation are 7274 and 8429 respectively, while for Norway these moments are 256 and 266. This again explains the poor fit of the Poisson distribution, since an equal mean and variance are assumed for the Poisson distribution.

Tab. 5.15: The AIC for different distributions for Germany and Norway

Country	Distribution	AIC
Germany	Normal	10461
Germany	Poisson	4762604
Germany	Negative Binomial	9462
Norway	Normal	6600
Norway	Poisson	131469
Norway	Negative Binomial	6076

Overlaying the number of infections with a negative binomial distribution and a normal distribution using the estimated parameters, as shown in Figure 5.34, reinforces the assumption of the negative binomial distribution, which is then used as the assumed probability of the data for the models calculated in this chapter. The graph for Norway can be seen in Figure 8.9.

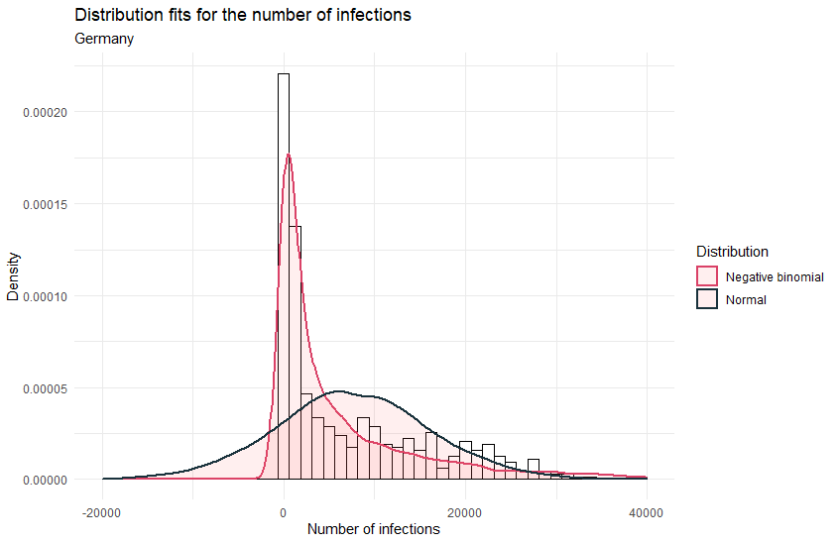

Fig. 5.34: Histogram for the number of cases in German municipalities with a normal and a negative binomial distribution overlayed.

5.7.2 Temporal models for Germany

For Germany, 502 data points are available, starting on 7 January 2020 and ending on 22 May 2021. Of these 502 data points, the first 482 are used for model training, while the last 20 are used for the analysis. Due to the temporal effect, a random split would not make sense as this may lead to a look-ahead bias as the model would use information that was not yet known or available during the analysed period. After removing variables using the VIF method already used for the non-temporal models, the trained model used the following variables:

- Measures taken in relation to contact tracing.
- Measures taken in relation to restrictions on internal movement.
- Measures taken in relation to the use of face masks.
- Measures taken in relation to the requirement to remain at home.
- The mobility at workplaces.
- The season of the year.
- The relative frequency of the main strain of Covid-19.
- The relative frequency of variant 20E.
- The relative abundance of variant 20L, also known as B.1.1.7

A second-order random walk is chosen for the temporal effect. The performance measures of this model and the temporal baseline models are shown in Table 5.16. While the AR(1) process has by far the lowest MAE in training, it has the second highest MAE in testing. The temporal model, besides the AR(1), has the lowest MAE during training and the overall lowest MAE during testing, indicating that the model fits the data best.

Tab. 5.16: The performance measures for different types of temporal models for Germany.

Model	DIC	WAIC	CPO	$\text{MAE}_{\text{train}}$	MAE_{test}
Only date as covariate	8815	8816	-4408	4915	27450
Random walk of second order	7522	7519	-3760	2009	6180
AR(1) process	5950	5924	-3877	18	7524
Temporal model	7524	7521	-3777	1931	5821

Looking at the predicted number of infections from the temporal model in Figure 5.35, it can be seen that the predictions for the training data follow roughly the

same path as the actual numbers, as three waves can be clearly seen. For the test data depicted more clearly in Figure 5.36, the predicted number of cases shows a very slow increase and it can also be seen that the "weekend effect", which means that fewer infections are reported on the weekend, has not been captured. Furthermore, the more recent a data point is in the test set, the greater the uncertainty associated with the prediction.

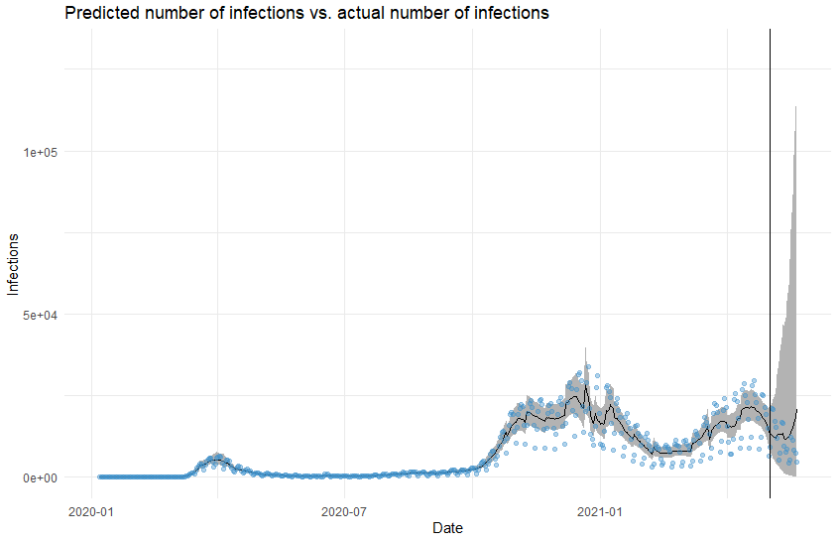

Fig. 5.35: The predicted number of infections in Germany according to the temporal model. The vertical line indicates where the test data begins.

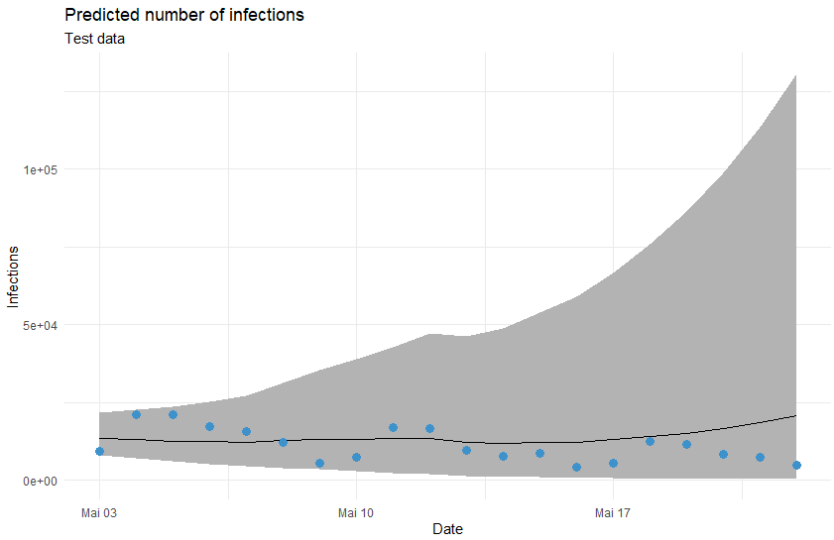

Fig. 5.36: The predicted number of infections in Germany according to the temporal model.

Another way to look at the predictions is to compare the 7-day incidence of the actual number of infections and the predicted number of infections, as daily variations such as the "weekend effect" should be smoothed out. Figure 5.37 shows that the predicted 7-day incidence again follows the actual data very well, even a little for the test data. However, this is because for the first few days of test data, the incidence depends on the predicted number of infections for a few days of training data. Once it is only test data, the 7-day decline slowly stops before a steep slope starts, in contrast to the actual 7-day incidence, which is still falling.

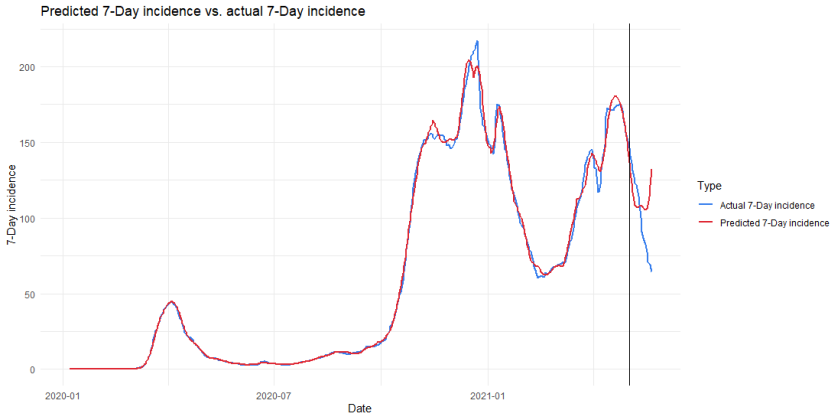

Fig. 5.37: The 7-day incidence of the actual number of infections and the predicted number of infections. The vertical line indicates where the test data begins.

Looking at the posterior temporal trend for Germany in Figure 5.38, the first two waves are clearly visible, with the second wave having two peaks, the first in early November 2020 and the second in mid-December 2020. The peak of the first and third waves is in late March 2020 and mid-April 2021, respectively. After the third peak, a steep increase sets in, as with the 7-day incidence in Figure 5.37, which is clearly not the case when looking at the actual data.

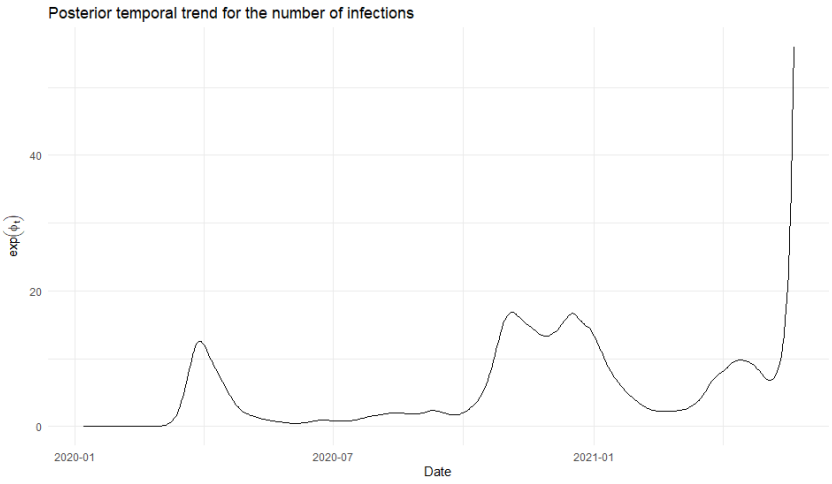

Fig. 5.38: The posterior temporal trend for the number of infections.

The covariates of the model are presented in Table 5.17. Apart from the intercept, the only significant effect is the mobility at workplaces.

