Approximate Bayesian Computation and compartmental metapopulation model to describe COVID-19 in Finland

Alejandro Ponce de León Chávez

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Otaniemi, Espoo 22-11-2021

Supervisor

Prof. Mikko Kivelä

Advisors

Prof. Lasse Leskelä

Prof. Tapio Ala-Nissilä

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the master's thesis

Author Alejandro Ponce de León Chávez

Title Approximate Bayesian Computation and compartmental metapopulation model to describe COVID-19 in Finland

Degree programme Computer, Communication and Information Sciences

Major Machine Learning, Data Science and Artificial Intelligence **Code of major** SCI3044

Supervisor Prof. Mikko Kivelä

Advisors Prof. Lasse Leskelä, Prof. Tapio Ala-Nissilä

Date 22-11-2021 **Number of pages** 72 **Language** English

Abstract

At the beginning of the Coronavirus Disease 2019 (COVID-19) pandemic, when there were no vaccines available, a considerable number of governments around the world imposed mobility restrictions to control the spread of the disease. Mobility restrictions are part of the so-called Non-Pharmaceutical Interventions (NPIs). NPIs are measures taken by individuals or mandated by governments to prevent the spread of the disease. Some of the most notable NPIs include social distancing, use of face masks, remote working, closure of non-essential businesses, and travel restrictions. In Finland, these and other NPIs were essential to control the disease during the first wave of the epidemic, from March 2020 to May 2020.

The aim of this thesis is to use a compartmental metapopulation model to study the COVID-19 epidemic and the role of NPIs in Finland. The model is informed with mobility data from a telecommunications operator and calibrated to describe the epidemic during the first wave. Furthermore, the model is calibrated using Approximate Bayesian Computing (ABC), a set of algorithms that make it possible to infer model parameters without an explicit function for the likelihood of data being produced by the model. The purpose of the model is to analyze the effect of NPIs in Finland through counterfactual scenarios. Counterfactual scenarios affect aspects of the calibrated model to present alternate scenarios of the original situation. This thesis uses two classes of counterfactual scenarios based on mobility data. The first class reduces or increases the number of contacts made inside the region. Similarly, the second class increases or reduces the cross-border traffic between the regions in Finland.

The analysis performed in this thesis confirms that NPIs applied in Finland were crucial to reducing the incidence in Finland during the first wave. Moreover, NPIs which prevent contacts inside regions, such as social distancing, have a more significant impact on reducing the incidence than those that limit traffic between the regions. Finally, the results of this analysis, i.e., the observed effect of NPIs and the software package developed for calibrating the model through ABC, could further analyze the situation of COVID-19 or future epidemics.

Keywords Approximate Bayesian Computation, Epidemic modelling, Coronavirus, Mobility , NPIs

Acknowledgements

First, I want to thank professors Mikko Kivelä, Lasse Leskelä, and Tapio Ala-Nissilä for their confidence in me and for inviting me to be part of the NordicMathCovid team. It has been a year in which I have learned a lot. This thesis is only a tiny reflection of their guidance and support.

I would also like to thank my family. Their love and support have and will continue to guide me through this life. You are always with me no matter where I go. You made me the person that I am today. *Muchas gracias siempre, orejudos, los quiero mucho.*

This would not be my thesis without also mentioning my friends. *A toda la banda.* Life and luck have allowed me to meet amazing people along the way. Although they may not know it, their friendship was also involved in the process of finishing this thesis.

The final words go to my girlfriend, *a mi maranita.* Thanks for always being there for me despite the distance and time. You are the only person that can put up with me and still smile.

Otaniemi, Espoo, 22-11-2021

Alejandro Ponce de León Chávez

Contents

Symbols and acronyms

Symbols

S	Susceptible
E	Exposed
P	Presymptomatic
A	Asymptomatic
I	Infected
R	Recovered or Removed
D	Deceased
V	Vaccinated
N	Total population size
β	Infectivity or transmission rate
γ	Recovery period or rate
λ	Latency period or rate
R_0	Basic reproduction number
R_{eff}	Effective reproduction number

Acronyms

ABC Approximate Bayesian Computation. 10, 17–21, 38–42, 45, 59, 60

CI Confidence Interval. 42–45

COVID-19 Coronavirus Disease 2019. 8–10, 37, 63

ERVA Hospital Catchment Area (in Finnish *erityisvastuualue*). 32, 33, 37, 44

HCD Hospital Care District. 32–45, 47–56, 58

MCMC Markov Chain Monte Carlo. 17, 21

NPI Non-Pharmaceutical Intervention. 8–10, 16, 18–21, 35, 43, 44, 46, 47, 50, 56–60, 63

O-D Origin-Destination. 25–27, 29, 48

THL Finnish Institute for Health and Welfare (in Finnish *Terveyden ja hyvinvoinnin laitos*). 10, 32, 36–46, 54, 55, 59, 60

WHO World Health Organization. 8

1 Introduction

The Coronavirus Disease 2019 (COVID-19) has had a tremendous impact on the life of everyone for the last two years. On 31 December 2019, China informed about cases of pneumonia of unknown cause (later confirmed as COVID-19) and, as of 3 January 2020, 44 patients were admitted into the Wuhan Hospital, 11 severely ill [66]. Two weeks later, on 13 January, the first detected coronavirus cases outside China were confirmed in Thailand [68]. The first reports warning about the severity of the disease, its symptoms, and the fact that it was human-to-human transmissible arrived soon at the end of January [56]. Despite the efforts, given the harsh characteristics of COVID-19 by 11 March 2020, the World Health Organization (WHO) cataloged coronavirus as a pandemic. At the time, there were 118,000 cases in 14 countries, and 4,291 people had died [67]. Shortly after that, Dr. Tedros Adhanom, the Director-General of the WHO, declared that Europe was the epicenter of the epidemic [41]. As a result, European countries declared a state of emergency to control the rising incidence and severity of the coronavirus. The primary measure adopted by European countries was to put the population into a lockdown that only permitted essential trips, such as to grocery shops or pharmacies. On 18 March 2020, there were 250 million people in lockdown because of Coronavirus Disease 2019 (COVID-19) only in Europe [63].

The lockdown was a measure implemented in several countries because of different reasons: (1) COVID-19 is human to human transmissible, (2) the transmission of the disease is through exposure to respiratory fluids carrying infectious viruses coming from droplets and aerosol particles [73], (3) there was no antiviral treatment for any coronavirus that had proven to be effective [56], (4) due to its novelty there was no available vaccine. These aspects suggested that the best measures to prevent new infections were physical distancing, community use of face masks, and avoidance of crowded indoor spaces [73]. The measures, combined with the increasing infections and the fatality of the disease among risk groups [42], forced the countries to implement a generalized lockdown, among other actions. These measures that do not rely on taking medication or getting vaccinated are denominated Non-Pharmaceutical Intervention (NPI) [43].

NPIs were the primary containment measure taken by most countries, and they proved to be helpful. Some of the NPIs commonly taken to avert the transmission of COVID-19 are social distancing, school closures, travel restrictions both external and internal in a country, and the use of face masks in public spaces. In summary, most measures attempted to decrease human mobility. Countries closed their borders to prevent importing new cases, and internally they closed public spaces to prevent contact. The strategy confirmed its effectiveness since the beginning of the epidemic in Wuhan. The Chinese government prohibited traveling to and from Wuhan on 23 January 2020. The authors in Ref. [64] estimate that the Wuhan travel restrictions delayed the arrival time of COVID-19 to some cities by 2.91 days which in turn gave those cities more preparation time, potentially saving lives. The same study also estimates that, without the travel ban and the intervention of the government, the number of confirmed cases by 19 February would have been around 744,000. With

NPIs, the confirmed number of cases was 29,839.

The situation in Europe was not different. The first country to implement major NPIs was Italy, and soon other countries followed. Italy ordered school closures on 5 March 2020, and lockdown followed soon on 11 March. The majority of European countries applied the same measures with some delay. Most ordered school closures on 13 March and a national lockdown from 16 March onward. NPIs saved the lives of about 3,100,000 people in eleven European countries, from the beginning of the epidemic until 4 May 2020, according to some estimates [52].

Finland was not exempted from the coronavirus epidemic nor from NPIs applied by the government. As of 17 September 2021, Finland has had 135,815 confirmed cases and 1,052 deaths attributed to COVID-19 [79]. A way to compare the number of confirmed cases and deaths to other countries is dividing by million inhabitants. Finland has had 24,000 confirmed cases per million inhabitants, while Spain has 105,000, and Germany 49,000. Finland is also the Nordic country with the least confirmed cases per million; Sweden has had 112,000, Norway 32,000, and Denmark 61,000. Revising the number of deaths per million inhabitants, Finland has had 180, Spain 1,800 and Germany 1,100. Among the Nordics, Finland is the second with the lowest number of deaths per million, the country with the lowest number of deaths is Norway with 150, Denmark with 450.52, and Sweden with 1450.77 [46]. From these numbers, the spread of the disease in Finland has been more controlled than in other European countries. Hence the interest to analyze the NPIs in Finland. The objective of NPIs taken by the Finnish government to control the epidemic was to reduce mobility, especially at the beginning of the epidemic in the spring of 2020.

Finland declared a state of emergency on 16 March 2020. At that moment, there were 272 confirmed cases of COVID-19. The state of emergency entailed immediate school closure, the suspension of passenger traffic to Finland except for returning Finnish citizens, the ban of gatherings of more than ten people, and the closure of museums, libraries, and other public spaces [69]. These imposed measures were the first and most significant mobility restrictions. The second most significant mobility restriction was the border closure of Uusimaa region on 27 March; Uusimaa is the capital region where Helsinki is located. It has 1.7 million inhabitants, a little less than a third of the population of Finland [72]. Uusimaa had reported 865 of the 1396 cases nationally at the time of the border closure. The government decided to close the borders of Uusimaa to keep the majority of the infections in the capital area and prevent the arrival of the coronavirus to other regions. Uusimaa borders remained closed until 15 April 2020 [71]. The event marked the gradual reopening of society and the lift of mobility restrictions. By 1 July 2020, the government partially lifted most of the restrictions; even public outdoor events of more than 500 people were allowed [70].

Coronavirus studies and data informed the decisions of the governments and helped them choose the NPIs that were applied. As mentioned earlier, some works studied the impact of NPIs [52], [59], [64]. The common conclusion was that social distancing measures were crucial for controlling the spread of the epidemic. Another aspect studied was the way coronavirus extended across countries using airport travel information [45]. This work also studied the importance of international travel bans

and when it is adequate to make such bans. A similar study that studied the timing of NPIs and their implications is the one presented in Ref. [44]. In this study, the authors used a model informed by mobile phone data to discuss the impact of NPIs in different neighborhoods in a city, taking into account the level of income of such neighborhoods. All of these studies greatly depend on the available data. Fortunately, precisely because of this very reason, there were several efforts to make different kinds of data available. Many governments and health institutes made public the daily incidence, hospital occupation, and the number of deaths [46], [50], [53], [79]. Also, companies such as Google [55] and Facebook [48] released information that described how mobility was changing in each country and region.

The main objective of this Thesis is to take advantage of the available data sources to model the spread of the COVID-19 epidemic in Finland during the spring of 2020 and then evaluate the role of NPIs taken by the government through counterfactual scenarios. A counterfactual scenario presents an alternate version of what happened. For example, it can present the scenario when the government did not take any NPIs or the opposite case, in which the restrictions taken by the government were stricter. By comparing the results of these counterfactual scenarios, it is possible to determine the factors that most influenced the progression of COVID-2019.

The counterfactual scenarios posed in this Thesis will analyze different mobility patterns during spring 2020, the period of the first coronavirus outbreak in Finland. The purpose is to establish which were the mobility patterns that affected the most progression of coronavirus. The mobility patterns are analyzed through a compartmental metapopulation model that captures mobility between twenty-one regions in Finland. This model investigates the interactions and movements of people between and within the regions. In this case, the calibration of the model is done using Approximate Bayesian Computation (ABC), mobile phone data provided by a major telecommunications operator, and the reported incidence by the Finnish Institute for Health and Welfare (in Finnish *Terveyden ja hyvinvoinnin laitos*) (THL).

The knowledge acquired during this epidemic can help control the next one by having models in place, providing adequate data for such models, and knowing which NPIs are more effective in controlling the disease. The present Thesis is a contribution to this knowledge.

1.1 Outline of the Thesis

The next Chapter 2 presents the theoretical aspects of epidemic modeling. It also discusses the theory behind inference with ABC. At the end of the Chapter, there are examples of studies similar to this Thesis analyzing coronavirus and NPIs. Chapter 3 presents the metapopulation model and its calibration to the coronavirus outbreak in Finland during spring 2020. Chapter 4 presents the description of the counterfactual scenarios and the comparison of such scenarios to the calibrated model. Finally, Chapter 5 discusses the effectiveness of NPIs according to the counterfactual scenarios, and Chapter 6 summarizes and concludes the Thesis.

2 Background

2.1 Epidemic modeling

Humankind, through its existence, has suffered any number of epidemics. Some have taken the life of a significant part of the world population; a classic example is a plague, also referred to as *Black Death*. The plague, during the fourteenth century, caused an estimated 50 million deaths. Half of those deaths were in Africa and Asia, and the other half in Europe. Those 25 million deaths in Europe represented a quarter of the European population at the time [20]. One more recent and typical example of an epidemic is influenza. Influenza is a complex and changing virus; it has developed new strains that appear periodically and at irregular intervals causing pandemics in different periods and countries. There have been 31 documented influenza pandemics; the so-called *Spanish flu* was one of the most virulent [20].

The purpose of modeling infectious diseases is to understand how the pathogens that cause epidemics spread, predict the outcome of an outbreak, and evaluate strategies to control the damage. Daniel Bernoulli conducted the first mathematical model of the spread of a disease in 1766 [21]. In this work, Bernoulli presented the advantages of a global inoculation against smallpox to influence the general public to get inoculated. However, it was not until the twentieth century that the first compartmental models were proposed.

2.1.1 Compartmental epidemic models

Compartmental models try to mimic the behavior of an epidemic by moving individuals from one compartment to another. Compartments are states that any given individual can experience during an epidemic. The most simple example is a model with two compartments: Susceptible to be infected (S) and Infected (I). All individuals start from the S compartment, and after some given time, if they contact an individual from the I compartment, they transition to the I compartment. The use of compartmental models in epidemics is more than a century old. One of the first works in the matter was published in 1916 by Ross [1]. In this work, the population was divided only into two compartments, similarly as in the example given before, affected and not affected. Interestingly, although there are only two compartments, this seminal work also considers other factors in the model such as birth rate, death rate, and immigration and emigration rates. Ross and Hudson expanded this work in Ref. [2] and Ref. [3]. Typically, the model does not consider the birth and death rates since the duration of the epidemic is assumed to be sufficiently short in time for these factors to be negligible [4], [29].

The first compartmental Susceptible, Infected, and Removed (SIR) model was proposed in 1927 by Kermack and McKendrick [4]. This work took the previous one by Ross and Hudson [1]–[3], but it took the death and birth rates as constants and introduced the third compartment. The Removed (R) compartment was for those who either recovered from the disease and developed immunity or those who died after contracting the disease. Figure 1 shows a graphic representation of the SIR model.

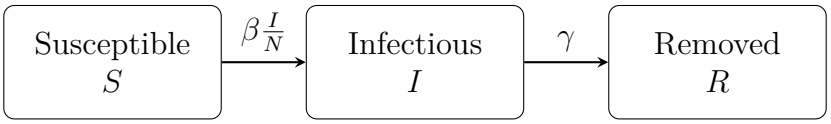

Figure 1: Graphical representation of the SIR compartmental model. The Susceptible S individuals transition to the Infected compartment I at a rate $\beta\frac{I}{N}$. In turn, the Infected I transition to the Removed R compartment at a rate γ.

Kermack and McKendrick proposed the following differential equations to describe the change over time of the compartments S, I, and R under constant infectivity rate β and recovery rate γ:

$$\begin{aligned} \frac{dS}{dt} &= -\beta\frac{SI}{N}; \\ \frac{dI}{dt} &= \beta\frac{SI}{N} - \gamma I; \\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{1}$$

where t represents time, $S+I+R = N$ and N is the total size of the population. Note that there is no entry into the S compartment and no exit from the R compartment, i.e., the population size is kept constant through the epidemic and is only moving within the compartments.