Tab. 5.17: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	mean_p	$\exp(\text{mean}_p)$	$\exp(\text{q0025}_p)$	$\exp(\text{q0975}_p)$	sig.
(Intercept)	-1.592	0.228	0.079	0.518	*
Variant 20L	0.658	2.117	0.836	4.476	
Variant 20E	0.320	1.407	0.926	2.059	
Restrictions internal movement	0.235	1.298	0.813	1.970	
Movement restricted Workplace mobility	0.168	1.186	1.029	1.359	*
Season winter	0.112	1.161	0.653	1.918	
Season spring	0.034	1.082	0.574	1.858	
Stay home requirements	0.022	1.044	0.687	1.517	
Recommended Facial coverings	0.018	1.070	0.551	1.884	
Recommended Facial coverings	-0.008	1.109	0.398	2.506	
Required in some public spaces					
Stay home requirements	-0.014	1.009	0.650	1.492	
Required (exc. essent.)					
Restrictions internal movement	-0.021	1.004	0.632	1.522	
Recommended Contact tracing	-0.047	0.976	0.630	1.454	
Limited tracing					
Main variant	-0.134	0.920	0.470	1.622	
Season summer	-0.341	0.739	0.412	1.216	
Contact tracing No tracing	-2.095	0.228	0.010	1.035	

5.7.3 Temporal models for Norway

For Norway, 473 data points are available, starting on 5 February 2020 and ending on 22 May 2021. Of these 473 data points, the first 453 are used for model training, while the last 20 are used for the analysis. After the removal of some variables, the following features remained:

- Measures taken in relation to restrictions on internal movement.
- Measures taken in relation to the requirement to remain at home.

- The mobility at workplaces.
- The mobility at groceries and pharmacies.
- The season of the year.
- The relative frequency of the main strain of Covid-19.
- The relative frequency of variant 20E.
- The relative abundance of variant 20L, also known as B.1.1.7

Again, a second-order random walk is chosen for the temporal effect. The performance measures of this model and the temporal baseline models are shown in Table 5.18. The lowest MAE during training was again observed for the AR(1) process, followed by the temporal model and the second-order random walk. The temporal model then showed the best performance in terms of MAE for the test set, followed by the second-order random walk and the AR(1) process.

Tab. 5.18: The performance measures for different types of temporal models for Norway.

Model	DIC	WAIC	CPO	MAE_{train}	MAE_{test}
Only date as covariate	5567	5568	-2784	130	446
Random walk of second order	4603	4603	-2387	55	119
AR(1) process	4602	4599	-2315	38	110
Temporal model	4613	4613	-2432	53	100

For the predicted number of infections shown in Figure 5.39, 4 peaks are seen for the training data, just like for the actual data. The rest of the predicted infection numbers also follow the actual infection numbers very closely, which is not too surprising given the low MAE shown in Table 5.18. For the test data shown in Figure 5.40, it is very similar to the graph for Germany in Figure 5.36, as the "weekend effect" is not captured, a slow increase in the predicted infection numbers can be seen, as well as an increasing uncertainty in the predictions.

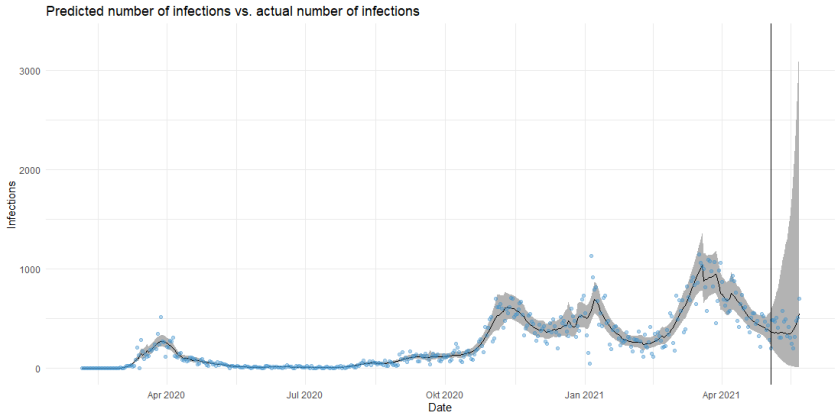

Fig. 5.39: The predicted number of infections in Norway according to the temporal model. The vertical line indicates where the test data begins.

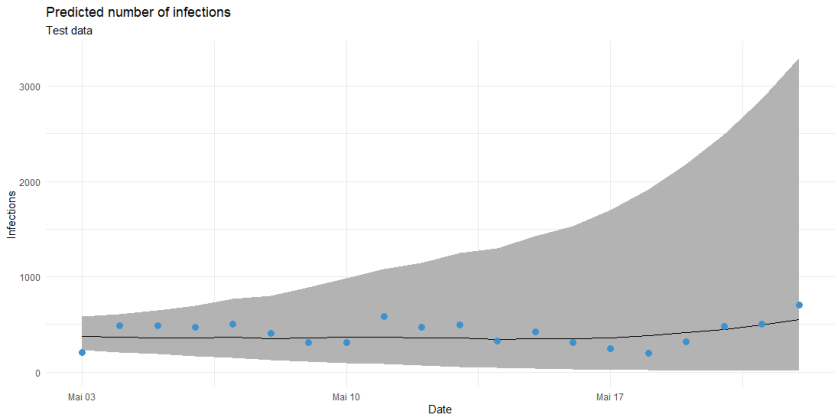

Fig. 5.40: The predicted number of infections in Norway according to the temporal model.

Comparing the 7-day incidence of the predicted data and the actual data in Figure 5.41, the incidences of the training data and the actual data follow the same pattern. However, the predicted 7-day incidence for the test data initially falls before rising, while the actual 7-day incidence initially rises and then falls before rising again. Although still not ideal, the predicted incidence for Norway looks slightly better than that for Germany, which can be seen in Figure 5.37.

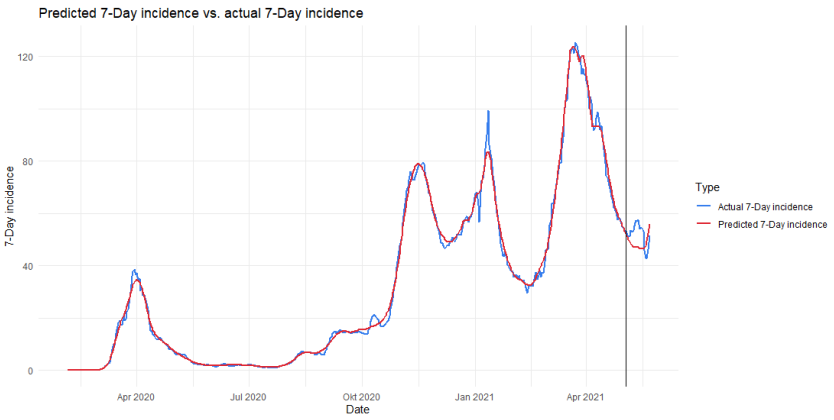

Fig. 5.41: The 7-day incidence of the actual number of infections and the predicted number of infections. The vertical line indicates where the test data begins.

The posterior temporal trend for Norway in Figure 5.42 clearly shows all three waves, including the double peak of the second wave. It can also be seen that the third wave is clearly the worst in terms of infection numbers. On the other hand, however, the first and second waves appear to be equally bad in this graphic, while in reality the second wave in Norway was worse than the first. Just as with Germany in Figure 5.38, a steep increase sets in after the third wave.

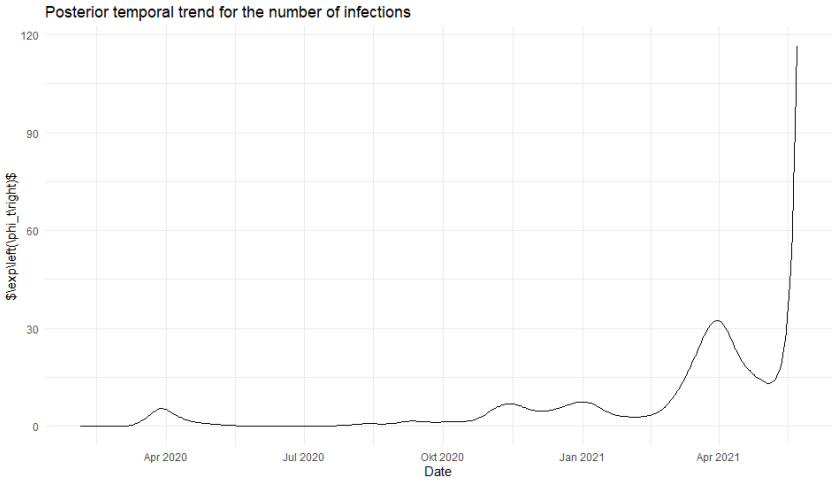

Fig. 5.42: The posterior temporal trend for the number of infections.

The covariates of the model are presented in Table 5.19. Apart from the intercept, the only significant effect is the mobility at workplaces.

Tab. 5.19: The fixed effects for the model. Values are rounded. A * denotes a significant effect.

Variable	mean_p	$\exp(\text{mean}_p)$	$\exp(\text{q0025}_p)$	$\exp(\text{q0975}_p)$	sig.
(Intercept)	-0.926	0.408	0.249	0.631	*
Workplace mobility	0.179	1.201	1.011	1.417	*
Season winter	0.109	1.159	0.645	1.924	
Main variant	0.056	1.073	0.760	1.482	
Variant 20E	0.041	1.043	0.942	1.149	
Restrictions internal movement	0.012	1.063	0.545	1.854	
Recommended Stay home requirements	-0.058	1.033	0.413	2.163	
Recommended Mobility grocery and pharmacy	-0.081	0.923	0.831	1.022	
Variant 20L	-0.086	0.963	0.499	1.684	
Season spring	-0.089	0.962	0.489	1.706	
season summer	-0.178	0.886	0.436	1.624	
Restrictions internal movement	-0.209	0.842	0.477	1.384	
Movement restricted					

6 Discussion

After the models have been calculated, the next step is to critically examine and evaluate the results. The evaluation of the spatial models follows a simple scheme. First, a brief look is taken at the models without the spatial component, before the spatial models are reviewed and compared with the non-spatial models. Next, it is examined which factors significantly influence the risk of infection and why these factors might have an impact. For this, the coefficient of the BYM2 model are analysed. Finally, a look at area-specific risk is taken to see which regions in a given country are most at risk.

6.1 Discussion of the (Non)-Spatial Models

6.1.1 Discussion of the (Non)-Spatial Models for Norway

The non-spatial model for Norway identified five significant effects:

- The number of unemployed immigrants
- The total number of immigrants
- The urban density
- The number of aerodromes
- The proportion of females

Moreover, the intercept was significant.
Comparing the spatial models with the non-spatial models using Table 5.11, it can be seen that BYM2, Besag and the non-spatial model performed almost equally well in terms of the DIC and WAIC, while the Leroux model performed best in terms of these two metrics. In terms of CPO, all spatial models performed better than the non-spatial model, with the Leroux model again performing best. However, the MAE shows that the non-spatial model has the best predictive performance, ahead of the Besag model, the BYM2 model and the Leroux model. This shows that the Leroux model overfits the training data more than the other models, a characteristic that

126

can also be seen in Figure 5.12 and Figure 5.13. The fact that the non-spatial model performs better than the spatial models is an indication that the spatial effect in Norway is not strong and that a different class of models may be better suited to identify critical factors affecting infection rates. Table 5.12 also shows that adding the spatial effect did not suddenly cause any variables to become significant, while all variables, including the intercept, that were significant in the non-spatial model remained significant. Nevertheless, significant effects were found and have to be discussed.