Other works built more elaborated models to account for different states by taking the SIR compartmental model as a base. For example, the $SIRD$ model adds the Deceased D compartment. Essentially, it differentiates when a person has recuperated from the disease and transitions into a state in which it acquires immunity (R) and when it does not recover from the disease and dies (D). Similarly, the $SIRV$ accounts for people being Vaccinated V and acquiring immunity. One of the most common variations of the SIR model is the $SEIR$ model. The $SEIR$ model adds a compartment E to account for the exposed but not yet infectious individuals [35] (Figure 2).

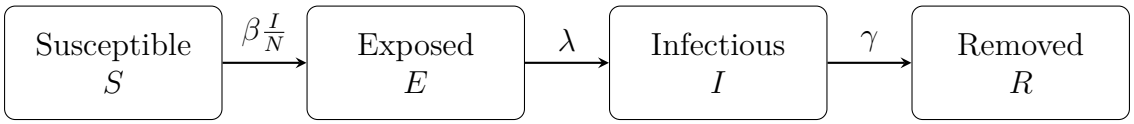

Figure 2: Graphical representation of the $SEIR$ compartmental model. The main difference from this model to the SIR model in Figure 1 is the introduction of a third compartment E for individuals who are exposed but not yet infectious.

The system of differential equations for the $SEIR$ model is the following,

$$\frac{dS}{dt} = -\beta\frac{SI}{N};$$
$$\frac{dE}{dt} = \beta\frac{SI}{N} - \lambda E;$$
$$\frac{dI}{dt} = \lambda E - \gamma I;$$
$$\frac{dR}{dt} = \gamma I.$$

We note that there is an extra parameter λ that represents the latent period. The latent period is the average time an individual transitions from being exposed but not infectious (compartment E) to infectious (compartment I).

Adding more compartments to the model is not the only change that can be applied. The model can also be stochastic and have some inherent randomness. Such stochastic compartmental models are reviewed next in Section 2.1.2.

2.1.2 Stochastic compartmental models

So far, the discussed models have been deterministic. In the deterministic setting is assumed that, over a specific period, the incidence and deaths will always be the same. This assumption applies when the infectious individuals form a small fraction of a large population. Thus, there will be slight fluctuations considering the individuals as discrete units, with a probability to be infected or not, or fractionally [29]. However, this is not the case in real life; there is always a probability of dying or getting infected that needs to be considered [8]. The first probability approaches were Greenwood's in 1931 [5] and Reed-Frost in 1928 but were not published until 1952 [7]. These works study the spread of disease in households using a binomial distribution. However, using a binomial distribution is not the only way to approach stochastic epidemic models; other approaches can be considered [25].

Going back to the SIR model but now presenting the model with one of the possible stochastic approaches, it is possible to randomly pick, at each time step t, the number of infected individuals $X^{(1)}(t)$ and the number of recovered individuals $X^{(2)}(t)$. Then, the SIR deterministic model in Section 2.1.1 can be expressed in a similar but stochastic fashion,

$$S(t+\Delta t) = S(t) - X^{(1)}(t);$$
$$I(t+\Delta t) = I(t) + X^{(1)}(t) - X^{(2)}(t);$$
$$R(t+\Delta t) = R(t) + X^{(2)}(t).$$

Here the total population also remains constant, $N = S(t) + I(t) + R(t)$. The probability of getting infected is given by,

$$\beta\Delta t\frac{I(t)}{N},$$

where the factor $\frac{I(t)}{N}$ represents the probability of a susceptible individual meeting an infected one in a homogeneous mixing population, β is the transmission rate, and Δt is a small time step discretization. The recovery probability is given by,

$$\gamma \Delta t,$$

where γ^{-1} is the average time an individual takes to recover or die from the disease. Using these two probabilities, we can pick the number of infected and removed individuals at time t using a binomial distribution $\text{Binom}(n, p)$ with n trials and timty p,

$$X_1(t) \sim \text{Binom}\Big(S(t), \beta \Delta t \frac{I(t)}{N}\Big);$$
$$X_2(t) \sim \text{Binom}(I(t), \gamma \Delta t).$$

Another common addition to the compartmental models, both in their deterministic and stochastic settings, is studying how different populations interact. For example, the so-called metapopulation models study the spatio-temporal interaction of populations separated geographically. The details of the metapopulation models are presented Section 2.1.3.

2.1.3 Compartmental metapopulation models

Compartmental epidemic models assume all populations have the same epidemic dynamics regardless of gender, age, or geographical location. To introduce heterogeneity, one common way to break this assumption is to divide the individuals into smaller groups depending on their age, gender, or spatial location. Usually, the groups get assigned different transmission rates β for contacts between groups and within the same group. The difference in transmission rate, or any other epidemic parameter, allows modeling the epidemic independently for each group and how the interaction among these groups will affect the outcome of the epidemic [11].

Compartmental metapopulation models break the homogeneous mixing assumption implicit into the traditional compartmental epidemic models. They do it by separating the population into smaller groups based on geographical regions. In the twentieth century, the advancements in transportation, such as the train, car, and airplane, shortened the travel times. Trips that used to take days or weeks were possible in hours. The shorter travel times provoked an increase in human mobility. The increase in human mobility facilitated the spread of diseases that were transmitted person-to-person. Before the twentieth century, diseases that geography limited to a small region resulted in national or worldwide infections [19]. The spread of diseases among geographical regions motivates metapopulation models that can capture the evolution of epidemics.

One of the first metapopulation studies that included data and compared their results to a disease outbreak was by Rvachev and Longini in 1985 [14]. The authors modeled the geographical spread of influenza. Since then, several studies have followed the same path and tried to model worldwide outbreaks of influenza using

available information on air travel [22], [24], [27]. These studies build a network using the database of the International Air Transport Association (IATA) [76]. The database contains the worldwide list of airport pairs connected by direct flights and the number of available seats on any given connection. With the information, the authors build a network between the commercial airports. Each airport is a node in the network. The network connects two airports with a weighted edge, in which the weight represents the passenger flow between the airports. Once the authors have built the network, they simulate the spread of influenza using a compartmental metapopulation model.

The mobility within a country has also been studied. For example, Ref. [23] studied the spread of influenza in the United States, Ref. [33] the case in France, and Ref. [26] analyzed the spread of measles in coastal cities of the United Kingdom. More recently, the authors in Ref. [47] studied the spread of influenza in Bangladesh. Here, the authors divide Bangladesh into 544 sub-regions. In this case, the mobility between the 544 sub-regions is taken from mobile phone data. Every sub-region i has a local stochastic compartmental model; the model has five compartments. To make an analogy of the model in Ref. [47] to the one presented previously in Section 2.1.2, which was stochastic but not compartmental, the five compartments of the model in Ref. [47] are simplified to three,

$$\begin{aligned} S_i(t+\Delta t) &= S_i(t) - X_i^{(1)}(t); \\ I_i(t+\Delta t) &= I_i(t) + X_i^{(1)}(t) - X_i^{(2)}(t); \\ R_i(t+\Delta t) &= R_i(t) + X_i^{(2)}(t), \end{aligned}$$

and

$$\begin{aligned} X_i^{(1)}(t) &\sim \text{Binom}\left(S_i(t), \beta_i \Delta t \frac{I_i(t)}{N_i}\right); \\ X_i^{(2)}(t) &\sim \text{Binom}(I_i(t), \gamma \Delta t). \end{aligned}$$

As previously mentioned, $\text{Binom}(n, p)$ is the binomial distribution with n trials and probability p and γ^{-1} the average time to recovery. Note that each sub-region i has in independent transmission rate β_i. The independent transmission rate β_i allows modeling the case where the infection is more severe in some region i compared to another region j. Also, it is worth mentioning that the population in region i does not remain constant, it varies in time according to the number of ingoing and outgoing people,

$$N_i(t) = N_i + \sum_{j=0}^{K} w_{j,i}(t) - \sum_{j=0}^{K} w_{i,j}(t).$$

The variable N_i is the initial population in region i, $w_{j,i}(t)$ is the number of people traveling from region j to region i at time step t and K the total number of regions of the model. The population in region i is being affected at every time step t by the ingoing people $\sum_j w_{j,i}$ and the outgoing $\sum_j w_{i,j}(t)$.

Metapopulation models that aim to study real-world scenarios and fit actual epidemic breaks need to be infused by mobility data. Flight data between airports and telecommunications data are suitable for this purpose. However, there is not always the possibility to get such accurate data. Other models such as the radiation model [31] and the gravity model [6] have served as approximations to the real mobility. In Ref. [34], there is an evaluation of the role of different mobility proxies when modeling epidemics.

Another critical component in the metapopulation models are the parameters that govern the epidemic. The parameters of the epidemic drive the central quantity in epidemics, the reproduction number R_0. The definition of R_0 and the role the parameters play is discussed in the next section.

2.1.4 Reproduction number of an epidemic

The basic reproduction number R_0 is the average number of new cases caused by a typical infectious individual in a completely susceptible population. The importance of R_0 lies in the interpretation of its numerical value. If $R_0 > 1$ it is considered that a disease can invade (it is growing), if $R_0 < 1$ it cannot (it is shrinking) [16]. The last interpretation, especially throughout the coronavirus epidemic, has guided the efforts of the authorities. When $R_0 > 1$, the authorities introduced new NPIs to drive it below 1 [38].

In the SIR model used as an example in Section 2.1.1, R_0 is obtained by linearizing the system of differential equations in Equations (1) when $I \approx 0$. The result of the linearization results in the following equation,

$$\frac{dI}{dt} \approx (\beta - \gamma)I,$$

whose solution is,

$$I(t) \approx I(0)e^{(\beta-\gamma)t}.$$

From the last expression, one can note that the disease will grow exponentially only when $\beta - \gamma > 0$. Hence, for the disease to grow [36],

$$R_0 = \frac{\beta}{\gamma} > 1.$$

More complicated models, such as the compartmental metapopulation models, use a similar concept to calculate R_0; it is calculated through the spectral radius of the Next Generation Matrix [29], which also uses the linearization of the system of differential equations that describes the epidemic.

Analyzing the expression obtained above for R_0 of the SIR model, the parameter γ^{-1} represents the average recovery period of an infectious individual. A longer recovery period will yield a larger R_0 with a fixed transmission rate β. If an infectious individual takes longer to recover or transition to a state where it can no longer keep infecting, it will infect more individuals driving R_0 up. On the other hand, β is

proportional to R_0. Thus with a fixed γ, if the transmission rate increases, also R_0 will and vice versa.

Sometimes, from previous studies, one can obtain the values for parameters γ and β (or any other extra parameter) associated with a disease. Depending on the purpose of the model, one could also fix R_0 and obtain a hypothetical β with a fixed γ. Otherwise, if the disease parameters are not available from the literature, it is necessary to estimate them through an inference method. An inference method that has been used recently for this purpose is analyzed next in Section 2.2.

2.2 Approximate Bayesian Computation (ABC)

Approximate Bayesian Computation (ABC) is an alternate approach to Markov Chain Monte Carlo (MCMC) techniques used for Bayesian inference. With Bayesian inference, one can draw conclusions from unobserved parameters, given some known or observable quantities, using probability models and Bayes theorem [32].

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)};$$

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

where y are the observable quantities or data and θ the unobserved parameters. The term $p(y|\theta)$ is called likelihood, $p(\theta)$ the prior and $p(\theta|y)$ the posterior distribution. The term $p(y)$, with fixed y, does not depend on θ; thus, it can be considered a constant. Just as the name suggests, the prior distribution summarizes our prior knowledge of the unknown parameter θ. In Bayesian inference, the objective is to derive the posterior distribution $p(\theta|y)$ either in a closed form or with some approximation. Using the estimated posterior distribution, one can make predictions on some unseen data $\tilde{y}$,

$$p(\tilde{y}|\theta, y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta.$$

The distribution $p(\tilde{y}|\theta, y)$ is called the posterior predictive distribution.

MCMC techniques approximate the distribution $p(\theta|y)$ directly, this distribution is also called the target distribution. Some of the most notable examples of algorithms that are part of the MCMC family are the Gibbs sampler [12], the Metropolis-Hastings algorithm [9], [10], and the Hamiltonian Monte Carlo algorithm [15]. MCMC algorithms rely on evaluating the likelihood function, while ABC algorithms do not. Hence, ABC algorithms are sometimes also called likelihood-free techniques.

ABC is used in problems where the likelihood of $p(\theta|y)$ is unavailable. The likelihood may be unavailable either for mathematical reasons when it is not available in closed form, or for computational reasons when it is too expensive to make an approximation [30]. Rubin proposed the first ideas behind ABC in 1984 [13]. Rubin proposes an algorithm that chooses relevant statistics to summarize distributions and uses the summaries to evaluate how well proposed posterior predictive distributions resemble the observed data. In essence, this is the spirit behind ABC, namely to find

a set of parameters with the ability to produce predictions that resemble as much as possible the observed data.

The first formal ABC algorithm was proposed by Pritchard et al. in 1999 [18] (see Algorithm 1), although it builds on an idea proposed by Tavaré et al. in 1997 [17].

Algorithm 1: Outline of the original ABC rejection algorithm proposed in Ref. [18].

$N \leftarrow$ desired number of samples;
$n \leftarrow 0$ Current number of samples;
while $n < N$ **do**
    Sample θ' from prior $p(\cdot)$;
    Simulate $\tilde{y}$ using $f(\theta')$;
    if $d(S(\tilde{y}), S(y)) < \epsilon$ **then**
        Save θ' as one accepted set of parameters;
        $n \leftarrow n + 1$;

In Algorithm, 1 $f(\cdot)$ is a function that can produce data $\tilde{y}$ that is similar to the observed data y given a set of parameters θ. $S(\cdot)$ is a function that defines one or more summary statistics that will be applied to the simulated and observed data. Finally, $d(\cdot, \cdot)$ is a distance measure over the summary statistics. If the distance between the summaries is less than some small threshold ϵ, the proposed set of parameters will be accepted. The idea is that if the summary statistics $S(\cdot)$ are representative and ϵ is small enough, then the generated parameters will approximate the posterior distribution $p(\theta|y)$.

The algorithm described was originally proposed in studies related to DNA [17] and chromosomes [18], but it has found success under different settings. There is a good compilation of several studies that use ABC in evolutionary biology and ecology in Ref. [28]. Given the success of ABC, there have been modifications and improvements made to the original algorithm. Most of them have been reviewed in Refs. [28], [30], [37].

In this Thesis, ABC is going to be used to calibrate a compartmental metapopulation model to the reported cases of coronavirus in Finland. In section 2.3, there is a review of some studies that have previously used ABC for epidemics simulation and other relevant work on the modeling of the coronavirus epidemic.

2.3 Coronavirus modeling

The coronavirus pandemic has brought a lot of renewed attention to the field of epidemiology. As a result, there are many studies and a considerable coronavirus interest, how to model it, and the effects NPIs had in society [77]. In this section, I will focus on presenting some studies done in the context of coronavirus and have the same objective as this Thesis: to evaluate the effect of NPIs and their role throughout the epidemic.

The term Non-Pharmaceutical Intervention (NPI) refers to a wide range of measures, from bottom-up measures that emerge from the society in response to the disease to top-down restrictions imposed by the authorities. The first ones include better personal hygiene, avoidance of social gatherings, and social distancing. The latter include border closures, traffic restrictions, lockdowns, and school closures [77]. Sometimes, there is no clear distinction to the origin of an NPI taken by a person, e.g., an individual may use a face mask because of a natural fear of the disease or because the government mandates it. However, no matter the origin, using a face mask is considered an NPI. Moreover, sometimes top-down NPIs can influence the adoption of bottom-up measures. For example, when a government mandates the border closure of a single region, it may indirectly influence people in other regions to enforce stricter social distancing measures.

At the beginning of the coronavirus epidemic, in 2020, NPIs were the primary control measures, both emerging from the society and mandated by governments. Therefore, understanding their effects on the outcome of the epidemic became crucial, and can also be beneficial for future epidemics. According to a literature compilation of 2020 [77], the three most cited studies at the time were Refs. [45], [52], [59]. All of them study the effect of NPIs, how they affected the course of the epidemic in different contexts, and pose counterfactual scenarios. Counterfactual scenarios analyze what would have happened in hypothetical situations with different restriction levels introduced by NPIs.

The study in Ref. [45] uses a metapopulation stochastic compartmental model seeded by flight and local epidemic information. The model is fitted to data using ABC. Subsequently, the authors perform counterfactual scenarios to reduce the interaction inside each region by manually altering the effective reproduction number R_{eff}. The authors also reduce traffic between regions by reducing the number of flights. The approach is very similar to the approach that will be taken later in this Thesis within Finland. Their findings suggest that reducing traffic between regions delayed the spread of the disease to other regions. However, after this first effect, infectious cases traveled out of the infectious region without being detected. Moreover, the authors in Ref. [45] expect that interventions aimed at reducing the transmission are more beneficial than travel restrictions in the long term.

Nonetheless, there are differences between the counterfactual scenarios proposed in Ref. [45] and the ones proposed in this Thesis. The scenarios that are proposed in this Thesis attempt to be as close to an alternate reality as possible using the available data. In contrast, Ref. [45] does a manual downscaling of the parameters and amount of traffic in flights.

The study in Ref. [59] presents a compartmental model that considers the age of individuals and the social mixing patterns. The objective of the study is to analyze the effect of social distancing measures in the city of Wuhan. The model of the study does not consider different geographic regions but analyzes the effect of interventions on the result of the epidemics through theoretical contact matrices. As the name implies, a contact matrix describes the contacts of an individual in high mixing places, where the disease quickly spreads, such as schools, workplaces, and residences. The authors in Ref. [59] use the theoretical matrices to produce counterfactual

scenarios that scale down the contacts of a contact matrix in regular times, without restrictions. With the theoretical contact matrices Ref. [59] theorizes on what would have happened by introducing different levels of restrictions at different times. The results in Ref. [59] suggest that physical distancing measures in Wuhan were crucial to reduce the number of infections and delay the epidemic's peak. Although this work does not use a compartmental model, it does present different counterfactual scenarios. It does it in a similar way to [44] by scaling the level of restrictions.

The counterfactual scenarios in Ref. [52] compare the number of deaths when the governments implement NPIs, which was the case in 2020, to a scenario in which they did not introduce NPIs. In this work, the R_{eff} for different countries is estimated through a Bayesian approach, although not using ABC. The model and estimation of R_{eff} are informed by the number of reported deaths in eleven European countries. The authors assume that every implemented NPI affects contacts; hence they infer a new R_{eff} every time a country introduces a new major NPI. For the counterfactual scenario, they keep the estimated R_{eff} for the beginning of the epidemic, i.e. they assume that none of the introduced NPIs affected the original R_{eff}, and calculated the number of deaths under this scenario. Then, they compared the number of deaths from the counterfactual scenario to the original number of deaths calculated by the model that does take into account NPIs. This way the study presents the number of averted deaths due to NPIs, which is significant. Based on the number of averted deaths in the model due to NPIs, the authors conclude that NPIs were crucial to reduce $R_{\text{eff}} < 1$. Furthermore, they find that the NPI that had the most substantial effect was a generalized lockdown, in which people is only permitted to do essential trips, a measure whose ultimate goal is to prevent as many contacts as possible.

Different kinds of data inform the proposed models. For example, mobile phone data informs the model proposed in Ref. [44]. The authors present a compartmental metapopulation model with mobile phone data in which people go from their residence to an outside location, maybe a store, a workplace, or a restaurant. First, the parameters of the model are fitted to match the reported cases. After fitting the model, they show counterfactual scenarios that can change the mobility in two ways. The first way to change the mobility is to reduce it by some fraction. The second way is to shift the date when the government introduced major NPIs. Using these two ways of changing mobility, they show the change in the number of reported cases with the counterfactual scenarios with respect to the fitted model. They find that policy-makers must be careful with the timing and the reopening society; a prompt and uncontrolled reopening may lead to a rise in the cases due to the increase of contacts. Another aspect that the authors analyze is the change in the number of visits to businesses, such as restaurants or convenience stores, with each counterfactual scenario. The purpose is to analyze which businesses were the most affected by the NPIs. This analysis is possible due to the high specificity of the mobile phone data. For this Thesis, mobile phone data is available but not as detailed as in Ref. [44]; with the available mobile phone data, I can build counterfactual scenarios at a regional level, as explained in the upcoming chapters.

The authors in Ref. [54] present a similar study that shows counterfactual scenarios based on mobility but at the regional level. Here, the authors present

a compartmental metapopulation model for Italy. In the model, Italy is divided into 107 sub-regions connected through mobility data derived from census mobility fluxes. The model is calibrated to data to match the number of hospitalized cases and the number of deaths using MCMC. The work presents the counterfactual scenarios as changes in mobility, much like this Thesis. In Ref. [54], the authors present the number of hospitalizations that would have occurred if NPIs had not been implemented. Similar to Ref. [54], the authors of Ref. [52] recognize the lockdown imposed by the government as the main factor in reducing transmission rates. They also recognize that the available data limit the extent and accuracy study and point out the fact that other authors have successfully used mobile phone data for their studies.

Throughout this section, I have presented studies focused on the study of NPIs and their role during the coronavirus epidemic. This Thesis will focus on the same objective under a different context and with different data. As pointed out, the available data often limit the extent of this type of studies. In this case, there are mobility data available that inform how people travel between municipalities in Finland. Using the available data and a compartmental metapopulation model calibrated with ABC, the Thesis poses as realistic counterfactual scenarios as possible.

In the following Chapter 3, I will describe the methods used to calibrate the epidemic model. The methods presented in the next chapter are based on the theory and studies revised in this chapter.

3 Compartmental metapopulation model

The objective of modeling the spread of COVID-19 is to understand and explain it through the parameters and factors in the model. In this Thesis, the main factor under analysis is the mobility between the regions of Finland, which motivates the use of a metapopulation model; metapopulation models study the spread of disease by separating it into regions and observing the interactions between them.

This chapter will present the epidemic model used to describe the spread of coronavirus in Finland and how the model is calibrated to data. At the end of the chapter, I show the result of calibrating the model to spring 2020, when COVID-19 arrived in Finland and had its first outbreak.

3.1 Epidemic model

The coronavirus epidemic is modeled using a stochastic compartmental metapopulation model published originally in Ref. [47]. In the model, a region, such as a country, is divided into K sub-regions. Each sub-region k has its own set of compartments. The model divides the individuals into Susceptible to the disease (S), Exposed (E) individuals to the disease but not yet infectious, Presymptomatic infectious (P), Asymptomatic infectious (A), Infectious and symptomatic (I), and Removed (R). A more traditional $SEIR$ model, with only three compartments, is not used since studies suggest that both Asymptomatic [40] and Presymptomatic [65] individuals contribute to the spread of the disease. Therefore, using P and A compartments brings the model closer to the actual progression of an individual in the disease than using a SIR or an $SEIR$ model. Figure 3 shows a graphical representation of the $SEPAIR$ compartments and the parameters of the model.

The set of stochastic $SEPAIR$ equations of the model that determine the situation in region k at time $t + \Delta t$ is the following,

$$\begin{aligned}
S_k(t+\Delta t) &= S_k(t) - X_k^{(1)}(t) - X_k^{(2)}(t) - X_k^{(3)}(t);\\
E_k(t+\Delta t) &= E_k(t) + X_k^{(1)}(t) + X_k^{(2)}(t) + X_k^{(3)}(t) - X_k^{(4)}(t);\\
P_k(t+\Delta t) &= P_k(t) + X_k^{(5)}(t) - X_k^{(6)}(t);\\
A_k(t+\Delta t) &= A_k(t) + X_k^{(4)}(t) - X_k^{(5)}(t) - X_k^{(8)}(t);\\
I_k(t+\Delta t) &= I_k(t) + X_k^{(6)}(t) - X_k^{(7)}(t);\\
R_k(t+\Delta t) &= R_k(t) + X_k^{(7)}(t) + X_k^{(8)}(t),
\end{aligned} \tag{2}$$

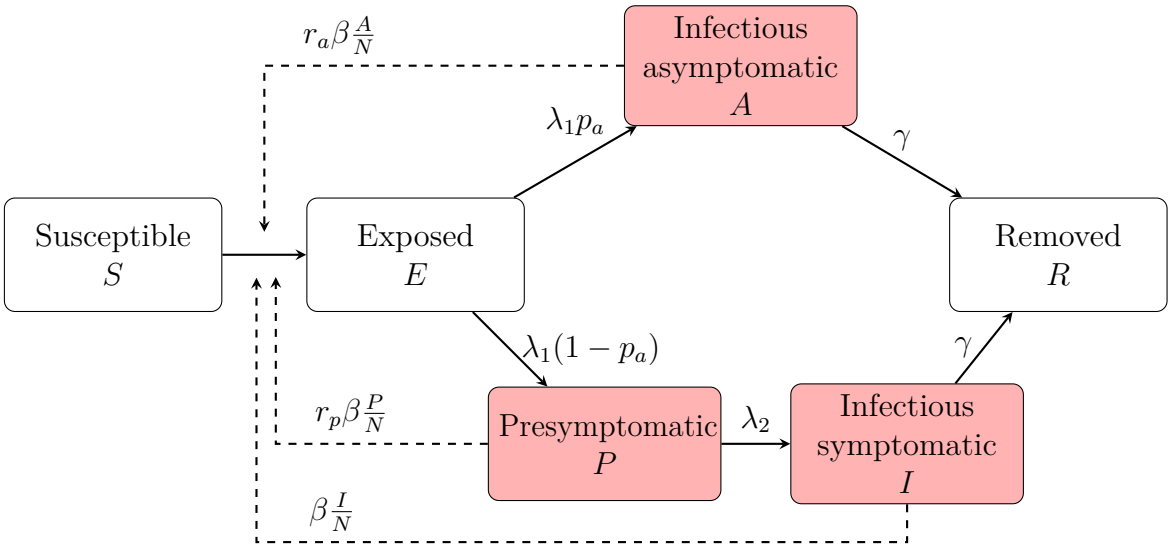

Figure 3: Graphical representation of the stochastic compartmental $SEPAIR$ model. The individuals from the compartments in red can spread the disease. A Susceptible individual S may become infected when interacting with individuals from the Presymptomatic compartment P, the Asymptomatic compartment A, or the Infectious compartment I. If infected, the Susceptible individual will transition to the Exposed compartment E and after that progress to Asymptomatic or Presymptomatic. If the individual becomes Presymptomatic, it will eventually progress to the Infectious symptomatic compartment. At the end of the disease, the Asymptomatic and Infectious individuals will end up in the Removed compartment R. One individual can be Removed either because the person recovered from the disease or because he or she died. The probability of an individual transitioning from one compartment to the next one is indicated on top of the arrows of the diagram. The values for the parameters of the model are discussed later in Section 3.3.

in which,

$$
\begin{aligned}
X_k^{(1)}(t) &\sim \text{Binom}\left(S_k(t), \beta_k(t)\Delta t \frac{I_k(t)}{N_k}\right);\\
X_k^{(2)}(t) &\sim \text{Binom}\left(S_k(t), r_a\beta_k(t)\Delta t \frac{A_k(t)}{N_k}\right);\\
X_k^{(3)}(t) &\sim \text{Binom}\left(S_k(t), r_p\beta_k(t)\Delta t \frac{P_k(t)}{N_k}\right);\\
X_k^{(4)}(t) &\sim \text{Binom}(E_k(t), \lambda_1\Delta t);\\
X_k^{(5)}(t) &\sim \text{Binom}(X_k^{(4)}(t), 1-p_a);\\
X_k^{(6)}(t) &\sim \text{Binom}(P_k(t), \lambda^{(2)}\Delta t);\\
X_k^{(7)}(t) &\sim \text{Binom}(I_k(t), \gamma\Delta t);\\
X_k^{(8)}(t) &\sim \text{Binom}(A_k(t), \gamma\Delta t).
\end{aligned}
\tag{3}
$$

In Equations (3), Binom(n, p) is the binomial distribution with n trials and success probability p. As mentioned earlier, individuals in the P and A compartments also contribute to the spread of the disease. The factors r_a and r_p account for this contribution. The factor r_a is the relative infectiousness an individual in the A compartment has with respect to an individual in the I compartment, i.e., setting $r_a = 1$ would mean an individual in the A compartment is as infectious as an individual in the I compartment. In turn, r_p represents the same but for the individuals in the P compartment. The parameter γ^{-1} is the average number of days an individual stays infectious before transitioning to R. Similarly, λ_1^{-1} is the average latent period and λ_2^{-1} the average presymptomatic period, both expressed in days. The parameter p_a is the probability an individual has of becoming part of the A compartment. Finally, the parameter β is the transmission rate an individual from I has. As expressed by the equations, β varies across the k regions and time step t. Thus, the variation of the value of β captures the progression of the disease at different regions and times. Finally, the parameter Δt is the time interval between epidemic realizations expressed in days. For example, if the time interval between t and $t + \Delta t$ is 6 hours, then $\Delta t = \frac{1}{4}$ since 6 hours is one-fourth of a day.

There is a set of $SEPAIR$ compartments per region k. Furthermore, at every time step t, people are traveling between the k regions. To have travelers between regions allows capturing the situation in which an infected person coming from a region k may go to another region l, interact with the population in region l, and transmit the disease. In Figure 4, there is a graphical representation of the model and the travelers with $K = 3$ sub-regions.

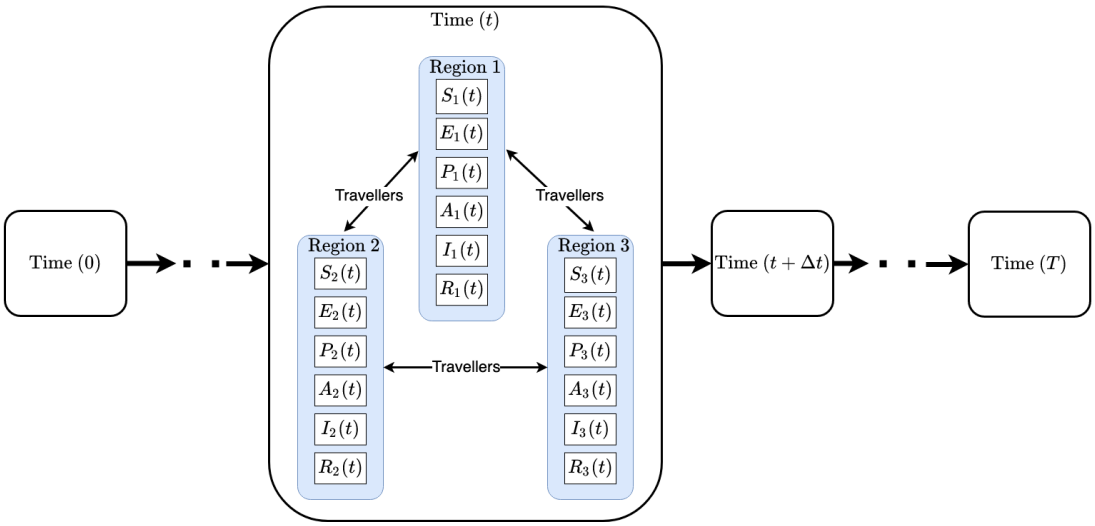

Figure 4: Interaction of $K = 3$ sub-regions in the compartmental model at time t. Each region, at every time step t, has its own set of $SEPAIR$ compartments, and there is a certain number of people traveling between the regions.

The number of people present in region k, and the people traveling into that region at time t, determine the number of people in each compartment at time $t + \Delta t$ and region k as seen in Equations (2). How traveling between regions affects the

population in each compartment is discussed next in Section 3.1.1.

3.1.1 Travelling between regions

Travelling affects the population of region k at every time step t. This is because people travel between regions before performing the stochastic step from t to $t + \Delta t$. The number of people traveling is taken from an Origin-Destination (O-D) matrix. Going back to the example where $K = 3$ from Figure 4, an O-D matrix at time t would look like the matrix in Table 1. The entry of the O-D matrix at cell with row k and column l indicates the number of travelers going from region k to region l.

O-D(t)	Region 1	Region 2	Region 3
Region 1	3	10	20
Region 2	5	8	13
Region 3	15	4	7

Table 1: Example of an O-D matrix with $K = 3$ regions at time step t. The values of the matrix are not based on real data. In the matrix, each cell shows the number of travelers between two regions. For example, 10 travelers are going from Region 1 to Region 2 and 20 from Region 1 to Region 3.