The exponentiated intercept implies a risk rate of -60.8% across Norway. The factor that influences the relative risk most is urban density, where an increase of 1 standard deviation leads to an increase in the relative risk of 22.5%. The relationship between a higher number of residential buildings in a given area and the number of infections is probably due to the fact that with an increasing number of residential buildings comes an increasing number of inhabitants. A look at the Bravais-Pearson correlation coefficient confirms this assumption with $\rho = 0.7070$. Furthermore, the correlation between urban density and population density is 0.8037. This is convenient since studies analysing the relationship between urban density and Covid-19 usually define urban density as the number of inhabitants per square kilometre. Because of the positive relationship between urban density and population density, urban density acts as a proxy for population density in these models, so the relationship between higher population density and higher case numbers should also account for the higher infection rates in areas with higher urban density. According to Jamshidi et al. (2020) and Whittle & Diaz-Artiles (2020), higher urban/population density and higher infection numbers are due to a decrease in proximity between people and an increase in the likelihood of interpersonal contact (Jamshidi2020global, Whittle2020ecological). Sigler et al. (2020) also found that dense urban environments provide more opportunities for the virus to spread, but that higher density had a stronger effect earlier and decreased in strength and importance over time(Sigler et al., 2020).

Factors related to immigration also play a crucial role in the relative risk of developing Covid-19. A one standard deviation increase in the number of unemployed immigrants leads to a 20.8% increase in risk and a one standard deviation increase in the total number of immigrants leads to a 20.3% increase in risk. Unfortunately, there are no studies analysing the relationship between unemployed immigrants and the risk of developing Covid-19, as studies mostly focus on the relationship between unemployment in general and the risk of disease. Unemployment, however, turned out to be non-significant in the BYM2 model. Looking at older studies, Elkeles and Seifert (1996) conducted a longitudinal study that looked at the unemployment status and health of labour migrants in Germany. They found that immigrants were

often employed in jobs with higher health risks and stress and therefore had poorer health (Elkeles & Seifert, 1996). Sia et al. (2019) analysed the association between immigration status and unemployment and men's and women's health using data from the Canadian Health Measures Survey. The study provided evidence of biological associations between unemployment and the likelihood of common chronic diseases, inflammation and possible malnutrition, with unemployed immigrants, and particularly unemployed immigrant women, being more prone to chronic diseases (Sia et al., 2019). Thus, there is precedent for an association between unemployment and immigrants and a higher risk of disease, however, further research would need to be conducted to analyse the association between unemployed immigrants and the risk of Covid-19 infection.

A higher risk among immigrants in Norway was already found in a study by Indseth et al. (2020). The reasons given are barriers to adequate information due to low health literacy in certain groups and misconceptions about Covid-19 or test criteria, as well as other socio-economic and environmental factors (Indseth et al., 2020). Therefore, the results of the BYM2 model are consistent with this study.

Looking at the factors that reduce the risk of infection for Covid-19, for airports, a 1 standard deviation increase leads to a 12.2% reduction in risk and a 1 standard deviation increase in the proportion of women leads to an 18.2% reduction in the risk of infection. Neither result is consistent with current research. Gaskin et al. (2021) found that counties in the United States closer to an airport have higher Covid-19 rates than counties further from an airport (Gaskin et al., 2021). Daon et al. (2020) found that different airports have different risks of being a source of an outbreak, especially airports in low- to middle-income countries (Daon et al., 2020). Many studies have investigated the relationship between gender and Covid-19 and most come to the same conclusion. While the risk of infection is the same in men and women, men tend to experience a higher severity and mortality rate for Covid-19 compared to women (Gausman & Langer, 2020; Kopel et al., 2020; Mukherjee & Pahan, 2021; Spagnolo et al., 2020).

A look at the relative risk of infection in Figure 6.1 shows that the relative risk is below 1 in most of Norway, which is not too surprising considering that Norway has managed the pandemic quite well so far. The two municipalities with the highest relative risk are Iveland and Ålesund. However, a look at the posterior probability in Figure 6.2 shows that the posterior probability of the risk being greater than 1 is less than 0.5 for Ålesund (0.47) and less than 0.75 for Iveland (0.62). Looking at the log posterior mean of the random effects, there is an increased risk in the regions around Oslo and in large parts of southern and central Norway, while the risk tends to be lower in the northern parts of the country.

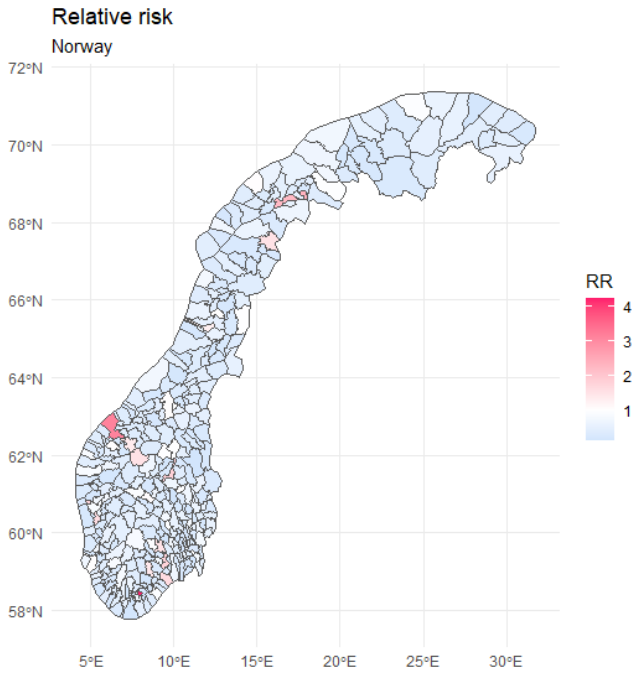

Fig. 6.1: Relative risk of contracting Covid-19 in Norway.

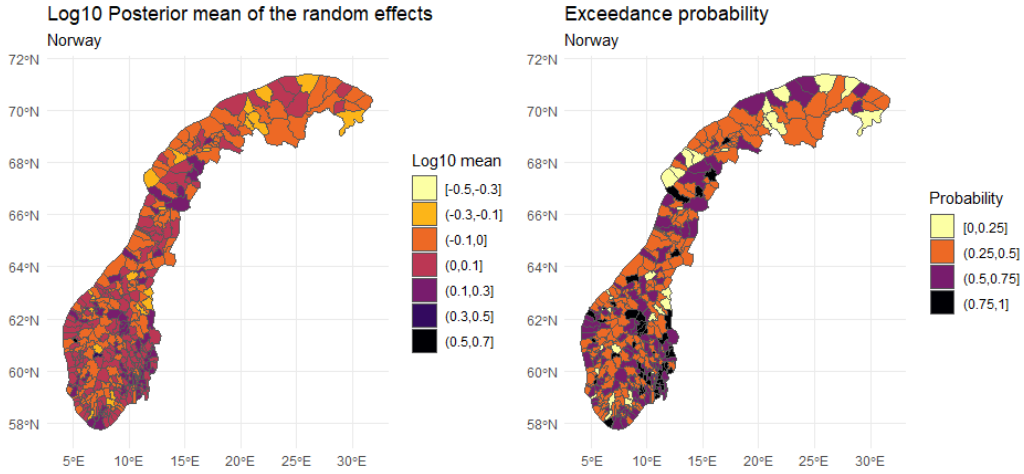

Fig. 6.2: Posterior mean of the municipality-specific relative risks $\zeta = \exp(\xi)$ compared with the whole of Norway (left) and posterior probability $\mathbb{P}(\zeta_i > 1|\boldsymbol{y})$

6.1.2 Discussion of the (Non)-Spatial Models for Germany

In the non-spatial model for Germany, six coefficients were significant, in addition to the intercept:

- The percentage of the vote for the right-wing populist AfD
- The population density
- The logarithmic trade tax
- The percentage of the vote for the SPD
- The percentage of the vote for the left-wing party "Die Linke"
- The percentage of the vote for the Green party

A look at Table 5.9 shows that the spatial models outperform the non-spatial model in terms of the DIC and the WAIC, while all models perform about equally well in terms of the CPO. The predictive performance was best for the BYM2 model, just ahead of the Besag model, as indicated by the lowest values for the MAE. When adding the spatial term for the BYM2 model, several variables lose their significance, namely the percentage of the vote for the SPD, "Die Linke" and the Greens, while the intercept, the percentage of the vote for the AfD, population density and the logarithmic trade tax remain significant.
The exponentiated intercept implies a risk rate of -6.8% across Germany. The greatest influence on the relative risk was determined to be the percentage of the vote for the far-right AfD. An increase in this variable by 1 standard deviation leads to an increase in the relative risk by 24.8%. The AfD openly criticises the measures taken by the government in Germany to prevent the spread of Covid-19, which leads to a large proportion of the party's voters not taking the measures seriously and refusing to keep a safe distance or wear a mask in public spaces. Several studies have taken a look at the response of right-wing parties to the pandemic and how people who vote for these types of parties have reacted. Wondreys and Mudde (2020) point out that these parties were quick to warn about the virus, but once cases spiked, they criticised the measures taken to contain the spread of the virus. They also noted that right-wing parties often rejected the measures proposed by the leading parties because they themselves were part of the opposition (Wondreys & Mudde, 2020), as is the case with the AfD in Germany. Vieten (2020) shows how the far-right mobilises people for anti-hygiene or anti-lockdown protests and how this is used to normalise the global far-right (Vieten, 2020). Eberl et al. (2020) analysed data from the Austrian Corona Panel Project to test whether populism

attitudes and belief in conspiracy theories related to Covid-19 are correlated. Using structural equation modelling, they show that populism indirectly influences Covid-19 conspiracy beliefs through trust in political and scientific institutions. They find that populist attitudes have a negative correlation with trust in the government and the parliament.Furthermore, they find that higher populist attitudes are negatively correlated with trust in science, a factor that reduces belief in Covid-19 conspiracy theories (Eberl et al., 2020). Finally, Farias and Pilati (2020) conducted a study in Brazil to predict social distancing violation intention and past non-compliance during the Covid-19 pandemic, controlling for the effects of interolerence of insecurity and socio-demographic variables. Their results included that individuals who support right-wing parties are more likely to violate social distancing measures (Farias & Pilati, 2020).

An increase in population density by 1 standard deviation leads to a risk increase of 11.2%. Reasons why a higher population density correlates positively with the number of infections have already been discussed in Section 6.1.1.

The last significant effect is the logarithmic trade tax with an increase of 1 standard deviation leading to a 6.7% increase in risk. There are no studies that specifically analyse the relationship between the trade tax and the risk of infection, but some studies have analysed infection rates in different spatial areas while controlling for factors such as income. In general, a negative relationship was found between areas with higher income and infection rates, meaning that areas with higher income had fewer cases of Covid-19. Cordes and Castro (2020) found that postcode areas in New York City with a low proportion of positive tests had higher incomes and also tested less compared to lower income areas. They also found that people in lower income areas were more likely to be without health insurance (Cordes & Castro, 2020). Coven and Gupta (2020) found that New York City residents who come from wealthier neighbourhoods are more likely to flee the city, and that people who live in low-income neighbourhoods are more likely to have frontline occupations and visit retail shops more often, increasing their exposure to Covid-19 (Coven & Gupta, 2020). The situation seems to be the same in Europe, as Baena et al. (2020) found that districts in Barcelona with a lower average income had a higher Covid-19 incidence. They found that the incidence in the district with the lowest income was 2.5 times higher than in the district with the highest income (Baena-Dıez et al., 2020). Again, the results of the BYM2 model are not consistent with current research.

The relative risk of infection shown in Figure 6.3 is highest in eastern Germany, more specifically in the federal state of Saxony. Saxony has established itself as the political stronghold of the AfD in recent years, which has even led to the ruling party, the Union, moving further to the right on the political spectrum. In addition to

Saxony, the traditionally highly conservative Bavaria and the populous Ruhr region also have a relative risk of over 1. Figure 6.4 shows that for most of these regions the posterior mean of the random effects is also over 1, mostly with a posterior probability of at least 0.75.

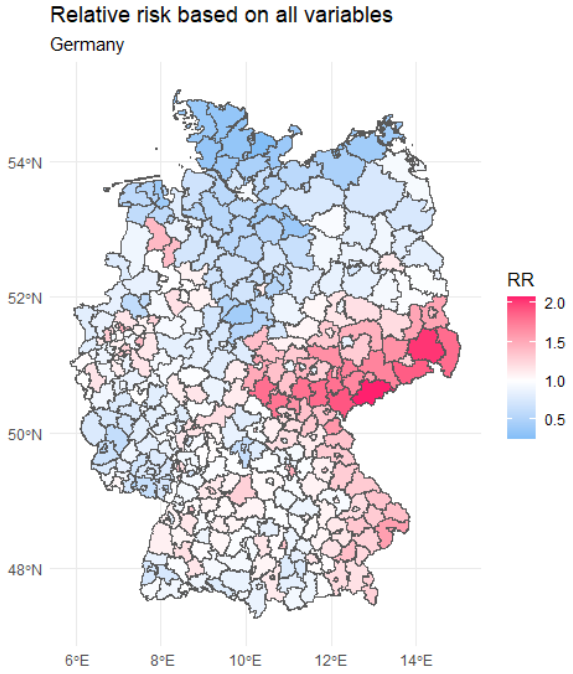

Fig. 6.3: Relative risk of contracting Covid-19 in Germany.

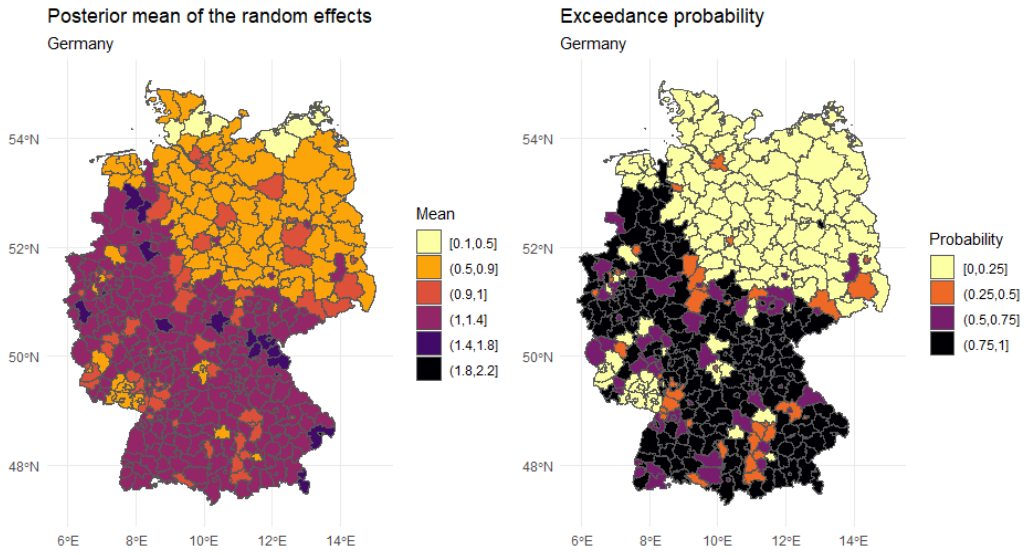

Fig. 6.4: Posterior mean of the municipality-specific relative risks $\zeta = \exp(\xi)$ compared with the whole of Germany (left) and posterior probability $\mathbb{P}(\zeta_i > 1|\boldsymbol{y})$

7 Conclusion

This research aimed to identify factors that have a significant influence on the spread of Covid-19 in Germany and Norway using Bayesian modelling. Based on the calculated models, it can be concluded that the population density, the percentage of the vote for the right-wing party AfD and the logarithmic trade tax are important factors influencing the spread of the viral disease in Germany. For Norway, it was found that the urban density, the number of unemployed immigrants, the total number of immigrants as well as the number of aerodromes in a municipality and the female to male ratio are important factors. The results indicate that in different countries different factors are influencing infection numbers. It has to be said however, that some of the findings are not in line with current research, specifically the influence of the logarithmic trade tax, the number of aerodromes and the proportion of females. blabla why spatio-temporal models were not possible blabla predictive models might make sense for norway bc spatial effect not that strong

133

8 Appendix

8.1 Probability Distributions and the Exponential Family

8.1.1 The Exponential Family

In statistics and probability theory, the *exponential family* is a parametric set of probability distributions of a specific form. The distribution of a random variable $\boldsymbol{y}$ belongs to the exponential family if the discrete or continuous density with respect to a σ-finite measure of $\boldsymbol{y}$ has the form

$$f(\boldsymbol{y}|\boldsymbol{\theta},\lambda) = \exp\left(\frac{\boldsymbol{y}^T\boldsymbol{\theta} - b(\boldsymbol{\theta})}{\lambda} + c(\boldsymbol{y},\lambda)\right), \tag{8.1}$$

with $c(\boldsymbol{y},\lambda) \geq 0$. $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$ is the *natural* or *canonical* parameter of the exponential family, while $\lambda > 0$ is a *dispersion* or *nuisance* parameter (Holland & Leinhardt, 1981). The natural parameter space Θ is the set of all $\boldsymbol{\theta}$ satisfying

$$0 < \int \exp\left(\left(\boldsymbol{y}^T\boldsymbol{\theta} - b(\boldsymbol{\theta})\right)/\lambda + c(\boldsymbol{y},\lambda)\right) d\boldsymbol{y} < \infty \tag{8.2}$$

Moreover, $b(\boldsymbol{\theta})$ is a twice differentiable function and all moments of $\boldsymbol{y}$ exist. Specifically,

$$\mathbb{E}_{\boldsymbol{\theta}}(\boldsymbol{y}) = \mu(\boldsymbol{\theta}) = \frac{\partial b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{8.3}$$

$$\text{Cov}_{\boldsymbol{\theta}}(\boldsymbol{y}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \lambda \frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \tag{8.4}$$

The covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is positive definite in Θ^0, therefore $\mu : \Theta^0 \to M = \mu\left(\Theta^0\right)$ is injective. By substituting the inverse function $\boldsymbol{\theta}(\mu)$ into $\frac{\partial^2 b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$, the variance function

$$v(\mu) = \frac{\partial^2 b(\boldsymbol{\theta}(\mu))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \tag{8.5}$$

is given and the covariance can be written as

$$\text{Cov}_{\boldsymbol{\theta}}(\boldsymbol{y}) = \lambda v(\mu). \tag{8.6}$$

134

Important members of the exponential family are the normal, binomial, Poisson, gamma and multivariate normal distribution (Fahrmeir & Tutz, 2013, p. 433).

8.1.2 The Normal Distribution

The normal distribution is an important type of continuous probability distribution in stochastics. The special significance of the normal distribution is based, among other things, on the central limit theorem, according to which distributions that result from the additive combination of a large number of independent influences are approximately normally distributed under weak conditions.
The density is given by

$$f\left(\boldsymbol{x}|\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\boldsymbol{x}-\mu}{\sigma}\right)^{2}\right). \tag{8.7}$$

The first two moments of the distribution are given by

$$\mathbb{E}\left[X\right]=\mu \tag{8.8}$$
$$\text{Var}\left[X\right]=\sigma^{2}. \tag{8.9}$$

The graph of this density function has a "bell-shaped form" and is symmetrical with μ as the centre of symmetry (Fahrmeir et al., 2016, pp. 83–85).

8.1.3 The Multivariate Normal Distribution

The density of a normally distributed random variable $\boldsymbol{y}=\left(y_{1},...,y_{n}\right)^{T},n<\infty$ with mean vector $\boldsymbol{\mu}$ $(n\times1)$ and SPD covariance matrix $\boldsymbol{\Sigma}$ $(n\times n)$ is

$$\pi\left(\boldsymbol{y}\right)=\left(2\pi\right)^{-n/2}\left|\boldsymbol{\Sigma}\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{\mu}\right)^{T}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{\mu}\right)\right),\ \ \boldsymbol{y}\in\mathbb{R}^{n} \tag{8.10}$$

Here, $\mu_{i}=\mathbb{E}\left[y_{i}\right]$, $\Sigma_{ij}=\text{Cov}\left(y_{i},y_{j}\right)$, $\Sigma_{ii}=\text{Var}\left(y_{i}\right)>0$ and $\text{Corr}\left(y_{i},y_{j}\right)=\Sigma_{ij}/\left(\Sigma_{ii}\Sigma_{jj}\right)^{1/2}$.
This is written as $\boldsymbol{y}\sim\mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$. For $n=1$, $\mu=0$ and $\Sigma_{11}=1$ the standard normal distribution is obtained.
$\boldsymbol{y}$ is now split up into $\boldsymbol{y}=\left(\boldsymbol{y}_{\boldsymbol{A}}^{T},\boldsymbol{y}_{\boldsymbol{B}}^{T}\right)$ and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are divided accordingly:

$$\boldsymbol{\mu}=\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{A}}\\\boldsymbol{\mu}_{\boldsymbol{B}}\end{pmatrix}\qquad\text{and}\qquad\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\Sigma}_{\boldsymbol{AA}}&\boldsymbol{\Sigma}_{\boldsymbol{AB}}\\\boldsymbol{\Sigma}_{\boldsymbol{BA}}&\boldsymbol{\Sigma}_{\boldsymbol{BB}}\end{pmatrix}.$$

Some basic properties of the multivariate normal distribution are the following.

1. $\boldsymbol{y_A} \sim \mathcal{N}(\boldsymbol{\mu_A}, \boldsymbol{\Sigma_{AA}})$.
2. $\boldsymbol{\Sigma_{AB}} = \mathbf{0}$ precisely when $\boldsymbol{y_A}$ and $\boldsymbol{y_B}$ are independent.
3. The conditional distribution $\pi(\boldsymbol{y_A}|\boldsymbol{y_B})$ is $\mathcal{N}\left(\boldsymbol{\mu_{A|B}}, \boldsymbol{\Sigma_{A|B}}\right)$, where

$$\begin{aligned}\boldsymbol{\mu_{A|B}} &= \boldsymbol{\mu_A} + \boldsymbol{\Sigma_{AB}}\boldsymbol{\Sigma_{BB}^{-1}}(\boldsymbol{y_B} - \boldsymbol{\mu_B}) \text{ and} \\ \boldsymbol{\Sigma_{A|B}} &= \boldsymbol{\Sigma_{AA}} - \boldsymbol{\Sigma_{AB}}\boldsymbol{\Sigma_{BB}^{-1}}\boldsymbol{\Sigma_{BA}}.\end{aligned}$$

4. If $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{y'} \sim \mathcal{N}(\boldsymbol{\mu'}, \boldsymbol{\Sigma'})$ are independent,

   then $\boldsymbol{y} + \boldsymbol{y'} \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\mu'}, \boldsymbol{\Sigma} + \boldsymbol{\Sigma'})$ (Rue & Held, 2005, pp. 19–20).

8.1.4 The Poisson Distribution

The Poisson distribution is a discrete probability distribution that can be used to model the number of events that occur independently of each other at a constant mean rate in a fixed time interval or spatial area.
The density is given by

$$f(k) = \mathbb{P}(X = k) = \begin{cases} \frac{\lambda^k}{k!}\exp(-\lambda) & \text{for } x \in \{0, 1, ...\} \\ 0 & \text{else} \end{cases} \tag{8.11}$$

with λ representing the expected value of X.
The first two moments of the distribution are given by

$$\mathbb{E}[X] = \lambda \tag{8.12}$$

$$\text{Var}[X] = \lambda. \tag{8.13}$$

For $\lambda \geq 10$ the distribution becomes approximately symmetrical and can thus be approximated by a normal distribution (Fahrmeir et al., 2016, p. 243).