As mentioned before, traveling occurs between step t and step $t + \Delta t$. For example, this allows for people in the infected compartment in region k ($I_k(t)$) to travel to a new region l and be part of population $I_l(t)$ before the realization of time step $t + \Delta t$. Thus, the infectious travelers $I_k(t)$ will affect the epidemic outcome in region l by infecting people. Therefore, at every time step t, the number of people at compartment $C \in \{S, E, P, A, I, R\}$ in region k is determined by the number of residents of k currently in k and the number of visitors from the rest of the regions l in k,

$$C_k(t) = \sum_{l \in K} C_{l,k}(t).$$

There are two aspects to take into consideration when moving travelers between regions:

- There are two types of population in region k at all times: people resident in the region and the travelers in the region.
- Every person at region k, resident or traveler, may belong to any of the six compartments.

A person in region k may belong to one of $6K$ classes. It may be in one of the the six compartments of the model and belong to two types of population (resident and traveler). The O-D matrix shows the total number of travelers between regions. However those travelers may belong to any of the $6K$ classes. Hence the need for a prioritization algorithm that chooses individuals from the $6K$ classes in an specific

order and moves the individuals between the regions. The way the algorithm handles traveling, considering that an individual may belong to $6K$ classes, has two cases, when $t = 0$ and when $t > 0$. Algorithm 2 shows the first case, when $t = 0$, and Algorithm 3 the second case.

When $t = 0$, before performing the first stochastic step, all the population is in the S compartment. Since all the population is in S and there are still no travelers from the other regions, the algorithm will move people between the regions according to the O-D matrix at $t = 0$. However, for $t > 0$, some people may have gotten infected, and there will be travelers present in the regions; this is why there is a separate algorithm to handle this case.

Algorithm 2: Algorithm to distribute the travelers at $t = 0$ between the K regions. The algorithm first moves the Susceptible S individuals between the regions before taking the first time step.

```
K ← number of regions;
O-D (0) ← O-D matrix at t = 0;
for k ← 1 to K do
    for l ← 1 to K do
        /* Get the c travellers going from k to l          */
        c ← O-D_{k,l}(0);
        /* Remove c from k and add to l                    */
        S_{k,k}(0) ← S_{k,k}(0) − c;
        S_{l,k}(0) ← S_{l,k}(0) + c;
```

On the other hand, when $t > 0$ (Algorithms 3 and 4), the algorithm checks how many travelers c need to be moved from region k to l according to the O-D matrix. Then, the algorithm will pick residents and travelers in region k until getting c to move to region l. The order in which the algorithm will pick the people is the following:

1. Travelers from l currently in k.
2. Residents of k currently in k.
3. Travelers from other regions that are not l currently in k.

The data is not perfect and is imbalanced. This means that, at the end of a day, all the people that travel from region k to region l may not return to the origin region. Hence, if the algorithm is ran for a long time horizon, a region may get depleted of population. If this situation happens, and there is still c people required to go from k to l, the algorithm will create the missing people as residents of k in the S compartment and immediately send them to l.

Note that in Algorithms 3 and 4, in several steps, it is mentioned that it is necessary to pick a specific number n of individuals from the compartments $S_*, E_*, P_*, A_*, I_*, R_*$. The number of individuals s, e, p, a, i, r from each compartment respectively must be

Algorithm 3: Algorithm to distribute the travelers at every time step t between the K regions. At any given time step, there can be v visitors from region l in region k. On the other hand, according to the O-D matrix, c people need to go from k to l. Following these conditions, there can be three cases (1) $c = v$, the same amount of people returning to l that travel from l, (2) $c < v$, fewer people returning than traveled, i.e. there will be long-stayers, and (3) $c > v$, more people returning than traveled. This algorithm presents the first two cases, and Algorithm 4 presents the last one.

$K \leftarrow$ number of regions;
$T \leftarrow$ time step $t \neq 0$;
O-D $(t) \leftarrow$ O-D matrix at t;
for $t \leftarrow \Delta t$ **to** T **do**
  for $k \leftarrow 1$ **to** K **do**
    for $l \leftarrow 1$ **to** K **do**
      /* c travellers going from k to l */
      $c \leftarrow$ O-D$_{k,l}(t)$;
      /* v travellers currently in k that belong to l */
      $v \leftarrow S_{l,k}(t) + E_{l,k}(t) + P_{l,k}(t) + A_{l,k}(t) + I_{l,k}(t) + R_{l,k}(t)$;
      if $c = v$ **then**
        /* Return every traveller from l to home */
        Add $S_{l,k}(t), S_{l,k}(t), E_{l,k}(t), P_{l,k}(t), A_{l,k}(t), I_{l,k}(t), R_{l,k}(t)$ to the respective compartments in l ;
        Remove $S_{l,k}(t), S_{l,k}(t), E_{l,k}(t), P_{l,k}(t), A_{l,k}(t), I_{l,k}(t), R_{l,k}(t)$ from the respective compartments in k ;
      /* Less people going to l than current visitors */
      else if $c < v$ **then**
        Pick randomly s, e, p, a, i, r individuals from $S_{l,k}(t), E_{l,k}(t), P_{l,k}(t), A_{l,k}(t), I_{l,k}(t), R_{l,k}(t)$ respectively such that $s + e + p + a + i + r = c$;
        Add s, e, p, a, i, r to the respective compartments in l ;
        Remove s, e, p, a, i, r from the respective compartments in k ;
      /* More people going to l than current visitors */
      else if $c > v$ **then**
        This is a special case since also resident individuals need to be chosen to go from k to l. See Algorithm 4 ;

picked such that $s+e+p+a+i+r = n$. The algorithm performs this operation using the Hypergeometric(N, K, n) distribution that describes the number of k successful draws from an object with a specific feature, in n draws without replacement, from a total population size N that contains exactly K objects with the required feature. For example, to pick s individuals from the S_* compartment, the distribution parameters would be Hypergeometric(N_*, S_*, n), where $N_* = S_* + E_* + P_* + A_* + I_* + R_*$.

Algorithm 4: Last case of Algorithm 3. More people returning to l than people who traveled from l, i.e., $c > v$

$c \leftarrow$ O-D$_{k,l}(t)$;
$v \leftarrow S_{l,k}(t) + E_{l,k}(t) + P_{l,k}(t) + A_{l,k}(t) + I_{l,k}(t) + R_{l,k}(t)$;
/* Return every traveller from l to home */
Add $S_{l,k}(t), S_{l,k}(t), E_{l,k}(t), P_{l,k}(t), A_{l,k}(t), I_{l,k}(t), R_{l,k}(t)$ to the respective compartments in l ;
Remove $S_{l,k}(t), S_{l,k}(t), E_{l,k}(t), P_{l,k}(t), A_{l,k}(t), I_{l,k}(t), R_{l,k}(t)$ from the respective compartments in k ;
/* m people still need to be sent from k to l */
$m \leftarrow c - v$;
/* p residents of k currently in k */
$p \leftarrow S_{k,k}(t) + E_{k,k}(t) + P_{k,k}(t) + A_{k,k}(t) + I_{k,k}(t) + R_{k,k}(t)$;
/* Enough residents of k to be sent to l */
if $p \geq m$ **then**
    Pick randomly s, e, p, a, i, r individuals from $S_{k,k}(t), E_{k,k}(t), P_{k,k}(t), A_{k,k}(t), I_{k,k}(t), R_{k,k}(t)$ respectively such that $s + e + p + a + i + r = m$;
    Add s, e, p, a, i, r to the respective compartments in l ;
    Remove s, e, p, a, i, r from the respective compartments in k ;
else
    /* Send all residents from k to l */
    Add $S_{k,k}(t), E_{k,k}(t), P_{k,k}(t), A_{k,k}(t), I_{k,k}(t), R_{k,k}(t)$ to the respective compartments in l ;
    Remove all residents from the compartments in k ;
    /* m people still need to be sent from k to l */
    $m \leftarrow m - p$;
    /* Travellers of all regions k' that are not from l */
    $v^{-l} \leftarrow S_{k',k}(t) + E_{k',k}(t) + P_{k',k}(t) + A_{k',k}(t) + I_{k',k}(t) + R_{k',k}(t)$;
    /* Enough travellers from all k' to be sent to l */
    if $v^{-l} \geq m$ **then**
        Pick randomly s, e, p, a, i, r individuals from $S_{k',k}(t), E_{k',k}(t), P_{k',k}(t), A_{k',k}(t), I_{k',k}(t), R_{k',k}(t)$ respectively such that $s + e + p + a + i + r = m$;
        Add s, e, p, a, i, r to the respective compartments in l, keep their resident location k' ;
        Remove s, e, p, a, i, r from the respective traveller compartment of k' in k ;
    else
        Add $S_{k',k}(t), E_{k',k}(t), P_{k',k}(t), A_{k',k}(t), I_{k',k}(t), R_{k',k}(t)$ to the respective compartments in l, keep home location k' ;
        Remove all travellers from all regions k' in all compartments in k ;
        /* m people still need to be sent from k to l */
        $m \leftarrow m - v^{-l}$;
        Create m Susceptible individuals in region k, $S_{k,k}$;
        Send the newly created $S_{k,k}$ to region l ;

The algorithms describe how to handle the travelers between regions specified by the O-D matrices. Data from a major telecommunications operator in Finland is processed to construct the O-D matrices.

3.2 Mobility data

Mobility data available from a major telecommunications operator in Finland is processed to determine the O-D matrices needed for the metapopulation model. The mobility data consists of the number of people moving between municipalities in Finland from 1 February 2019 to 31 May 2019 and a second period from 1 February 2020 to 31 May 2020. The two time periods are the same but in different years, 2019 and 2020. Therefore, the periods allow comparing the behavior of people when traveling between municipalities in a year without coronavirus (2019) and during the first wave of the epidemic (2020).

The data from the operator presents the number of travelers between two municipalities on a specific day and determined hour bin. All days have 4 hour bins, from 0:00 to 5:59, 6:00 to 11:59, 12:00 to 17:59 and 18:00 to 23:59. On the other hand, Finland is divided into 309 municipalities. Thus on any given day there can be four entries, one per hour bin, between any two municipalities. Table 2 presents an example of the mobility data.

Date	Origin Mun	Destination Mun	Hour bin	Trips
2019-02-01	Mun A	Mun B	0:00-5:59	10
2019-02-01	Mun A	Mun B	6:00-11:59	50
2019-02-01	Mun A	Mun B	12:00-17:59	50
2019-02-01	Mun A	Mun B	18:00-23:59	10

Table 2: Example of the mobility data. The values shown in the rows are not real. The first column of the Table indicates the date on which the trips occur. The second and third columns are the origin and destination municipalities of the trips, respectively. Then, the fourth column indicates the hour bin of the trips. Finally, the fifth and last column shows the number of trips between the origin and destination municipalities for that day and hour bin.

3.2.1 Data preprocessing

To obtain the data in the format of Table 2, the telecommunications operator counts the number of registered users it has that started moving in Municipality A and ended up in Municipality B at the specific day and hour bin. If the user starts moving and then does not move for more than 20 minutes, the user ends its trip. The operator also scales the number of trips it registered to represent the total market instead of only its market share and present the data as accurately as possible. Lastly, the operator also mentions that, for some days, there are municipalities with low signal quality. To deal with the low signal quality, the operator removes the trips originating

from municipalities with low signal quality, effectively removing the municipality as an origin for the day in which the signal had low quality. After analyzing the data, almost all days there is at least one removed municipality. However, since there are 309 municipalities, the removed municipalities do not affect the overall quality of the data, especially when aggregated. Hence I can directly use the data for most days, even if some days have these removed municipalities. Nonetheless, the number of removed municipalities causes bias in the data for the minimum number of days. The days that are not considered from the data are determined by visually inspecting the number of total trips between all municipalities per day (see Figure 5).

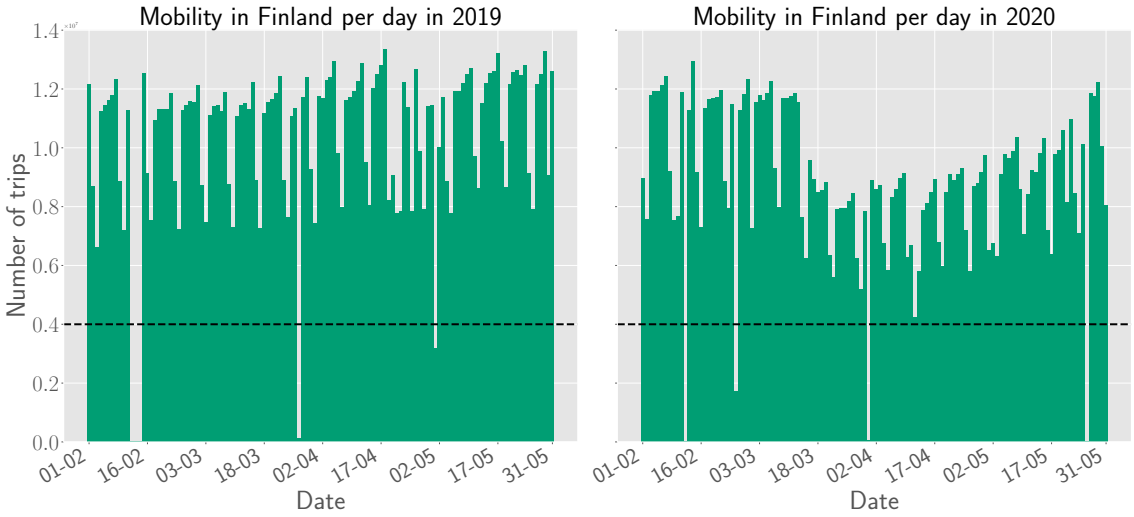

Figure 5: Mobility in Finland summing the total number of trips per day from all municipalities. The number of trips per day in 2019 is in the left panel and 2020 in the right panel. The figures reveal some days with out-of-the-ordinary low mobility in both years. These days are not considered if they are not above a threshold; the threshold is the discontinuous black line in the figures.

A threshold is set in order to discard days with low mobility that are out of the ordinary. If the total number of trips in a day is less than 4×10^6, the day is discarded. Also, I verified that the discarded days had a significant number of reported missing municipalities due to low signal quality. The removed days have a significantly low number of trips compared to the average number of trips per day in a year. The average number of trips per day in 2019 is 10,340,396. The days that are discarded from data are 2019-02-12 with 8,398 trips, 2019-02-13 with 2873 trips, 2019-02-15 with 20,325 trips, 2019-03-27 with 115,671 trips and 2019-05-01 with 3,164,628. For 2020 the average number of trips is 8,834,824. The discarded days are 2020-02-12 with 3,731 trips, 2020-02-25 with 1,703,217 trips, 2020-03-31 with 40,520 and 2020-05-26 with 3,653 trips.

The analysis done with the metapopulation model requires mobility data for all days. Instead of completely removing days with low signal quality, they are replaced with an average measure of mobility. In specific they are replaced with the average number of trips for the weekday of the removed day. From the data is observed

that mobility patterns for a weekday are similar in a given period. For example, all Mondays of autumn have similar mobility patterns, and the same applies for the rest of the weekdays in the autumn, all Tuesdays, Wednesdays, and so forth. However, it may not apply for a different period such as all summer Mondays. The difference in mobility patterns between two periods, for example, autumn and summer, is because during autumn the schools are open and a few people go out on vacation, whereas during summer, the schools close, and a significant amount of people go on vacation. Using this assumption that all weekdays of a given period have similar mobility patterns, the days with low mobility are replaced with the average number of trips for that weekday in the correspondent period. The average number of trips $\bar{tr}$ between any two regions k and l, for a weekday x and hour bin h, is calculated as follows,

$$\bar{tr}_{k,l}^{(x,h)} = \frac{1}{D} \sum_{d \in D} tr_{k,l}^{(d,h)},$$

where D groups all the weekdays for a certain period together, e.g., $D = \{\text{Mondays of the autumn}\}$. This way, I constructed an average mobility week $\bar{w}$ in which all the weekdays have the average number of trips between regions, for an hour bin, in a given period.