8.1.5 The Negative Binomial Distribution

The negative binomial distribution is a univariate probability distribution that belongs to the discrete probability distributions. It models the number of trials required

to achieve a given number of successes in a Bernoulli process.
The density is given by

$$f(k, r, p) = \mathbb{P}(X = k) = \binom{k + r - 1}{r - 1} (1 - p)^k p^r, \tag{8.14}$$

with r the number of successes, k the number of failures, and p the probability of success.
The first two moments of the distribution are given by

$$\mathbb{E}[X] = \frac{pr}{1 - p} \tag{8.15}$$

$$\text{Var}[X] = \frac{pr}{(1 - p)^2}. \tag{8.16}$$

For large values of r, the negative binomial distribution can be approximated by a normal distribution (Haldane, 1941).

8.2 Moments

8.2.1 Skewness

Skewness is a statistical indicator that describes the type and strength of the asymmetry of a probability distribution. It shows whether and how strongly the distribution is skewed to the right (right-skewed, left-skewed, negative skewness) or to the left (left-skewed, right-skewed, positive skewness). Any non-symmetrical distribution is called skewed.
The skewness of a random variable X is the third standardized moment, defined as

$$\text{Skew}[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\left(\mathbb{E}\left[(X-\mu)^{3/2}\right]\right)^2} = \frac{\mu_3}{\sigma^3}, \tag{8.17}$$

with μ_3 the third central moment and σ the standard deviation (Doane & Seward, 2011; Wilkins, 1944).

8.2.2 Kurtosis

Kurtosis is a measure of the slope of a probability distribution of a random variable. Distributions with low kurtosis scatter relatively evenly; for distributions with high kurtosis, the scatter results more from extreme but rare events.
The kurtosis of a random variable X is the fourth standardized moment, defined as

$$\text{Kurt}[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\left(\mathbb{E}\left[(X-\mu)^2\right]\right)^2} = \frac{\mu_4}{\sigma^4}, \tag{8.18}$$

with μ_4 the fourth central moment and σ the standard deviation (DeCarlo, 1997; Wilkins, 1944).

8.3 Distribution Fits

8.3.1 Distribution Fits for Germany

8.3.1.1 Fits for the Non-Temporal Models

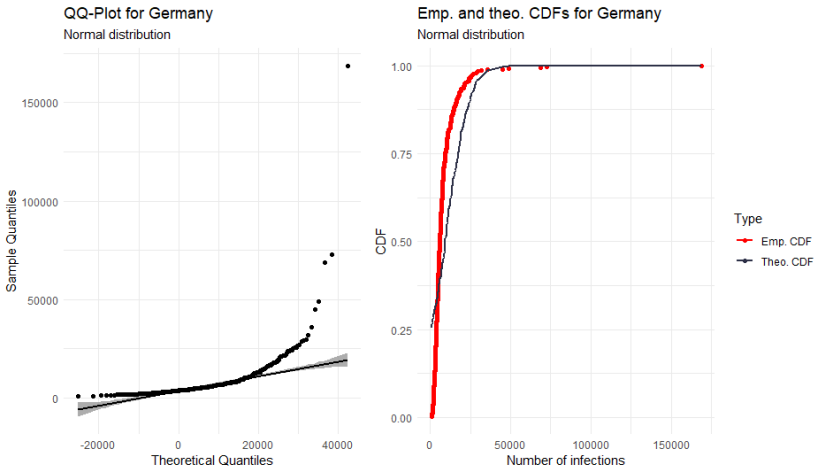

Fig. 8.1: A normal fit to the number of cases in German municipalities

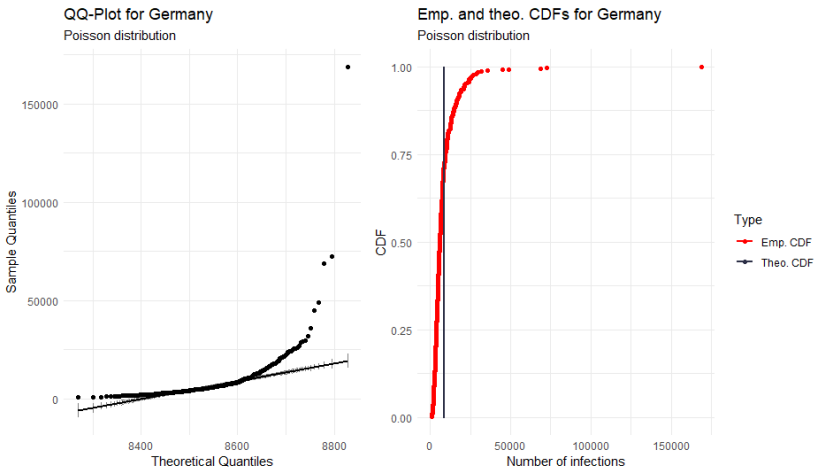

Fig. 8.2: A Poisson fit to the number of cases in German municipalities

8.3.1.2 Fits for the Temporal Models

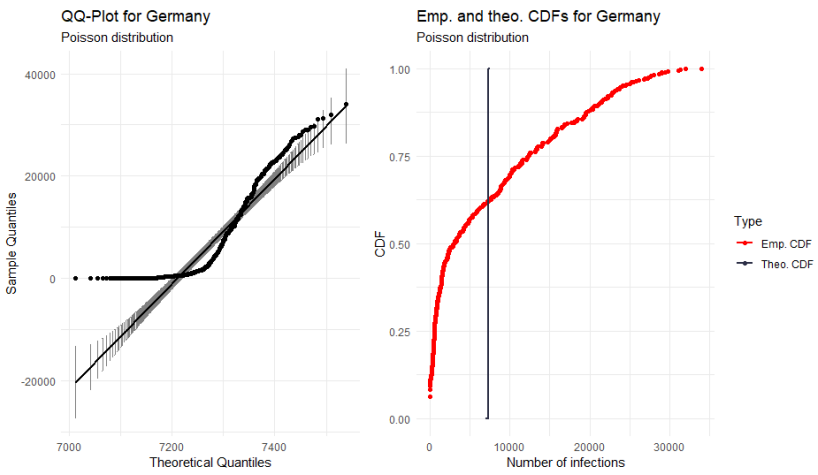

Fig. 8.3: A Poisson fit to the number of cases in German municipalities

8.3.2 Distribution Fits for Norway

8.3.2.1 Fits for the Non-Temporal Models

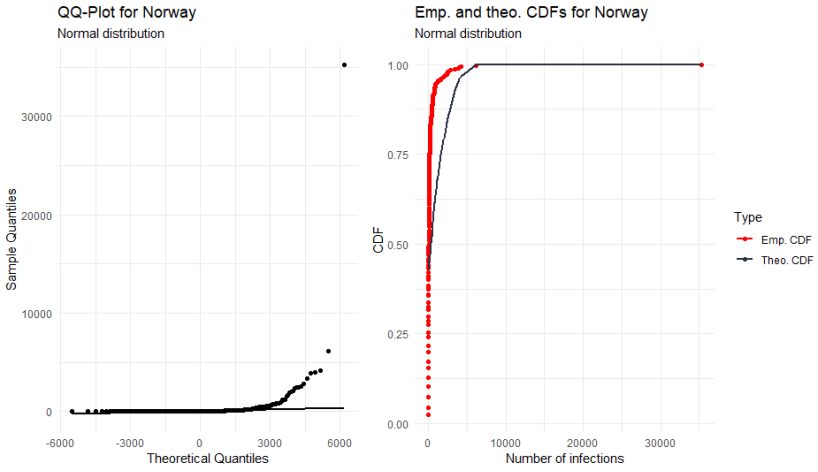

Fig. 8.4: A normal fit to the number of cases in Norwegian municipalities

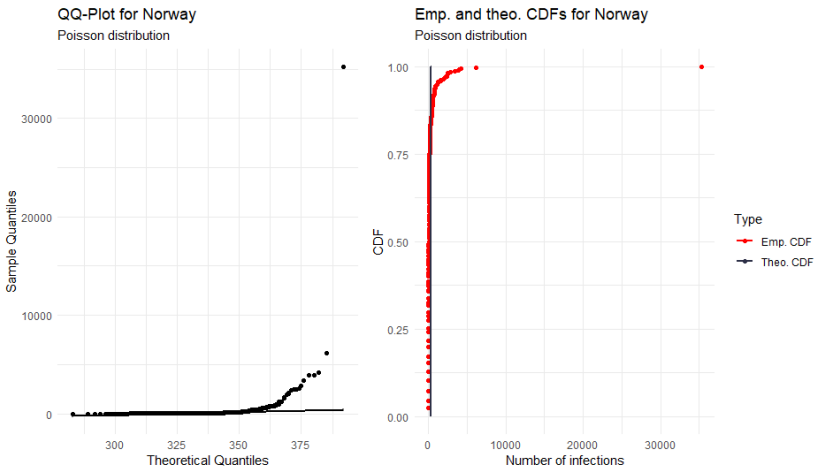

Fig. 8.5: A Poisson fit to the number of cases in Norwegian municipalities

8.3.2.2 Fits for the Temporal Models

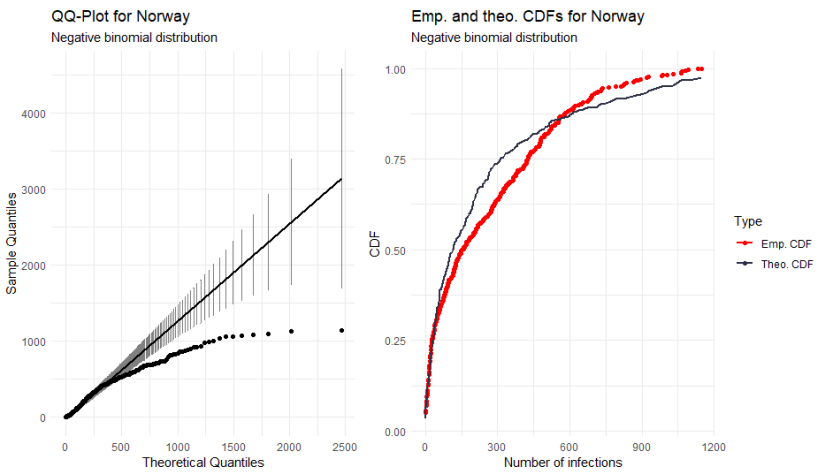

Fig. 8.6: A negative binomial fit to the number of cases in Norwegian municipalities

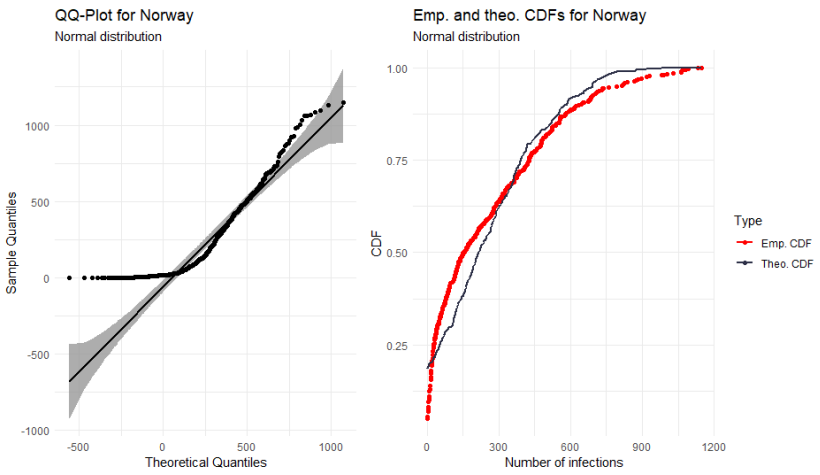

Fig. 8.7: A normal fit to the number of cases in Norwegian municipalities

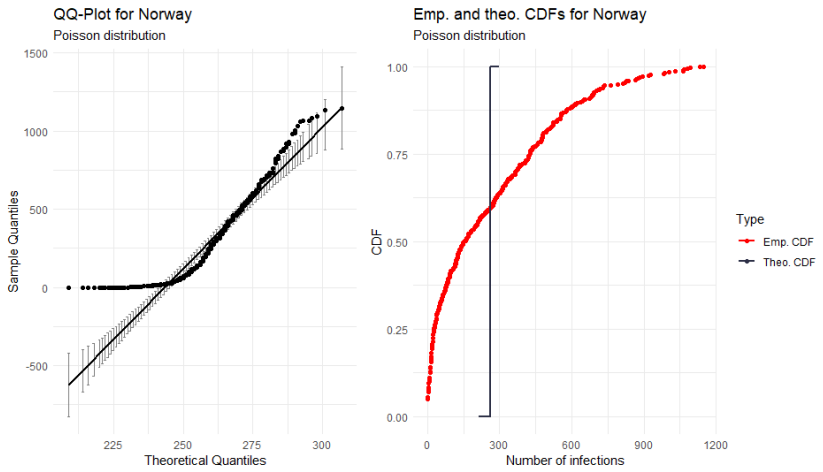

Fig. 8.8: A Poisson fit to the number of cases in Norwegian municipalities

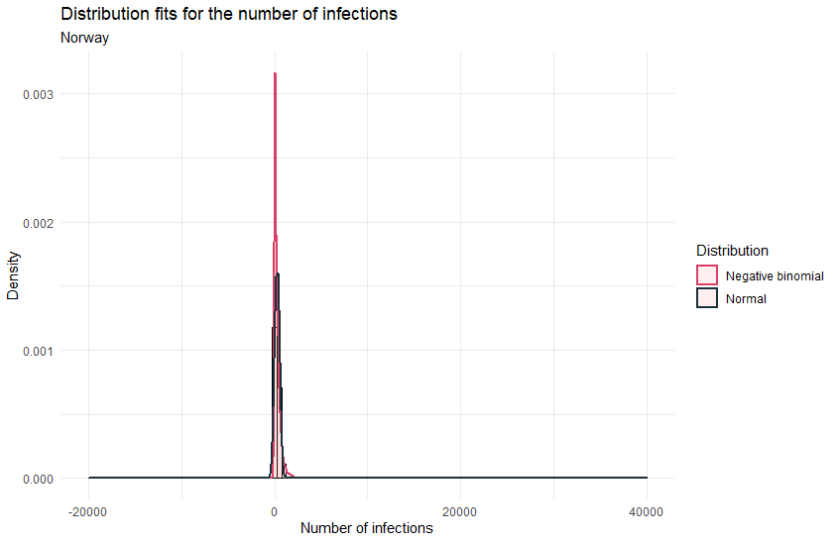

Fig. 8.9: Histogram for the number of cases in Norwegian municipalities with a normal and a negative binomial distribution overlayed.

8.4 Choice of Hyperpriors for Germany

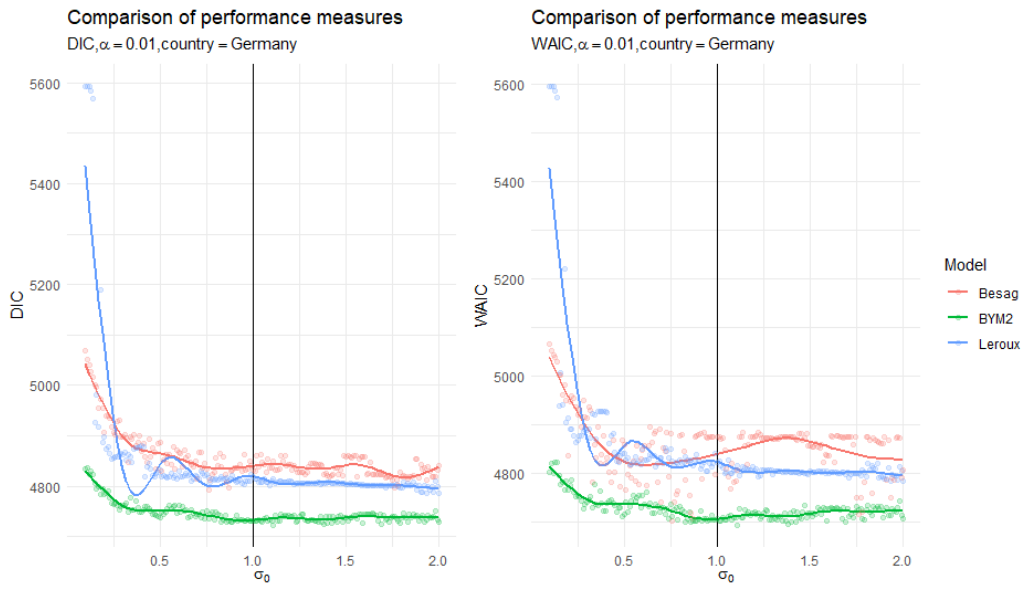

Fig. 8.10: Values of the DIC and the WAIC when changing the value for σ_0. The black line highlights the values for $\sigma_0 = 1$.

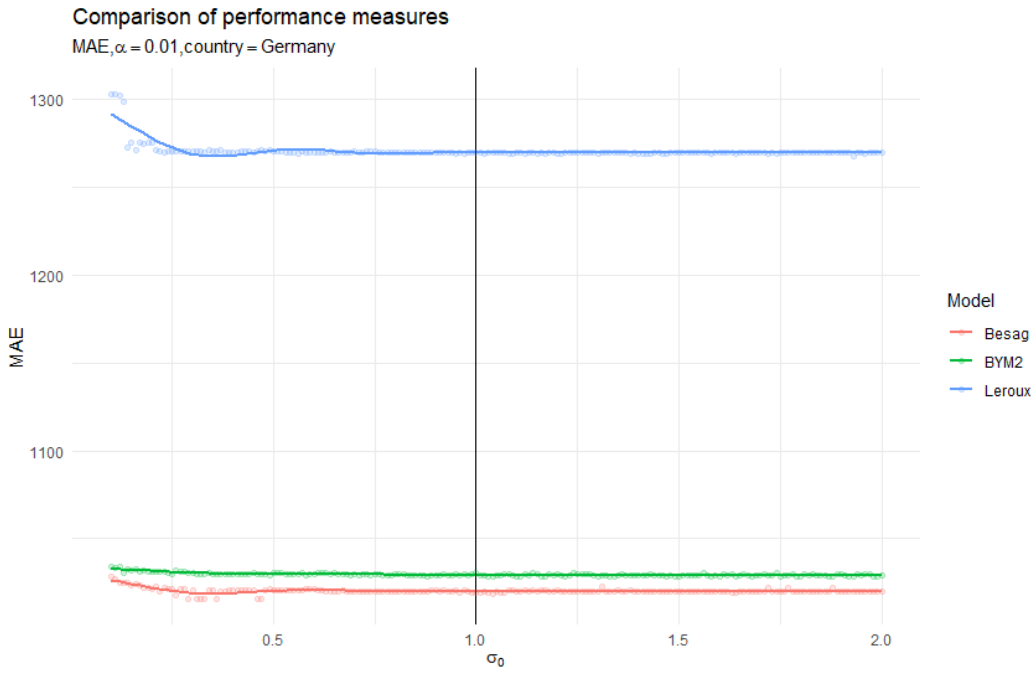

Fig. 8.11: Values of the MAE when changing the value for σ_0. The black line highlights the values for $\sigma_0 = 1$.

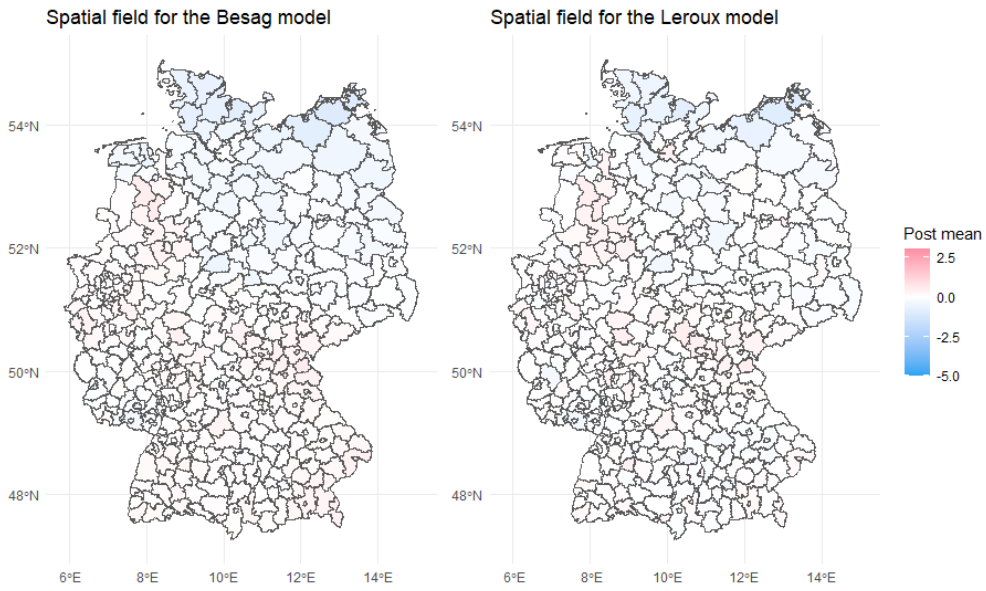

Fig. 8.12: Spatial field for a Besag model and a Leroux model.

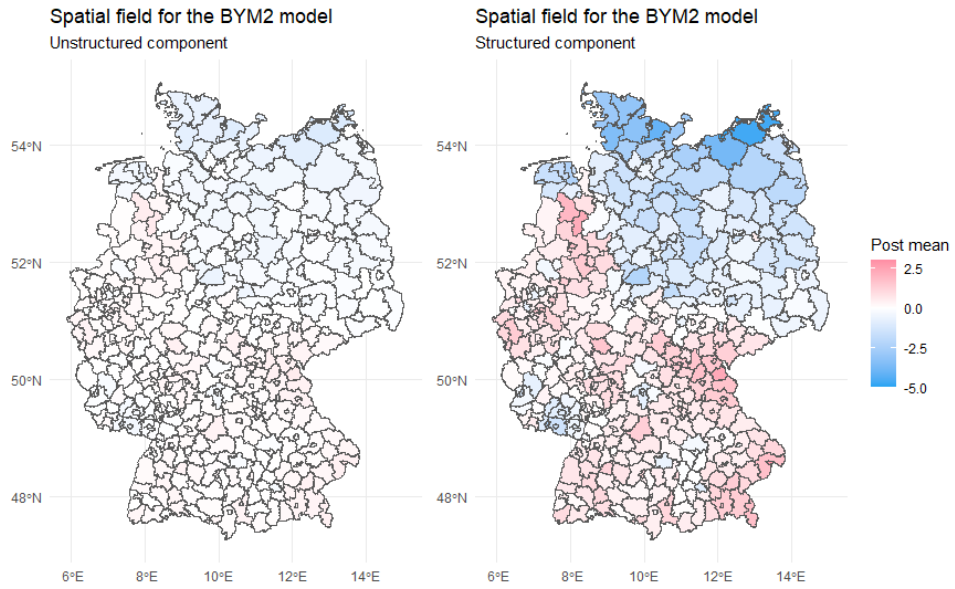

Fig. 8.13: Spatial fields for a BYM2 model.