Using the method that was just described, I built six weeks $\bar{w}$ for six different periods. The periods with similar mobility are the following:

- $\bar{w}^{(2019)}$. The period from 2019-02-01 to 2019-05-31. The whole year 2019 is considered to have similar mobility. Therefore, it is considered a single period.
- $\bar{w}^{(\text{before})}$. The period from 2020-02-01 to 2020-03-15. Period before the coronavirus pandemic. On 2020-03-16, the government declared a state of emergency in Finland [69], in which it applied several mobility restrictions.
- $\bar{w}^{(\text{quarantine})}$. The period from 2020-03-16 to 2020-03-26. The period after the state of emergency was declared. The state of emergency entailed the closure of museums, libraries, and other public spaces. Meetings of more than ten people were prohibited, and also passenger traffic to Finland was suspended.
- $\bar{w}^{(\text{uusimaa})}$. The period from 2020-03-27 to 2020-04-15. The period where the border of the region of Uusimaa was closed. On 2020-03-16, the government ordered the closure of the border of the region of Uusimma [72], restricting ingoing and outgoing traffic. The border closure of Uusimaa was the most significant restriction to mobility during 2020. The border was reopened on 2020-03-27 [71].
- $\bar{w}^{(\text{open})}$. The period from 2020-04-16 to 2020-05-13. The period after Uusimaa border was opened and before the reopening of schools. After the reopening of the border of Uusimaa, there were still some mobility restrictions in effect. Some of them were lifted by 2020-05-14, in which also the government allowed the gradual reopening of elementary schools [51].

- $\bar{w}^{(\text{schools})}$. The period from 2020-05-14 to 2020-05-31. The period after the gradual reopening of elementary schools. There are no more available mobility data from the telecommunications operator after 31 May. There were no other restrictions lifted between the schools reopening on 14 May and 31 May 2020.

The week $\bar{w}^{(2019)}$ is used to replace all the days that have out of ordinary low mobility (2019-02-12, 2019-02-13, 2019-02-14, 2019-03-27, 2019-05-01). The week $\bar{w}^{(\text{before})}$ is used to replace the mobility for days 2020-02-12 and 2020-02-25. Next, the week $\bar{w}^{(\text{uusimaa})}$ is used to replace the mobility of 2020-03-31. Finally, the week $\bar{w}^{(\text{schools})}$ is used to replace 2020-05-26.

3.2.2 Aggregation of mobility data to HCD

How the Finnish Institute for Health and Welfare (in Finnish *Terveyden ja hyvinvoinnin laitos*) (THL) reports the confirmed coronavirus cases is discussed later in detail in Section 3.4. Nevertheless, it is worth to cover how the mobility data will be aggregated. The mobility data are aggregated to match the geographical level at which THL reports the confirmed cases. THL reports the cases at the Hospital Care District (HCD) level. However, the mobility data available is in a finer geographical level, at the municipality, or in Finnish, *kunta* level. HCDs, or in Finnish *sairaanhoitopiirit*, enclose one or more municipalities. A larger geographical level that encloses HCDs is called Hospital Catchment Area (in Finnish *erityisvastuualue*) (ERVA), also mentioned later in Section 3.4.1. To see the different geographical levels in Finland, see Figure 6.

The mobility data is aggregated to the HCD level to match the geographical level at which THL reports the cases. That means first to identify to which HCD each municipality belongs to and then make the summation of trips between the HCDs,

$$tr_{k,l}^{(d,h)} = \sum_{k' \in K'} \sum_{l' \in L'} tr_{k',l'}^{(d,h)},$$

where $tr_{k,l}^{(d,h)}$ are the final aggregated trips from HCD k to l at day d and hour bin h. The variable K' is the set of municipalities that belong to HCD k, and similarly, L' is the set of municipalities that belong HCD to l. The mobility data will still look similar to those entries in Table 2, but instead of having origin and destination municipality, there will be origin and destination HCD.

3.2.3 Comparing mobility between 2019 and 2020

The government enforced a high number of mobility restrictions in 2020 due to the coronavirus epidemic. Comparing the mobility data between 2019 and 2020, one can detect the impact of the restrictions to mobility. To compare the mobility of 2020 to the mobility of 2019, I used the previously constructed week $\bar{w}^{(2019)}$ as a baseline. First, the total number of trips of a day d between every HCD pair k and l is going to be obtained summing the trips of every hour bin,

$$tr_{k,l}^{(d)} = \sum_{h=1}^{4} tr_{k,l}^{(d,h)}.$$

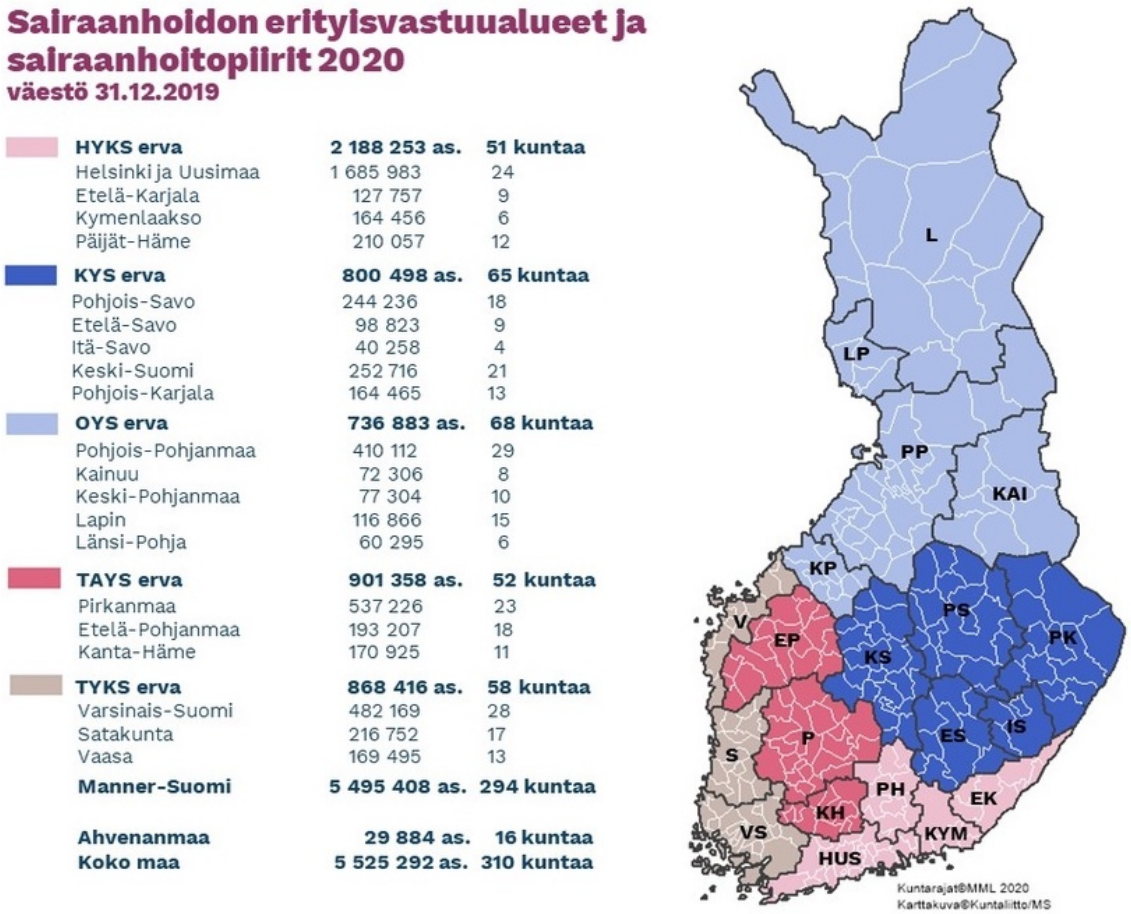

Figure 6: Geographical levels in Finland. The smallest level is a municipality (*kunta* in Finnish); on the map, it is delimited by white borders. The next level is HCD, with black borders between each on the map. Finally, the largest level visible level in the map is Hospital Catchment Area (in Finnish *erityisvastuualue*) (ERVA). Each ERVA has a different color on the map. On the left, the HCDs are listed, indicating how many municipalities belong to them and to which ERVA they belong. The image is taken from the Association of Finnish Municipalities with their permission (*Kuntaliitto*) [57].

The above expression will be applied for all weekdays in $\bar{w}^{(2019)}$ and every day in the mobility data of 2020. Then, the relative change in mobility r in 2020 with respect to 2019 per day d, origin k, and destination l can be obtained by,

$$r_{k,l}^{(d)} = \frac{tr_{k,l}^{(d)} - \bar{w}_{k,l}^{(x)}}{\bar{w}_{k,l}^{x}},$$

where $\bar{w}$ is the average mobility per weekday during 2019 ($\bar{w}^{(2019)}$) and x is the weekday of day d. Last, to get a measure of relative change in mobility per HCD k,

the average relative change for all the trips that originated in k is taken as,

$$r_k^{(d)} = \frac{1}{K} \sum_{l=1}^{K} r_d^{k,l}.$$

Finally, Finland's average relative change in mobility per day can be obtained by averaging $r_k^{(d)}$ across all HCDs. Figure 7 is a comparison of the relative change in mobility during 2020 with respect to 2019 in four regions: Finland, Helsinki and Uusimaa HCD (the most populated HCD), the HCD that most reduced its mobility (Ahvenanmaa), and the HCD that least reduced it (Itä-Savo).

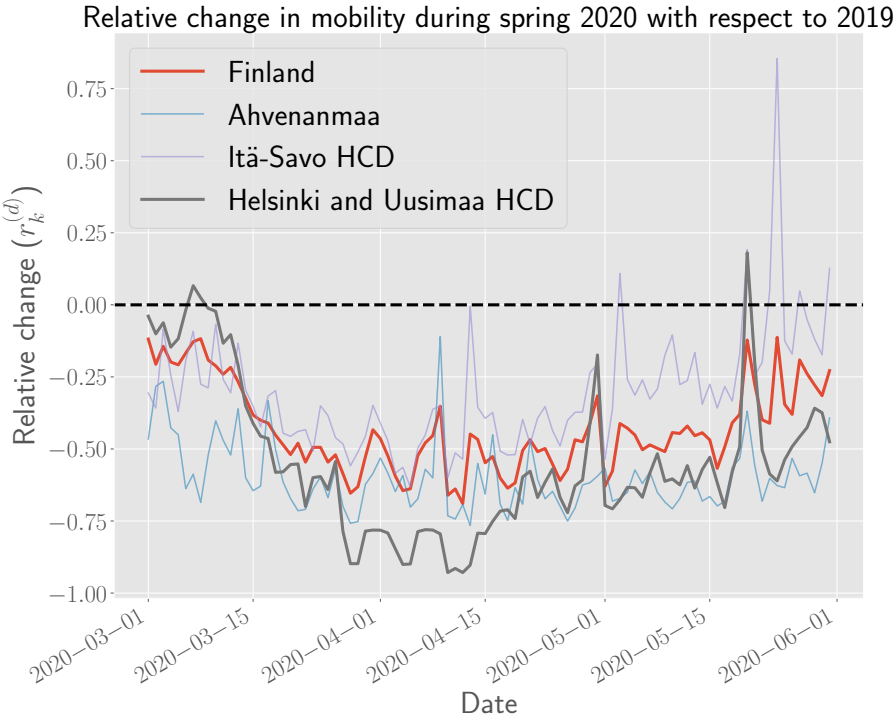

Figure 7: Relative change in mobility in Finland in 2020 with respect to 2019. Marked with a dashed line is the baseline, i.e., the mobility level of 2019. Helsinki and Uusimaa HCD is the most populated HCD, and its relative change is visible in black on the plot. The borders of Uusimaa were closed from 2020-03-27 to 2020-04-15; in the plot, we can see the effect of the mobility restriction. Ahvenanmaa was the HCD that reduced its mobility the most, and Itä-Savo HCD was the one that reduced it the least. The relative change in both can be seen in blue and purple, respectively. In red is the relative change in mobility in the whole of Finland.

Figure 7 presents the effect of the mobility restrictions in Finland. For example, a drop in the relative change can be seen when a state of emergency was declared in 2020-03-16. There was another drop for Helsinki and Uusimaa HCD when the border was closed on 2020-03-27, and then a slight increase when the border opened on 2020-04-15.

The mobility data reflects the mobility restrictions imposed by the government. The aim is to use this informative data and present counterfactual scenarios, i.e.

scenarios that involve modifying the mobility patterns of 2020 and seeing how these changes may affect the epidemic's evolution. One example is the scenario in which there would be no restrictions and the people continue their mobility patterns from 2019. It is expected that the counterfactual scenarios have consequences on the development of the epidemic, taking into account that NPIs were the primary resource of the government to prevent the spread of the disease. The way the counterfactual scenarios are set in detail is discussed in Section 4.1.

First, there is a need to calibrate the model to the epidemic situation in Finland during the first wave of the epidemic. To calibrate the model the mobility data and the values for the parameters in the epidemic model are needed. This section presents the mobility data; the next Section 3.3 will cover the parameters of the epidemic model.

3.3 Parameters of the model for coronavirus epidemic

In the stochastic Equations (3), there are several parameters. The values of those parameters are inherent to the characteristics of the coronavirus disease. Some of the values of the model parameters have been taken from the literature and set in the model. The parameters from the literature are $\lambda_1, \lambda_2, \gamma, r_a, r_p$ and p_a. The parameter Δt was set based on the periodicity of the available mobility data. The mobility data is analyzed in detail in Section 3.2. The parameter $\beta_k(t)$ is inferred from the number of reported cases in Finland. Section 3.5 discusses the method followed to infer the parameter. Table 3 presents the values for all the parameters of the model.

Parameter	Description	Value
$\beta_k(t)$	Transmission rate	Inferred
λ_1^{-1}	Average latent period	3 days
λ_2^{-1}	Average presymptomatic period	2 days
γ^{-1}	Average infectious period	5 days
r_a	Relative infectiousness of an A individual wrt I	0.1
r_p	Relative infectiousness of a P individual wrt I	1.25
p_a	Probability of being asymptomatic	0.4
Δt	Time interval between time steps	0.25 days
K	Number of regions	21 HCDs

Table 3: Parameters and their values in the $SEPAIR$ epidemic model. The values for the parameters $\lambda_1, \lambda_2, \gamma, r_a, r_p$ and p_a are from Ref. [49]. The parameter $\beta_k(t)$ is inferred from data, see Section 3.5. The parameter Δt is set based on the available mobility data. K is also set based on the available data; in this case, the confirmed cases reported by THL. More information about the reported cases is available in Section 3.4.

Finland is divided into sub-regions called Hospital Care Districts (HCDs). There are 21 HCDs, hence $K = 21$ in the model. The main reason to use HCDs as sub-regions of Finland and not apply the model to a finer level, for example, municipalities,

is because the Finnish government reports the number of confirmed cases at the HCD level. A more detailed explanation of how THL reports the cases is given next in Section 3.4.

3.3.1 The effective reproduction number

The effective reproduction number R_{eff} is computed in the same way as in Ref. [47],

$$R_{\text{eff}} = \beta\left(\frac{r_p(1-p_a)}{\lambda_2} + \frac{1-p_a}{\gamma} + \frac{r_a p_a}{\gamma}\right).$$

Most of the values for the parameters are fixed (see Table 3). However, the parameter β changes across HCDs and in time and is calibrated based on data, see Section 3.5.

3.4 Reported cases by THL

The $\beta_k(t)$ parameters in the compartmental metapopulation model presented in Section 3.1 are calibrated using the number of confirmed cases or incidence, reported by the Finnish Institute for Health and Welfare (in Finnish *Terveyden ja hyvinvoinnin laitos*) (THL). THL reports daily the incidence of coronavirus by Hospital Care District (HCD). The data are available on their website, and are free to download [79]. One or more municipalities form each HCD, and a code identifies it. Each HCD has a different population and area extent. Helsinki and Uusimaa HCD comprise a little less than a third of the population of Finland. The second biggest HCD population-wise is Pirkanmaa HCD, with almost 10% of the population of Finland. The population of each HCD is in Table 4.