8.5 Code Examples

8.5.1 Specifying the Different Types of Models

```
# set the seed
set.seed(420)
# draw a sample
test <- sample(
    seq_len(nrow(newest_numbers)),
    size = floor(0.2 * nrow(newest_numbers))
  )
# get the number of infections for the test data
test_value <- newest_numbers$value[test]
# set the number of infections to NA in the train data
newest_numbers$value[test] <- NA
# define the link function
link <- rep(NA, nrow(newest_numbers))
link[which(is.na(newest_numbers$value))] <- 1
# for the temporal model every 5th day is taken
germany <- germany[
  germany$Date %in% seq(
    from = min(germany$Date),
    to = max(germany$Date),
    by = 5
  ),
]
# draw a sample
test_temporal <- sample(
  seq_len(nrow(germany)),
  size = floor(0.2 * nrow(germany))
)
test_value_temporal <- germany$value[test_temporal]
germany$value[test_temporal] <- NA
link_temporal <- rep(NA, nrow(germany))
link_temporal[which(is.na(germany$value))] <- 1
# define the penalised prior
prior_1 <- list(
  prec = list(
    prior = "pc.prec",
    param = c(1, 0.01)
  )
)
# create the neighbordhood matrix
nb <- poly2nb(newest_numbers)
# save the matrix
nb2INLA("maps/map.adj", nb)
g <- inla.read.graph(filename = "maps/map.adj")
```

```
# define the C matrix for the Leroux model
Q <- Diagonal(x = sapply(nb, length))
for (i in 2:nrow(newest_numbers)) {
  Q[i - 1, i] <- -1
  Q[i, i - 1] <- -1
}

C <- Diagonal(x = 1, n = nrow(newest_numbers)) - Q
# formula for the model without the spatial component
formula_1 <- value ~
  pop_dens + urb_dens + sex + trade_tax + SPD + Gruene + FDP +
  die_linke + clinic + place_of_worship + nursing_home +
  aerodrome + platform + office + marketplace + higher_education
# formula for the besag model
formula_2 <- value ~
  pop_dens + urb_dens + sex + trade_tax + SPD + Gruene + FDP +
  die_linke + clinic + place_of_worship + nursing_home +
  aerodrome + platform + office + marketplace + higher_education +
  f(idarea_1, model = "besagproper", graph = g, hyper = prior_1)
# formula for the bym2 model
formula_3 <- value ~
  pop_dens + urb_dens + sex + trade_tax + SPD + Gruene + FDP +
  die_linke + clinic + place_of_worship + nursing_home +
  aerodrome + platform + office + marketplace + higher_education +
  f(
    idarea_1, model = "bym2", graph = g,
    scale.model = TRUE, hyper = prior_1
  )
# formula for the leroux model
formula_4 <- value ~
  pop_dens + urb_dens + sex + trade_tax + SPD + Gruene + FDP +
  die_linke + clinic + place_of_worship + nursing_home +
  aerodrome + platform + office + marketplace + higher_education +
  f(idarea_1, model = "generic1", Cmatrix = C, hyper = prior_1)
# formula for the temporal bym2 model
formula_5 <- value ~
  pop_dens + urb_dens + sex + trade_tax + SPD + Gruene + FDP +
  die_linke + clinic + place_of_worship + nursing_home +
  aerodrome + platform + office + marketplace + higher_education +
  f(
    idarea_1, model = "bym2", graph = g,
    scale.model = TRUE, hyper = prior_1
  ) +
  f(id_date_1, model = "rw2", hyper = prior_1)
# compute the models
# for the non-temporal models the code looks like this
res_1 <- inla(
  formula_1,
```

```
  family = "nbinomial",
  data = newest_numbers,
  E = expected_count,
  control.predictor = list(
    compute = TRUE,
    link = link
  ),
  Ntrials = newest_numbers$population,
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE)
)
# for the temporal models the code looks like this
res_5 <- inla(
  formula_5,
  family = "nbinomial",
  data = germany,
  E = expected_count,
  control.predictor = list(
    compute = TRUE,
    link = link_temporal
  ),
  Ntrials = germany$population,
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE)
)
```

Listing 8.1: Specifying different models in INLA.

8.5.2 Making Predictions for the Test Data

Listing 8.2: The code for making predictions in INLA.

8.5.3 Calculating the Posterior Mean

```
inla.emarginal(
      exp,
      res_1$marginals.fixed$`(Intercept)`
)
```

Listing 8.3: Calculating the posterior mean of a coefficent.

8.5.4 Calculating a Credibility Interval

```
inla.qmarginal(
      c(0.025, 0.975),
      inla.tmarginal(
        exp,
        res_1$marginals.fixed$`(Intercept)`
      )
)
```

Listing 8.4: Extracting the credibility interval for a coefficient

8.5.5 Calculating the Temporal Trend

```
temporal_car <- lapply(
  res_5$marginals.random$id_date_1,
  function(x) {
    marg <- inla.tmarginal(
      function(y) exp(y),
      x
    )
    inla.emarginal(mean, marg)
  }
)
```

Listing 8.5: Calculating the Temporal Trend of a Model

Bibliography

Ahmadi, M., Sharifi, A., Dorosti, S., Ghoushchi, S. J., & Ghanbari, N. (2020). Investigation of effective climatology parameters on covid-19 outbreak in iran. *Science of the Total Environment*, *729*, 138705 (cit. on p. 10).

Aitkin, M. (1991). Posterior bayes factors. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(1), 111–128 (cit. on p. 21).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723 (cit. on p. 50).

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, *191*, 104254 (cit. on p. 11).

Baena-Dıez, J. M., Barroso, M., Cordeiro-Coelho, S. I., Dıaz, J. L., & Grau, M. (2020). Impact of covid-19 outbreak by income: Hitting hardest the most deprived. *Journal of Public Health*, *42*(4), 698–703 (cit. on p. 131).

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press. (Cit. on pp. 34, 67).

Bayes, T. (1763). Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, (53), 370–418 (cit. on p. 18).

Bermudi, P. M. M., Lorenz, C., de Aguiar, B. S., Failla, M. A., Barrozo, L. V., & Chiaravalloti-Neto, F. (2021). Spatiotemporal ecological study of covid-19 mortality in the city of são paulo, brazil: Shifting of the high mortality risk from areas with the best to those with the worst socio-economic conditions. *Travel medicine and infectious disease*, *39*, 101945 (cit. on p. 11).

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, *14*(21-22), 2433–2443 (cit. on p. 66).

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 (cit. on p. 46).

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, *43*(1), 1–20 (cit. on pp. 3, 34, 47).

Bivand, R. S., Gómez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software*, *63*(20), 1–31 (cit. on p. 85).

Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal bayesian models with r-inla*. John Wiley & Sons. (Cit. on p. 45).

Box, G. E., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons. (Cit. on pp. 14, 19).

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., & Di Napoli, R. (2021). Features, evaluation, and treatment of coronavirus (covid-19). *Statpearls [internet]* (cit. on p. 7).

Castro, M. C., Kim, S., Barberia, L., Ribeiro, A. F., Gurzenda, S., Ribeiro, K. B., Abbott, E., Blossom, J., Rache, B., & Singer, B. H. (2021). Spatiotemporal pattern of covid-19 spread in brazil. *Science* (cit. on p. 11).

Chen, Y., Li, Q., Karimian, H., Chen, X., & Li, X. (2021). Spatio-temporal distribution characteristics and influencing factors of covid-19 in china. *Scientific Reports*, *11*(1), 1–12 (cit. on p. 9).

Chen, Z.-L., Zhang, Q., Lu, Y., Guo, Z.-M., Zhang, X., Zhang, W.-J., Guo, C., Liao, C.-H., Li, Q.-L., Han, X.-H., et al. (2020). Distribution of the covid-19 epidemic and correlation with population emigration from wuhan, china. *Chinese medical journal* (cit. on p. 8).

Chu, W., Ghahramani, Z., & Williams, C. K. (2005). Gaussian processes for ordinal regression. *Journal of machine learning research*, *6*(7) (cit. on p. 34).

Cordes, J., & Castro, M. C. (2020). Spatial analysis of covid-19 clusters and contextual factors in new york city. *Spatial and Spatio-temporal Epidemiology*, *34*, 100355 (cit. on p. 131).

Coven, J., & Gupta, A. (2020). Disparities in mobility responses to covid-19. *New York University* (cit. on p. 131).

Cox, D. (1980). Discussion of 'sampling and bayes' inference in scientific modelling and robustness. *JR Stat Soc Ser A*, *143*, 410 (cit. on p. 52).

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), 391–403 (cit. on p. 54).

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons. (Cit. on p. 35).

Daon, Y., Thompson, R. N., & Obolski, U. (2020). Estimating covid-19 outbreak risk through air travel. *Journal of travel medicine*, *27*(5), taaa093 (cit. on p. 128).

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, *41*(1), 1–15 (cit. on p. 19).

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological methods*, *2*(3), 292 (cit. on p. 138).

Dey, D. K., Ghosh, S. K., & Mallick, B. K. (2000). *Generalized linear models: A bayesian perspective*. CRC Press. (Cit. on p. 34).

Diaconis, P., & Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of statistics*, 269–281 (cit. on p. 22).

Diggle, P. J., Ribeiro, P. J., & Christensen, O. F. (2003). An introduction to model-based geostatistics. *Spatial statistics and computational methods* (pp. 43–86). Springer. (Cit. on p. 36).

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: A forgotten statistic? *Journal of statistics education*, *19*(2) (cit. on p. 138).

Eberl, J.-M., Huber, R. A., & Greussing, E. (2020). From populism to the 'plandemic': Why populists believe in covid-19 conspiracies (cit. on p. 131).

Elkeles, T., & Seifert, W. (1996). Immigrants and health: Unemployment and health-risks of labour migrants in the federal republic of germany, 1984–1992. *Social science & medicine*, *43*(7), 1035–1047 (cit. on p. 128).

Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & Tutz, G. (2016). *Statistik: Der weg zur datenanalyse*. Springer-Verlag. (Cit. on pp. 135, 136).

Fahrmeir, L., & Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *50*(2), 201–220 (cit. on p. 34).

Fahrmeir, L., & Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*. Springer Science & Business Media. (Cit. on pp. 34, 135).

Farias, J. E. M., & Pilati, R. (2020). Violating social distancing amid covid-19 pandemic: Psychological factors to improve compliance (cit. on p. 131).

Fink, D. (1997). A compendium of conjugate priors. *See http://www. people. cornell. edu/-pages/df36/CONJINTRnew% 20TEX. pdf*, *46* (cit. on p. 22).

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, *7*(1), 57–68 (cit. on p. 32).

Gaskin, D. J., Zare, H., & Delarmente, B. A. (2021). Geographic disparities in covid-19 infections and deaths: The role of transportation. *Transport policy*, *102*, 35–46 (cit. on p. 128).

Gausman, J., & Langer, A. (2020). Sex and gender disparities in the covid-19 pandemic. *Journal of women's health*, *29*(4), 465–466 (cit. on p. 128).

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press. (Cit. on p. 65).

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, *24*(6), 997–1016 (cit. on p. 53).

Gianquintieri, L., Brovelli, M. A., Pagliosa, A., Dassi, G., Brambilla, P. M., Bonora, R., Sechi, G. M., & Caiani, E. G. (2020). Mapping spatiotemporal diffusion of covid-19 in lombardy (italy) on the base of emergency medical services activities. *ISPRS International Journal of Geo-Information*, *9*(11), 639 (cit. on p. 9).

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632–648 (cit. on p. 67).