The HCD with the most prevalence until the end of May 2020 was the Helsinki and Uusimaa HCD. By itself, it accounts for over two-thirds of the total prevalence in Finland during the period. Southwest Finland HCD, where the city of Turku is located, is in a second place. However, it only accounts for a little bit less than 5% of the total prevalence. Moreover, 11 out of the 21 HCDs have less than a 1% contribution to the total prevalence.

The reported incidence is not a reflection of the actual number of infectious people. Not all infectious people present symptoms and are asymptomatic. Most asymptomatic individuals will not get tested, and THL will not count them in the reports. Nonetheless, from the numbers, it can be seen how Helsinki and Uusimaa HCD greatly influence the national epidemic numbers. If Helsinki and Uusimaa HCD has a large incidence, Finland will most likely also have a large incidence.

Figure 8 shows the progression of the epidemic in Finland during the spring period of 2020. The cases started increasing at the beginning of March 2020. The day with the most reported incidence, when the epidemic reached its peak, was 6 April 2020. On that day, THL reported 210 confirmed cases. After 6 April, the number of reported cases started to decrease, reaching almost 0 by the end of May 2020.

HCD Name	Code	Population	Population (%)	Cases	Cases (%)
Ahvenanmaa	0	29,884	0.54	13	0.18
Southwest Finland	3	482,169	8.72	347	4.96
Satakunta	4	216,752	3.92	53	0.75
Kanta-Häme	5	170,925	3.09	111	1.58
Pirkanmaa	6	537,226	9.72	260	3.73
Päijat-Häme	7	210,057	3.80	109	1.56
Kyymenlaakso	8	164,456	2.97	44	0.63
South Karelia	9	127,757	2.31	23	0.32
South Savo	10	98,823	1.78	52	0.74
Itä-Savo	11	40,258	0.72	13	0.18
North Karelia	12	164,465	2.97	25	0.35
North Savo	13	244,236	4.42	138	1.97
Central Finland	14	252,716	4.57	134	1.91
South Ostrobothnia	15	193,207	3.49	46	0.65
Vaasa	16	169,495	3.06	58	0.83
Central Ostrobothnia	17	77,304	1.39	23	0.32
North Ostrobothnia	18	41,0112	7.42	134	1.91
Kainuu	19	72,306	1.30	65	0.93
Länsi-Pohja	20	60,295	1.09	145	2.07
Lappi	21	116,866	2.11	74	1.05
Helsinki and Uusimaa	25	1,685,983	30.5	5,116	73.26
Total	-	5,525,292	100	6,983	100

Table 4: List of HCDs in Finland, including their code, population at 2019, and cumulative number of reported cases or prevalence until 31 May 2020. The prevalence is taken from the daily numbers reported by THL [79]. The population numbers were obtained from Statistics Finland [78].

3.4.1 Hospitalizations

Another common indicator to evaluate the severeness of COVID-19 is the number of hospitalizations. Unfortunately, THL does not report this by HCD. THL reports the number of hospitalizations to a more aggregated geographical level called Hospital Catchment Area (in Finnish *erityisvastuualue*) (ERVA). An ERVA contains more than one HCD. Since there is no available way to assign the number of hospitalized people by ERVA, and relate it to the HCD level, the number of hospitalizations is not going to be used to calibrate the compartmental model.

Next, Section 3.5 presents the complete process for calibrating the model using the incidence and the mobility data.

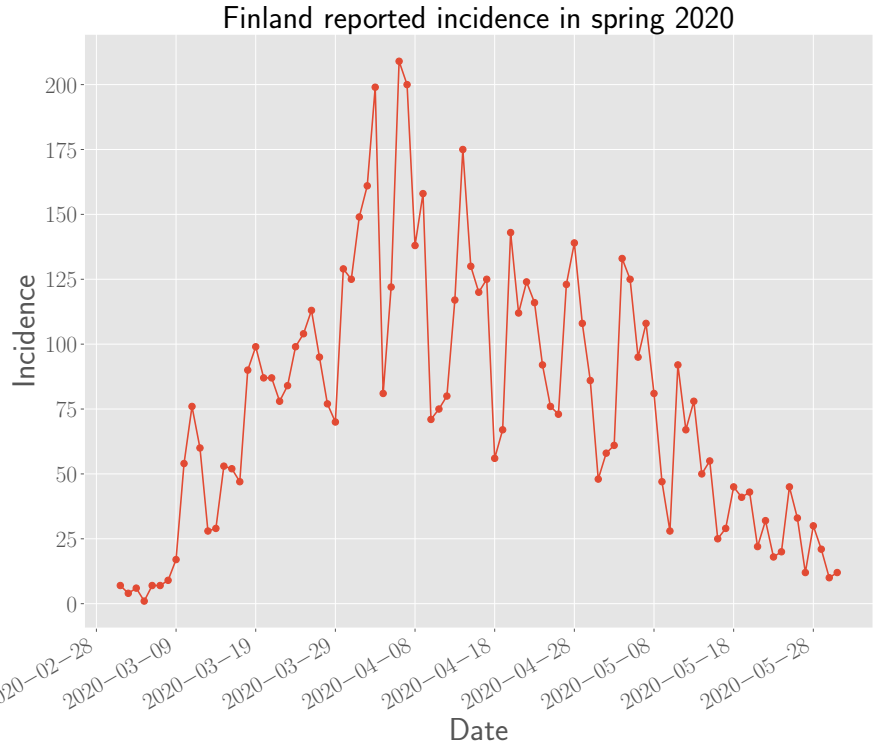

Figure 8: Incidence in Finland per day from 24 February 2020 to 31 May 2020. The peak number of reported cases was reached on 6 April 2020.

3.5 Calibration of the epidemic model

The parameters $\beta_k(t)$ from the model presented in Section 3.1 are inferred from the data. The parameter depends on HCD k and time t, reflecting the changing behavior of the disease across different HCDs and different times. The reported incidence by THL and the mobility data are used to calibrate the parameters $\beta_k(t)$. The specific inferring method used is Approximate Bayesian Computation (ABC), initially proposed in Ref. [17] and already explained in Chapter 2.

Chapter 2 presents the rejection algorithm for ABC. In the algorithm there are four essential components: the prior distribution for the parameters $p(\cdot)$, a simulator function $f(\cdot)$, the summary statistics $S(\cdot)$ and the distance measure $d(\cdot,\cdot)$. The basic idea is to extract a set of proposed parameters θ' from the prior distribution, make simulations of the data $\tilde{y}$ using $f(\theta')$ and accept the parameters θ' if the generated data $\tilde{y}$ is close enough to the observed data y according to the summary statistics and the distance function, $d(S(\tilde{y}), S(y))$. In the following sections, I will analyze the choice of the ABC elements for the coronavirus epidemic and the epidemic model.

3.5.1 Prior distributions

Priors are used to sample a set of proposed parameters θ' that will be later introduced into the simulator function $f(\cdot)$, as seen in Algorithm 1. For the case of the epidemic model presented earlier, the proposed parameters sampled θ' are going to be proposed

parameters $\beta_k(t)$ for the model. I used the exponential distribution as a prior distribution for the β parameters. The reason behind choosing an exponential distribution is because its support is $[0, \infty)$ and thus it can never adopt negative values. This same characteristic is also desired for the β parameter; a negative value of β would yield a negative R_{eff}, which is impossible. On the other hand, when $\beta = 0.22$, the effective reproductive number $R_{\text{eff}} = 1$. Therefore, to have no assumptions on the state of the epidemic, the mean for the prior exponential distribution is $\lambda^{-1} = 0.22$.

The parameter $\beta_k(t)$ depends on time t and HCD k. Therefore, during simulations, the algorithm first samples a national value $\beta(t)$, and then it samples the $\beta_k(t)$ values for each HCD conditionally on the national $\beta(t)$ obtained.

3.5.2 Simulating the data

The first step to simulate data is to sample some proposed $\beta_k(t)$ parameters from the prior distribution. The $\beta_k(t)$ parameters along with the mobility data (Section 3.2) and the rest of the parameters (Section 3.3) will be used to do realizations of the epidemic model over some time T to get $\tilde{y}$. If $\tilde{y}$ is similar enough to y, the set of proposed $\beta_k(t)$ parameters will be accepted.

Defining the simulated data. The true observed data y are the reported incidence by THL during time T. To simulate data $\tilde{y}$, I will use the epidemic model as the simulator function $f(\cdot)$. The incidence of individuals in the epidemic model's infectious compartment (I) will be taken as the simulated data $\tilde{y}$. Hence, I am assuming that all the individuals that are infected are also reported.

The number of required parameters. Note that the number of required $\beta_k(t)$ parameters depends on the time T of the simulations. For instance, if $T = 7$ days, and there is a different $\beta_k(t)$ used per day and HCD, then $7 \times 21 = 147$ parameters $\beta_k(t)$ parameters are needed. The number 21 comes from the number of HCDs. However, the value of $\beta_k(t)$ is not expected to vary that much daily. Even the reports done by health institutes tend to report a R_{eff} weekly instead of daily. Because of this, the same sampled $\beta_k(t)$ will be used for seven days. Adjusting the previous example when $T = 7$ days, only 21 parameters $\beta_k(t)$ are needed instead of 147.

Dividing the time T. Another detail is that to have a faster convergence of the ABC algorithm and more accurate $\beta_k(t)$ values, I do the simulations of the epidemic in steps of $p = 21$ days until reaching the complete time T rather than a single simulation for the complete duration of T. Algorithm 5 exhibits the complete flow of the process to obtain $\tilde{y}$.

3.5.3 Summaries and distance measure

The summary statistics and distance measure are the last components of the ABC algorithm. I used $K + T$ summary statistics and constructed a vector $\mathbf{s}$ of length

Algorithm 5: Simulating the epidemic in steps of maximum $p = 21$ days. Since the value of $\beta_k(t)$ is used for 7 days, and there are $K = 21$ HCDs, then at most, a total of $3 \times 21 = 63$ parameters $\beta_k(t)$ are sampled and calibrated at the same time using ABC (Algorithm 1).

```
p ← 21 ;
/* Variable to keep track of the simulated time                    */
sim ← 0 ;
T ← Total time;
while sim < T do
    /* Finding out how much time is missing to simulate            */
    if T − sim ≥ 21 then
        d ← 21;
    else
        d ← T − sim;
    Run ABC with the epidemic simulator for d days to get β_k(t);
    Save obtained β_k(t);
    sim ← sim + p;
```

$K + T$ with its result. The K statistics come from the sum over time T of the incidence of each HCD k. For a single HCD,

$$y_k = \sum_{t=0}^{T} y_k(t).$$

Similarly, the T statistics come from the sum of the incidence of all HCDs K at time t,

$$y(t) = \sum_{k=0}^{K} y_k(t).$$

Once the vectors for the observed reported cases $\mathbf{s}$ and the simulated reported cases $\tilde{\mathbf{s}}$ are obtained, they are compared using the Euclidean distance,

$$d(S(y), S(\tilde{y})) = d(s, \tilde{s}) = \sqrt{\sum (s_i - \tilde{s}_i)^2}.$$

3.5.4 Seeding the epidemic

The last aspect to consider when calibrating the parameters $\beta_k(t)$ is setting the initial infectious people. Without this, the disease will not spread. As a result, there is a need to set some initial number of infectious people per HCD k. This process is also sometimes referred to as seeding the epidemic.

I seeded the epidemic using the incidence reported by THL during the first three weeks of the time T. The period of three weeks, 21 days, was chosen to match the first step of $p = 21$ days when simulating the epidemic, as explained in the previous section.

The primary purpose of the Thesis is to analyze the mobility measures implemented by the government during the spring of 2020. Accordingly, I will calibrate the epidemic model from 2 March 2020 to 31 May 2020. Thus, the epidemic will be seeded using the reported incidence by THL during the first three weeks of the period, from 2 March to 22 March.

To match the day the cases were reported, I construct a matrix M with $K = 21$ rows and $D = 21$ columns. The columns represent the days in which the cases were reported. The entries of the seed matrix $m_{k,d}$ are the reported infections $i_{k,d}$ amplified by a factor ψ_k. The factor ψ_k is introduced to account for the undetected cases at the beginning of the epidemic at HCD k,

$$m_{k,d} = \psi_k i_{k,d},$$

The final value of $m_{k,d}$ represents the infectious individuals used as a seed in HCD k in day d. The simulator, in turn, will introduce the infectious seed individuals at the appropriate simulation day. The value of ψ_k is also going to be calibrated from data using ABC.

The complete reported prevalence for the three weeks is not used to seed the epidemic. Instead, the prevalence is divided into three weeks ω_i. An HCD k will only have as seed the reported cases of one of the three weeks. The incidence for the rest of the weeks is not going to be used. If HCD k has reported cases during the first week ω_1, those cases will be used to seed the epidemic. If k does not have reported cases at ω_1, then the cases from the second week ω_2 will be used. Still, some HCDs did not report cases during the first two weeks; for those HCDs, the cases for the third week, ω_3, are used. The used week-prevalence for seeding each HCD per week ω_i is in Table 5.

3.6 Calibrated model in Finland during spring of 2020

The model is calibrated over the period from 2 March 2020 to 31 May 2020. It was in this period when the epidemic reached Finland and had its first outbreak or wave. For simplicity, through the rest of this chapter, I will refer to this period as *spring 2020*.

The parameter $\beta_k(t)$ is the only parameter calibrated to data of *spring 2020*. The parameter $\beta_k(t)$ accounts for the transmission rate of an infectious individual. This parameter depends on time t and region k, i.e., the parameter's value varies across time and region. The value of $\beta_k(t)$ stays constant over seven days, one week, in a given region. Since *spring 2020* consists of 91 days and 21 HCDs, there are $(91/7) \times 21 = 273$ inferred parameters.

ABC is used to infer the parameters $\beta_k(t)$. ABC obtains 1000 samples for each $\beta_k(t)$. Each set of samples yields a realization of the epidemic model over *spring 2020*. To calibrate the model, the observable quantity of the model is the incidence in the I compartment. Thus the incidence in the I compartment will resemble the incidence in Finland at the end of the calibration. Hence there are 1000 realizations of the epidemic model that resemble the incidence of the epidemic in Finland in *spring 2020*.

HCD Name	ω_1	ω_2	ω_3
Ahvenanmaa	0	0	2
Southwest Finland	1	0	0
Satakunta	0	7	0
Kanta-Häme	3	0	0
Pirkanmaa	7	0	0
Päijat-Häme	0	1	0
Kyymenlaakso	0	1	0
South Karelia	1	0	0
South Savo	0	2	0
Itä-Savo	0	0	1
North Karelia	1	0	0
North Savo	3	0	0
Central Finland	1	0	0
South Ostrobothnia	0	1	0
Vaasa	0	1	0
Central Ostrobothnia	1	0	0
North Ostrobothnia	1	0	0
Kainuu	0	0	1
Länsi-Pohja	0	0	1
Lappi	0	1	0
Helsinki and Uusimaa	22	0	0

Table 5: Prevalence used for seeding the epidemic. The rows are the HCDs in Finland and the columns the weeks ω_i that were used for seeding. Within a row, only one column $\omega_i \neq 0$, i.e., only one week per HCD was used for seeding. The week ω_1 encompasses the period from 2 March 2020 to 8 March 2020, ω_2 from 9 March to 15 March, and ω_3 from 16 March to 22 March.

The average incidence over the 1000 realizations of the epidemic model will represent the overall incidence of the calibrated model. The 25 and 75 percentiles will delimit the 50% Confidence Interval (CI) and the 12.5 and 87.5 percentiles the 75% CI.

Figure 9 shows the incidence of the calibrated model in Finland and for the three HCDs with the most prevalence over *spring 2020*. The three HCDs with the most prevalence are Helsinki and Uusimaa, Southwest Finland, and Pirkanmaa. These three HCDs correspond with three major cities in Finland: Helsinki, Turku, and Tampere, respectively.

According to Figure 9, the calibrated model does resemble the reported incidence by THL. The calibration of the model is better for the Helsinki and Uusimaa HCD than for the other HCDs. It is expected that the Helsinki and Uusimaa HCD has a better fit since it is the HCD with the highest incidence and therefore has a more significant influence on the distance measure of the ABC algorithm.

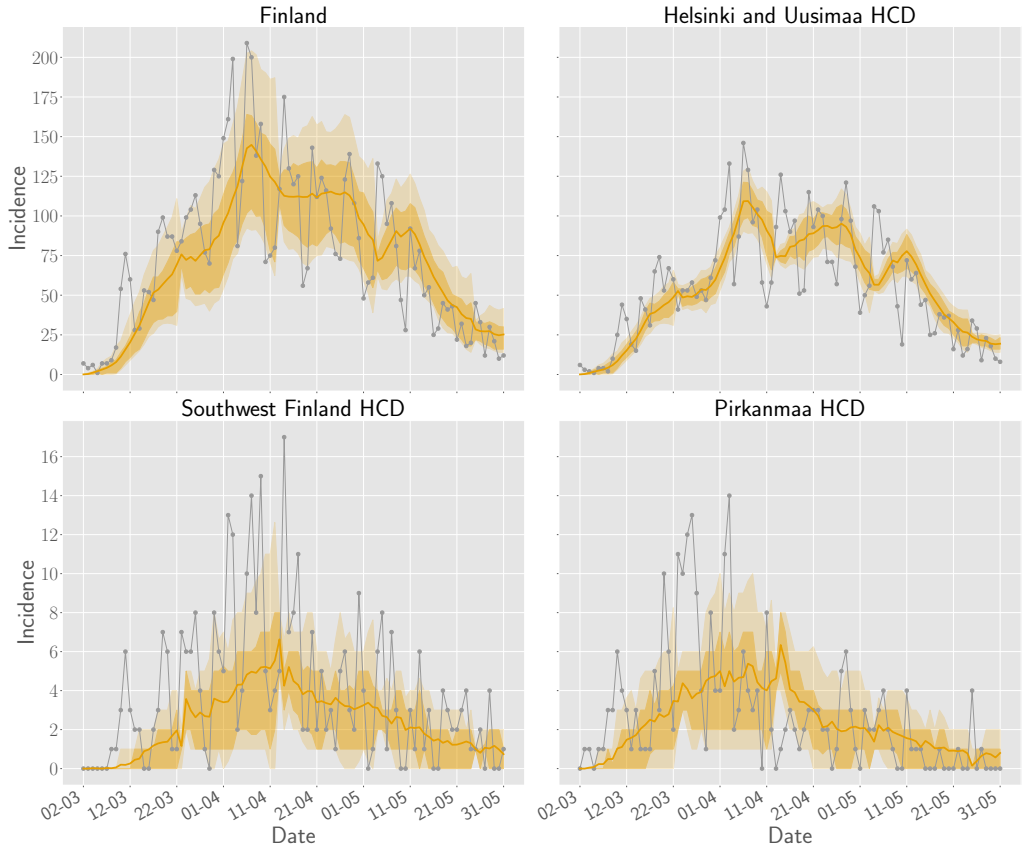

Figure 9: Calibrated metapopulation model from 2 March 2020 to 31 May 2020. The reported incidence by THL is in gray. The incidence of the model is in orange. The shadowed areas are the CIs of the model. The darker shadow is the 50% CI and the lighter the 75% CI.

3.6.1 R_{eff} during *spring 2020*

The effective reproduction number (R_{eff}) values can be retrieved using the samples of the $\beta_k(t)$ parameter. The Section 3.3.1 specifies the relation between R_{eff} and $\beta_k(t)$.

As before, Figure 10 displays the plotted values of the R_{eff} for Finland, Helsinki and Uusimaa, Southwest Finland, and Pirkanmaa HCDs. Helsinki and Uusimaa HCD has the highest R_{eff} value, $R_{\text{eff}} > 3$ at its highest point. This point was reached in the week of 9 March 2020. After this week, the government declared a state of emergency, and R_{eff} decreased had a slight increase in the week that started on 30 March. The week of 30 March is the week after the border of Uusimaa was closed. The slight increase of R_{eff} at this time suggests that, while the NPI of closing the border was to prevent infections at other HCDs, it promoted the contacts and infections inside Helsinki and Uusimaa.

In Finland, $R_{\text{eff}} < 2$ at the beginning of *spring 2020*. After the declaration of the state of emergency, and the implementation of NPIs, the R_{eff} value dropped near to unity. During the week of 30 March, R_{eff} had a slight increase. The reported incidence of that week explains the increase in R_{eff}. That week marked the peak of

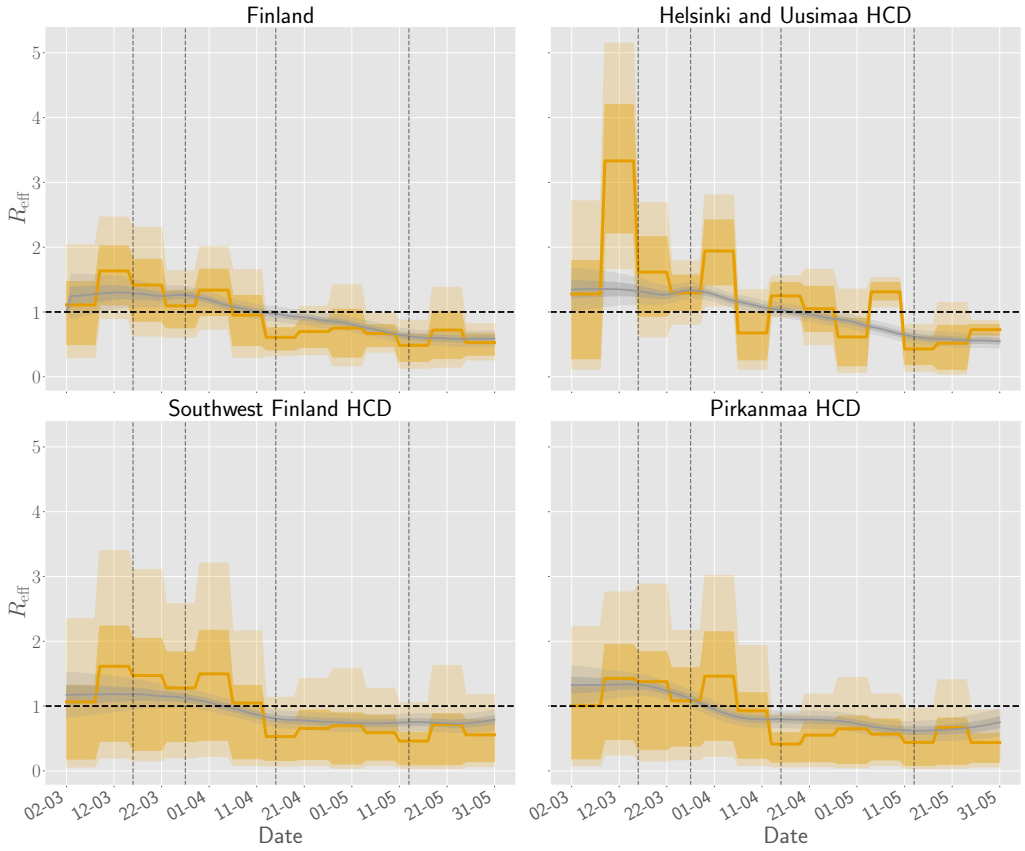

Figure 10: R_{eff} for the calibrated model from 2 March 2020 to 31 May 2020. The continuous orange line is the average value of R_{eff}, the darker and lighter shade are the 50% CI and 75% CI, respectively. In a discontinuous horizontal black line, $R_{\text{eff}} = 1$, the equilibrium of the epidemic, when it will not increase nor decrease. The discontinuous vertical lines mark dates of meaningful mobility changes. The first line marks when the state of emergency was declared in Finland on 16 March 2020. The second and third lines mark when Uusimaa border was closed, from 27 March to 15 April. The last black line is 13 May, when the gradual reopening of schools started along with the ease of other mobility restrictions. The gray line is the calculated R_{eff} in Ref. [61], and it is plotted for comparison.

the epidemic at a national level and in most HCDs. By the following week, the week of 6 April, R_{eff} dropped below 1, and it stayed below 1 until the end of *spring 2020*. This indicates that the NPIs at the beginning of spring worked. Based on this stable state of $R_{\text{eff}} < 1$, the government decided to ease the restrictions on 13 May.

The R_{eff} values that result from the calibrated model are compared with the results obtained in Ref. [61]. THL did not start reporting the weekly R_{eff} until 15 May 2020 [80]. However, Ref. [61] is a site that, since 3 March 2020 calculates the R_{eff} values for Finland and at the ERVA level using an $SEIR$ model. The site uses the reported incidence by THL to estimate the state of the $SEIR$ model using an unscented Rauch-Tung-Striebel smoother. Then, a value of R_{eff} is obtained from the estimated system state. The results obtained by Ref. [61] mostly agree with the ones

obtained by THL for the other dates; thus they are a good baseline. The results of Ref. [61] for *spring 2020* are very similar to the results obtained in this Thesis. Figure 13 shows that the R_{eff} values are similar except for the Helsinki and Uusimaa HCD at the beginning of spring 2020. In this case, the R_{eff} values of the calibrated model used in this Thesis are more pessimistic than those in Ref. [61].

Finally, the following section presents a final result with a scenario that tests the model and methods of the Thesis.

3.7 The epidemic in Finland beyond *spring 2020*.

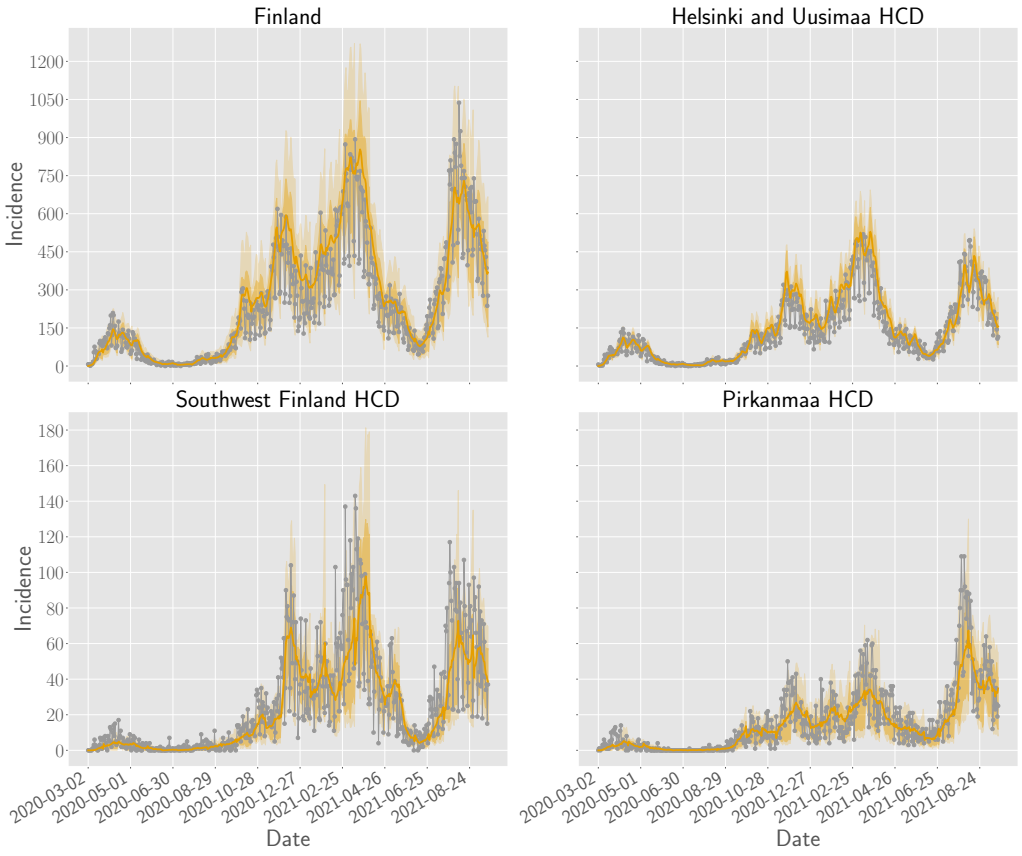

Figure 11: Calibrated metapopulation model from 2 March 2020 to 19 September 2021. The reported incidence by THL is in gray. The incidence of the model is in orange. The shadowed areas are the CIs of the model. The darker shadow is the 50% CI and the lighter the 75% CI.

To test the model's robustness and ABC as inference method, I calibrated the model to a longer time horizon. The model is calibrated from 2 March 2020 to 19 September 2021 in this final result. The objective is to test if the methods presented in the Thesis for calibrating the model could be applicable to another scenario.

Figure 11 shows the incidence for the same four regions as the previous results of this chapter but for a longer time horizon. The incidence by the calibrated model

does follow the reported incidence by THL. The result implies that the calibration and inference methods are robust.

Throughout this chapter, I have presented the model, introduced its different parameters and characteristics, its calibration to data, and the resulting R_{eff} given the calibration. In the coming Chapter 4, I will discuss how to adjust the calibrated model through mobility data to analyze different counterfactual scenarios. The objective of the counterfactual scenarios is to know how the NPIs implemented in Finland ultimately affected the outcome of the coronavirus epidemic in Finland.

4 Counterfactual scenarios

The main objective of the Thesis is to analyze the effect of NPIs through counterfactual scenarios on the coronavirus epidemic situation in Finland during spring 2020. A compartmental metapopulation model, calibrated with the incidence in each region in Finland, describes the epidemic situation (see Chapter 3). Each counterfactual scenario will affect the mobility of the calibrated model; this alteration will produce a different incidence than the one of the calibrated model. The new incidence is then compared to the incidence of the calibrated model to analyze and contrast the role of mobility and NPIs.

Most of the NPIs affected the daily mobility patterns. Mobility had a crucial impact in driving down the incidence and R_{eff}. Without a vaccine and effective medication against coronavirus, it was one of the few things that the government could control. If people do not meet, there is also no possibility of getting infected.

Counterfactual scenarios allow gaining the previously mentioned understanding of the impact that NPIs had on the outcome of the epidemic. With counterfactual scenarios, the possibility of modeling how mobility would have been if a given NPI was not taken and then analyzing the impact of that scenario in the outcome of the epidemic is available. The opposite is also possible, posing a scenario in which the control imposed by NPIs restricts more mobility than it did. The comparison of different scenarios can shed light on which NPIs were most effective in controlling the epidemic.

4.1 Counterfactual scenarios based on mobility data

The counterfactual scenarios that I will present will affect mobility since it is one of the most representative aspects of everyday life affected by NPIs. There are two ways in which data can affect mobility. The first one is to affect the ingoing and outgoing people of a specific HCD, changing the cross-border mobility between HCDs, explained next in detail in Section 4.1.1. The second one is to affect the internal mobility of the HCDs, impacting the way people interact with each other. To that end, a parameter (α) is introduced into the stochastic equations presented in Section 3.1; Section 4.1.2 explains the new parameter in detail. Figure 12 displays a small diagram illustrating the factors affecting the mobility in a HCD k.

4.1.1 Cross-border traffic

For the case of affecting mobility between HCDs it is enough to change the number of trips of the O-D matrix for a given number of days and hours. Thus the number of trips between HCDs is going to be altered in two ways.

Having no trips between HCDs. Not having any trips between HCDs accounts for a very severe scenario in which the borders of all the HCDs are closed so nobody can go outside of their HCD of residence. People would still have contacts and

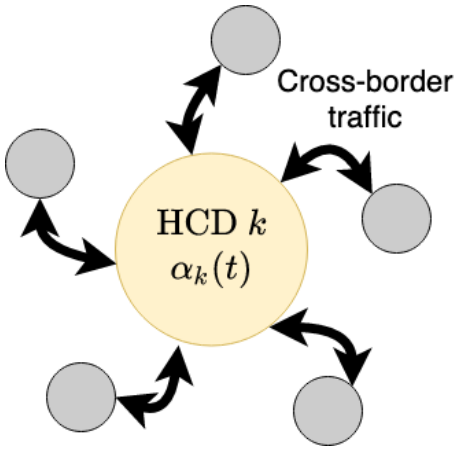

Figure 12: Schematic of the factors altering mobility in HCD k. Changing the number of trips ingoing and outgoing (cross-border traffic) will influence the interaction with other regions. On the other hand, the parameter $\alpha_k(t)$ will scale up or down the parameter $\beta_k(t)$ to account for how people interact inside a given region.

therefore spread the disease inside the HCD. Therefore, all entries in the O-D matrix will simply be set to zero.