Gross, B., Zheng, Z., Liu, S., Chen, X., Sela, A., Li, J., Li, D., & Havlin, S. (2020). Spatio-temporal propagation of covid-19 pandemics. *EPL (Europhysics Letters)*, *131*(5), 58003 (cit. on p. 9).

Guan, W.-j., Ni, Z.-y., Hu, Y., Liang, W.-h., Ou, C.-q., He, J.-x., Liu, L., Shan, H., Lei, C.-l., Hui, D. S., et al. (2020). Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, *382*(18), 1708–1720 (cit. on p. 8).

Haldane, J. B. (1941). The fitting of binomial distributions. *Annals of Eugenics*, *11*(1), 179–181 (cit. on p. 137).

Hansen, T. (2020). Public covid-19 data for norway (covid19data.no). (Cit. on p. 69).

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications (cit. on p. 30).

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the american Statistical association*, *76*(373), 33–50 (cit. on p. 134).

Indseth, T., Grøsland, M., Arnesen, T., Skyrud, K., Kløvstad, H., Lamprini, V., Telle, K., & Kjøllesdal, M. (2020). Covid-19 among immigrants in norway, notified infections, related hospitalizations and associated mortality: A register-based study. *Scandinavian Journal of Public Health*, 1403494820984026 (cit. on p. 128).

Irwin, M. (2005). Prior choice, summarizing the posterior. (Cit. on pp. 22, 23).

Kammann, E., & Wand, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *52*(1), 1–18 (cit. on p. 34).

Kasilingam, D., Sathiya Prabhakaran, S. P., Rajendran, D. K., Rajagopal, V., Santhosh Kumar, T., & Soundararaj, A. (2020). Exploring the growth of covid-19 cases using exponential modelling across 42 countries and predicting signs of early containment using machine learning. *Transboundary and Emerging Diseases* (cit. on p. 11).

Kitagawa, G., & Gersch, W. (1996). *Smoothness priors analysis of time series* (Vol. 116). Springer Science & Business Media. (Cit. on p. 34).

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, *19*(17-18), 2555–2567 (cit. on p. 66).

Kopel, J., Perisetti, A., Roghani, A., Aziz, M., Gajendran, M., & Goyal, H. (2020). Racial and gender-based differences in covid-19. *Frontiers in public health*, *8*, 418 (cit. on p. 128).

Korsvoll, A. I. S. (2020). Nye smittetilfelle i hyllestad- no har sju personar korona. *Firda* (cit. on p. 83).

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86 (cit. on p. 24).

Lang, S., & Brezger, A. (2004). Bayesian p-splines. *Journal of computational and graphical statistics*, *13*(1), 183–212 (cit. on p. 34).

Leroux, B. G., Lei, X., & Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *Statistical models in epidemiology, the environment, and clinical trials* (pp. 179–191). Springer. (Cit. on p. 48).

Lovelace, R., Nowosad, J., & Muenchow, J. (2019). *Geocomputation with r*. CRC Press. (Cit. on pp. 56, 58–61).

Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., & Pilla, F. (2021). Exploring spatiotemporal effects of the driving factors on covid-19 incidences in the contiguous united states. *Sustainable cities and society*, *68*, 102784 (cit. on p. 10).

Markov, A. A. (1906). Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, *15*, 135–156 (cit. on p. 27).

Marroquin, J. L., Velasco, F. A., Rivera, M., & Nakamura, M. (2001). Gauss-markov measure field models for low-level vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(4), 337–348 (cit. on p. 34).

Martínez-Beneito, M. A., López-Quilez, A., & Botella-Rocamora, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in medicine*, *27*(15), 2874–2889 (cit. on p. 65).

Martins, T. G., Simpson, D. P., Riebler, A., Fuglstad, G., Rue, H., & Sørbye, S. (2014). Penalising model component complexity: A principled, practical approach to constructing priors. *arXiv preprint arXiv:1403.4630* (cit. on pp. 23–25, 49).

Mehmood, K., Bao, Y., Abrar, M. M., Petropoulos, G. P., Soban, A., Saud, S., Khan, Z. A., Khan, S. M., Fahad, S., et al. (2021). Spatiotemporal variability of covid-19 pandemic in relation to air pollution, climate and socioeconomic factors in pakistan. *Chemosphere*, *271*, 129584 (cit. on p. 10).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092 (cit. on p. 30).

Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, *44*(247), 335–341 (cit. on p. 27).

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). Gis-based spatial modeling of covid-19 incidence rate in the continental united states. *Science of the total environment*, *728*, 138884 (cit. on p. 10).

Moraga, P. (2019). *Geospatial health data: Modeling and visualization with r-inla and shiny*. CRC Press. (Cit. on pp. 36, 44, 63–65, 67).

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17–23 (cit. on p. 64).

Mukherjee, S., & Pahan, K. (2021). Is covid-19 gender-sensitive? *Journal of Neuroimmune Pharmacology*, 1–10 (cit. on p. 128).

Nandy, A., Tiwari, C., & Kundu, S. (2021). Managing the covid-19 pandemic: Does social infrastructure matter? evidence from india. *Transforming Government: People, Process and Policy* (cit. on p. 11).

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384 (cit. on p. 51).

NTB. (2021). Ordfører: Ulvik-utbruddet spredte seg trolig blant barna. *Haugesunds Avis* (cit. on p. 83).

Openshaw, S. (1984). The modifiable areal unit problem. norwich, uk. (Cit. on p. 67).

Orea, L., & Álvarez, I. C. (2020). How effective has the spanish lockdown been to battle covid-19? a spatial analysis of the coronavirus propagation across provinces. *Documento de trabajo*, *3*, 1–33 (cit. on p. 11).

Padgham, M., Rudis, B., Lovelace, R., & Salmon, M. (2017). Osmdata. *The Journal of Open Source Software*, *2*(14) (cit. on p. 75).

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446 (cit. on p. 76).

Pedrosa, R. H. (2020). The dynamics of covid-19: Weather, demographics and infection timeline. *MedRxiv* (cit. on p. 10).

Petrov, A. N., Welford, M., Golosov, N., DeGroote, J., Degai, T., & Savelyev, A. (2020). Spatiotemporal dynamics of the covid-19 pandemic in the arctic: Early data and emerging trends. *International Journal of Circumpolar Health*, *79*(1), 1835251 (cit. on p. 9).

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(1), 175–184 (cit. on p. 52).

Raiffa, H., & Schlaifer, R. (1961). Applied statistical decision theory, division of research, graduateschool of business administration, harvard university, boston, 1961. *RaiffaApplied Statistical Decision Theory1961* (cit. on p. 22).

Riebler, A., Sørbye, S. H., Simpson, D., & Rue, H. (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research*, *25*(4), 1145–1165 (cit. on pp. 46–49, 53).

Robert, C., & Casella, G. (2013). *Monte carlo statistical methods*. Springer Science & Business Media. (Cit. on pp. 27–29, 31).

Robert, C. P., Marin, J.-M., & Rousseau, J. (2010). Bayesian inference. *arXiv preprint arXiv:1002.2080* (cit. on p. 22).

Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, *38*(2), 337–341 (cit. on p. 67).

Rue, H., & Held, L. (2005). *Gaussian markov random fields: Theory and applications*. CRC press. (Cit. on pp. 15–17, 19, 20, 34, 37, 38, 40–42, 136).

Rue, H., Martino, S., & Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, *71*(2), 319–392 (cit. on pp. 32–34, 42).

Saha, A., Gupta, K., Patil, M., et al. (2020). Monitoring and epidemiological trends of coronavirus disease (covid-19) around the world. *Matrix Science Medica*, *4*(4), 121 (cit. on p. 9).

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and covid-19 fatalities in the european region using spatial regression approach. *Sustainable cities and society*, *62*, 102418 (cit. on p. 11).

Schmid, C., Schiffels, S., Boog, J., Lange, M., & Aschoff, M. (2021). *Covid19germany: Load, visualise and analyse daily updated data on the covid-19 outbreak in germany* [R package version 0.1.1]. (Cit. on p. 69).

Sia, D., Miszkurka, M., Batal, M., Delisle, H., & Zunzunegui, M. V. (2019). Chronic disease and malnutrition biomarkers among unemployed immigrants and canadian born adults. *Archives of Public Health*, *77*(1), 1–10 (cit. on p. 128).

Sigler, T., Mahmuda, S., Kimpton, A., Loginova, J., Wohland-Jakhar, P., Charles-Edwards, E., & Corcoran, J. (2020). The socio-spatial determinants of covid-19 diffusion: The impact of globalisation, settlement characteristics and population (cit. on p. 127).

Snow, J. (1857). Cholera, and the water supply in the south districts of london. *British Medical Journal*, *1*(42), 864 (cit. on p. 2).

Spagnolo, P. A., Manson, J. E., & Joffe, H. (2020). Sex and gender differences in health: What the covid-19 pandemic can teach us. (Cit. on p. 128).

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 485–493 (cit. on p. 51).

Stone, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, 689–705 (cit. on p. 33).

Teutsch, S. M., Churchill, R. E. et al. (2000). *Principles and practice of public health surveillance*. Oxford University Press, USA. (Cit. on p. 55).

Vieten, U. M. (2020). The “new normal” and “pandemic populism”: The covid-19 crisis and anti-hygienic mobilisation of the far-right. *Social Sciences*, *9*(9), 165 (cit. on p. 130).

Wang, Y., Liu, Y., Struthers, J., & Lian, M. (2021). Spatiotemporal characteristics of the covid-19 epidemic in the united states. *Clinical infectious diseases*, *72*(4), 643–651 (cit. on p. 10).

Watanabe, S., & Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, *11*(12) (cit. on p. 51).

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20 (cit. on p. 76).

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations* [R package version 1.4.0]. (Cit. on p. 76).

Wilkins, J. E. (1944). A note on skewness and kurtosis. *The Annals of Mathematical Statistics*, *15*(3), 333–335 (cit. on p. 138).

Wondreys, J., & Mudde, C. (2020). Victims of the pandemic? european far-right parties and covid-19. *Nationalities Papers*, 1–34 (cit. on p. 130).

Xiong, Y., Wang, Y., Chen, F., & Zhu, M. (2020). Spatial statistics and influencing factors of the novel coronavirus pneumonia 2019 epidemic in hubei province, china (cit. on p. 10).

Yong, L. (2018). Loo and waic as model selection methods for polytomous items. *arXiv preprint arXiv:1806.09996* (cit. on p. 51).

Webpages

des Bundes und der Länder, S. Ä. (2020). *Regionaldatenbank deutschland*. Retrieved February 28, 2021, from https://www.regionalstatistik.de/genesis/online. (Cit. on p. 72)

Deutschland, E. (2020). *Kreisgrenzen 2017*. Retrieved February 28, 2021, from https://opendata-esri-de.opendata.arcgis.com/datasets/affd8ace4c204981b5d32070f9547eb9_0. (Cit. on p. 74)

Folkehelseinstituttet. (2020). *Statistikkbankene i fhi*. Retrieved April 5, 2021, from https://statistikk.fhi.no/. (Cit. on p. 71)

Geonorge. (2021). *Kartkatalogen*. Retrieved February 28, 2021, from https://www.geonorge.no/. (Cit. on p. 74)

OpenStreetMap contributors. (2017). *Planet dump retrieved from https://planet.osm.org*. (Cit. on p. 75).

Sentralbyrå, S. (2016). *Statbank norway*. Retrieved February 28, 2021, from https://www.ssb.no/en/statbank/. (Cit. on p. 72)

Smistad, E. (2020). *Konverter-norgeskart-projeksjon*. (Cit. on p. 74).

List of Figures

158

List of Tables

List of Listings

163