Have an equivalent number of trips as in 2019. The second scenario accounts for not having any restrictions. The year 2019 is a time without the presence of coronavirus and hence of normal mobility. The objective is to see how the disease would have spread under a normal cross-border flow. The O-D matrix will be constructed using $\bar{w}^{(2019)}$, defined in Section 3.2. Thus, the trips tr for a day d will be replaced with the trips of $\bar{w}^{(2019)}$ of weekday x of day d.

4.1.2 Inner mobility and the α parameter

The value of the parameter $\beta_k(t)$ is mainly responsible for determining the transmissibility inside an HCD k. This parameter encodes how many contacts an average person makes and the transmission probability of the disease during these contacts. The parameters $\beta_k(t)$ of each HCD are initially calibrated to emulate the situation of spring 2020 in Finland. Then, I introduce an $\alpha_k(t)$ parameter that will adjust the calibrated contacts of $\beta_k(t)$ depending on the counterfactual scenario modeled.

The data and situation allow three different counterfactual scenarios to adjust the contacts of the calibrated model to spring 2020:

1. Reduce the number of contacts.
2. Reproduce the number of contacts made in 2019.
3. Remove the contacts of people who stayed in their HCD of residence that, without coronavirus, would have traveled out.

Another plausible scenario is one in which, instead of removing completely the contacts of people who stayed in their HCD of residence, they make as many contacts in their HCD as they would have made if they had traveled out. Unfortunately, the last described scenario cannot be modeled due to the nature of the data. Currently, there is no way to determine from the data the average time a person from HCD k spends when he or she visits HCD l. Thus, it is also impossible to know the average contacts an individual makes in HCD l and transfer those contacts to their HCD of residence k. For these reasons, this scenario is not considered here.

As mentioned, the parameter $\alpha_k(t)$ will affect $\beta_k(t)$ in Equations (2). It does it in the following way,

$$\begin{aligned} X_k^{(1)}(t) &\sim \text{Binom}\left(S_k(t), \alpha_k(t)\beta_k(t)\Delta t\frac{I_k(t)}{N_k}\right);\\ X_k^{(2)}(t) &\sim \text{Binom}\left(S_k(t), r_a\alpha_k(t)\beta_k(t)\Delta t\frac{A_k(t)}{N_k}\right);\\ X_k^{(3)}(t) &\sim \text{Binom}\left(S_k(t), r_p\alpha_k(t)\beta_k(t)\Delta t\frac{P_k(t)}{N_k}\right). \end{aligned} \tag{4}$$

Since $\beta_k(t)$ is not present in $X^{(4)}(t)$ through $X^{(8)}(t)$ of Equations (2), these expressions will remain unchanged.

From Equations (4), it is observed that $\alpha_k(t) = 1$ would be equivalent to leaving the inner mobility unchanged and therefore not altering the calibrated $\beta_k(t)$ values. Thus, setting $\alpha_k(t) > 1$ would account for more contacts inside the HCD. Equivalently, $\alpha_k(t) < 1$ would be set for a scenario with fewer contacts.

In the coming paragraphs, I will explain how the values for $\alpha_k(t)$ were set according to the three counterfactual scenarios mentioned before.

Reducing the number of contacts. The first counterfactual scenario will set $\alpha_k(t)$ to reduce the number of contacts made in 2020. This scenario will allow modeling a situation where the population reduces its contacts by staying more at home or making fewer contacts due to even tighter restriction measures inside each HCD or a natural fear of the disease.

To model this scenario, I set $\alpha_k(t)$ to a **constant**,

$$\alpha_k^{(\text{const})}(t) = 0.8.$$

This value implies a 20% reduction in contacts with respect to the calibrated scenario in all HCDs and at all times t.

Reproduce the number of contacts made in 2019. The second counterfactual scenario will model $\alpha_k(t)$ to affect $\beta_k(t)$ such that it reproduces the number of contacts in 2019. The value of $\alpha_k(t)$ will encode the information of the difference in the contact patterns between 2019 and 2020. The purpose is to pose a situation in which the population did not reduce their contacts during 2020. Instead, it kept the contact patterns of 2019.

To this end, I will consider the trips that started in HCD k and ended in HCD k, the inner trips of HCD k during 2019 and 2020, to compute the values of $\alpha_k(t)$. These inner trips also include the within-municipality trips present in the data, aggregated to the HCD level. The difference in the inner trips between the years will be interpreted as the difference in the contact patterns. Thus, more inner trips during 2019 than 2020 will yield more contacts inside the HCD and vice versa.

To compute $\alpha_k(t)$, I will use the average mobility per weekday during 2019 $\bar{w}^{(2019)}$ presented in Section 3.2 and the trips tr during 2020 of the mobility data,

$$\alpha_k^{(2019)}(t) = \frac{\bar{w}_{k,k}^{(x,h)}}{tr_{k,k}^{(d,h)}},$$

in which $\bar{w} = \bar{w}^{(2019)}$, x is the weekday of day d, h denotes the hour bin of the trips, and $tr^{k,k}$ are the trips that started and ended in HCD k on 2020. It is worth mentioning that t matches h. Every time step t in simulation is equivalent to 6 hours; the hour bins h also account for 6 hours of the day.

Removing the contacts of people that stayed in their HCD of residence. The final counterfactual scenario will model the hypothetical but very possible situation in which the people who did not go out of their HCD k during 2020 but would have left without coronavirus do not participate in the contacts of the region to spread the disease. The assumption is that people forced to stay inside their HCD of residence stayed in lockdown and did not participate in contacts. Figure 13 confers a representation of the scenario.

There is an amount of travelers $\sum_{l\neq k}^{K} tr_{k,l}^{(2020)}$ that travel out HCD k during 2020. However, the amount of travelers $\sum_{l\neq k}^{K} tr_{k,l}^{(2019)}$ that travel out of the same HCD in 2019 is greater than in 2020, i.e.,$\sum_{l\neq k}^{K} tr_{k,l}^{(2019)} > \sum_{l\neq k}^{K} tr_{k,l}^{(2020)}$. Under normal conditions, without the coronavirus epidemic, one could expect $\sum_{l\neq k}^{K} tr_{k,l}^{(2019)} \approx \sum_{l\neq k}^{K} tr_{k,l}^{(2020)}$. NPIs and fear of the disease are responsible for this difference in the mobility between HCDs. One can theorize why the travelers were not able to make the usual trips; probably they were ordered to work from home or the avoided trips were not essential, or some other unknown reason.

Be that as it may, since the population of HCD k stays more or less constant between years, the population present in HCD k during 2020 is greater than the population present in 2019, $N_k - \sum_{l\neq k}^{K} tr_{k,l}^{(2019)} < N_k - \sum_{l\neq k}^{K} tr_{k,l}^{(2020)}$.

On the other hand, the model is calibrated assuming all the people present in HCD k during 2020 participated in contacts. Hence, the people who did not travel to other HCDs l are present and making contacts during calibration. To model the situation in which the people who stayed in their HCD k of residence did not participate in contacts, I assume that the population participating in contacts in HCD k during 2020 would be around the same as those who stayed at HCD k during

2019. The modeling of the assumption leads to the following expression,

$$\alpha_k^{(\text{lockdown})}(t) = \frac{\text{2019 population in } k \text{ at time } t}{\text{2020 population in } k \text{ at time } t};$$

$$\alpha_k^{(\text{lockdown})}(t) = \frac{N_k - \sum_{l \neq k}^{K} \bar{w}_{k,l}^{(x,h)}}{N_k - \sum_{l \neq k}^{K} tr_{k,l}^{(d,h)}},$$

in which N_k is the total population at HCD k, $\bar{w} = \bar{w}^{(2019)}$ the average number of trips in 2019 per weekday, x is the weekday of day d, h denotes the hour bin of the trips, and $tr_{k,l}$ are the trips starting at HCD k and ending in l in 2020. Figure 13 shows a representation of the scenario.

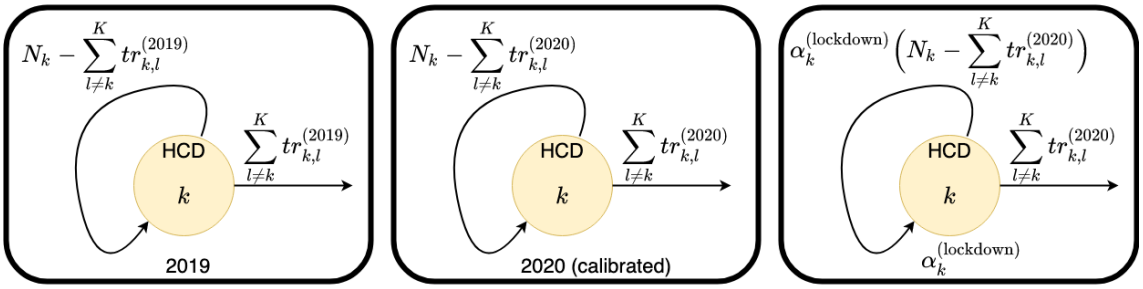

Figure 13: Graphical representation of the way $\alpha^{(\text{lockdown})}$ is modeled. Each square represents one scenario of the people that travel out from HCD k (arrow pointing to the right) and the people that stay inside the HCD and participate in contacts (recursive arrow). The leftmost square represents a scenario in which mobility in non-pandemic times is maintained, like in 2019. The central square is the scenario that happened in 2020 and to which the model is calibrated. The rightmost square is the situation modeled for $\alpha^{(\text{lockdown})}$. The outgoing trips are kept, but the people who participate in contacts inside the HCD are the same as in 2019.

Comparing the parameters α_k. I computed $\alpha_k(t)$ as described in the previous paragraphs. I also computed the values per day to get an idea of the values that $\alpha_k(t)$ will adopt. The values are computed by first aggregating by day the number of trips and substituting the values in the above-presented expressions, i.e., substituting $tr_{k,l}^{(d,h)}$ by $\sum_h tr_{k,l}^{(d,h)}$ and $\bar{w}_{k,l}^{(d,h)}$) by $\sum_h \bar{w}_{k,l}^{(d,h)}$). Figure 14 presents the daily numerical values of $\alpha_k(t)$.

As seen in Figure 14, $\alpha^{(2019)}$ will scale up β, accounting for a greater number of contacts in 2019 as compared to 2020. Also, we can see that, according to $\alpha^{(\text{lockdown})}$, on average, 10% of the people did not participate in contacts during 2020 since they were in lockdown. Finally, the scenario that will reduce the contacts the most is $\alpha^{(\text{const})}$.

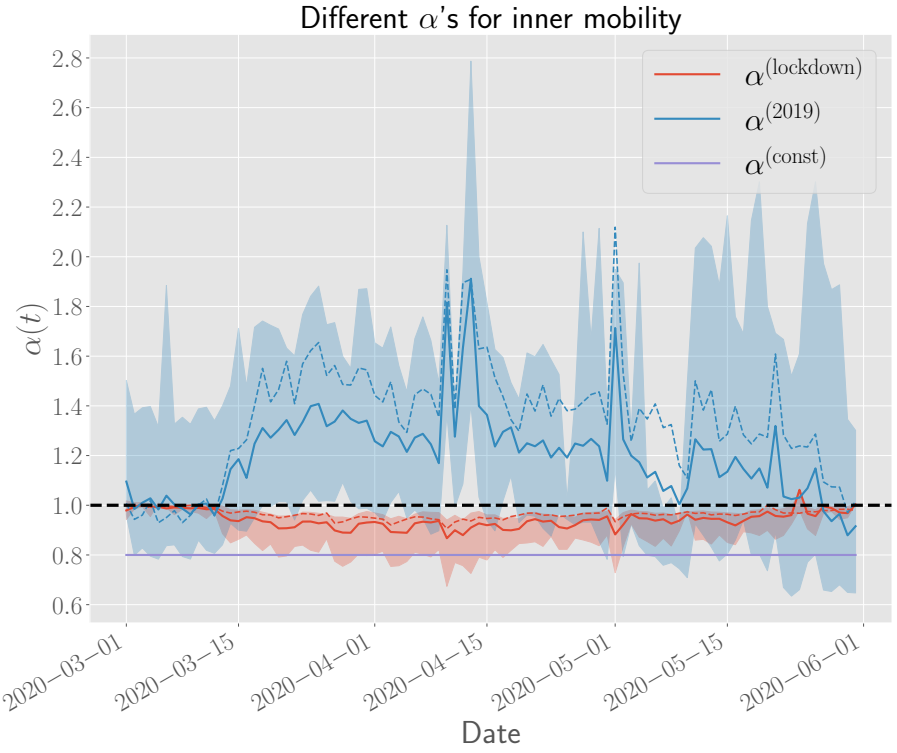

Figure 14: Numerical values of α per day. The a dashed black line denotes $\alpha = 1$, equivalent to not affecting the parameter β and leaving the model as calibrated. In a continuous line, for each color, are the average values of all the HCDs. Thus, it can be interpreted as a national numerical value of α. The dashed colored lines are the α values for the Helsinki and Uusimaa HCD under the different configurations of α. The shaded area is delimited by the HCDs that have, on average, the largest and smallest α. In the case of $\alpha^{(\text{lockdown})}$, the largest value corresponds to the North Karelia HCD and the smallest to the Kanta-Häme HCD. For $\alpha^{(2019)}$, the Vaasa and Itä-Savo HCD are the largest and smallest values, respectively.

4.2 Altering the calibrated model through counterfactual scenarios.

All possible combinations of $\alpha_k(t)$ and cross-border traffic were tried out to see which one had the most significant influence on the epidemic, inner contacts or cross-border traffic. Table 6 summarizes the combinations.

In total, there are eleven counterfactual scenarios. Figure 15 presents the results for the counterfactual scenarios in which $\alpha_k(t)$ is not $\alpha^{(2019)}$, and Figure 16 presents the results for $\alpha^{(2019)}$.

Figure 15 shows that, for the case of Finland, the scenarios that use $\alpha^{(\text{const})}$ are the ones with the lowest incidence. This case is equivalent to reducing 20% the number of contacts inside HCDs. Stricter lockdown measures are an example of how to achieve this reduction in contacts. For these scenarios with $\alpha^{(\text{const})}$, cross-border traffic does not significantly impact the incidence at a national level. The incidence

$\alpha_k(t)$	Cross-border traffic
1	2020
1	2019
1	0
const	2020
const	2019
const	0
2019	2020
2019	2019
2019	0
lockdown	2020
lockdown	2019
lockdown	0

Table 6: Summary of the available counterfactual scenarios of all possible combinations of $\alpha_k(t)$ and cross-border traffic. The top row is equivalent to not affecting mobility and leaving the model and parameters as calibrated. The subsequent rows show the counterfactual scenarios affecting mobility. The first column shows the way contacts inside HCDs can be affected by $\alpha_k(t)$, and the second column shows the cross-border traffic between HCDs.

is around the same despite the cross-border traffic.

In Finland, the scenario that would increase incidence the most is when cross-border traffic is zero, with no traffic between HCDs, and $\alpha_k(t) = 1$. Limiting the traffic to stop the spread of the disease can contain it in certain HCDs. However, the measure will promote the contacts inside HCDs in which the disease is already present, increasing the infections.

Notably, at a national level, all the scenarios with $\alpha^{(\text{lockdown})}$, in which the contacts of people who did not travel due to the pandemic were removed, report less incidence than those using $\alpha_k(t) = 1$, the calibrated model. In these scenarios with $\alpha^{(\text{lockdown})}$, the scenario with the most incidence is with the least cross-border traffic. The one with the fewest incidence is when cross-border mobility is at its highest.

The results for the Helsinki and Uusimaa HCD are very similar to the ones of Finland. This similarity confirms that the situation at a national level depends significantly on what happens in this HCD. Nonetheless, in the Helsinki and Uusimaa HCD, there is a small but noticeable difference in the incidence when $\alpha^{(\text{const})}$. When $\alpha^{(\text{const})}$, the worst scenario is when the cross-border traffic is equal to zero and the best when it is equivalent to the traffic of 2019. This difference is also explained by the fact that, without cross-border traffic, there are more contacts inside the HCD. With cross-border traffic similar to the one of 2019, contacts inside the region reduce with the disadvantage that the disease may escape the HCD and cause infections in other HCDs.

For the Southwest Finland HCD, the results are different when comparing them to Finland and the Helsinki and Uusimaa HCD. For this HCD, the lowest incidence

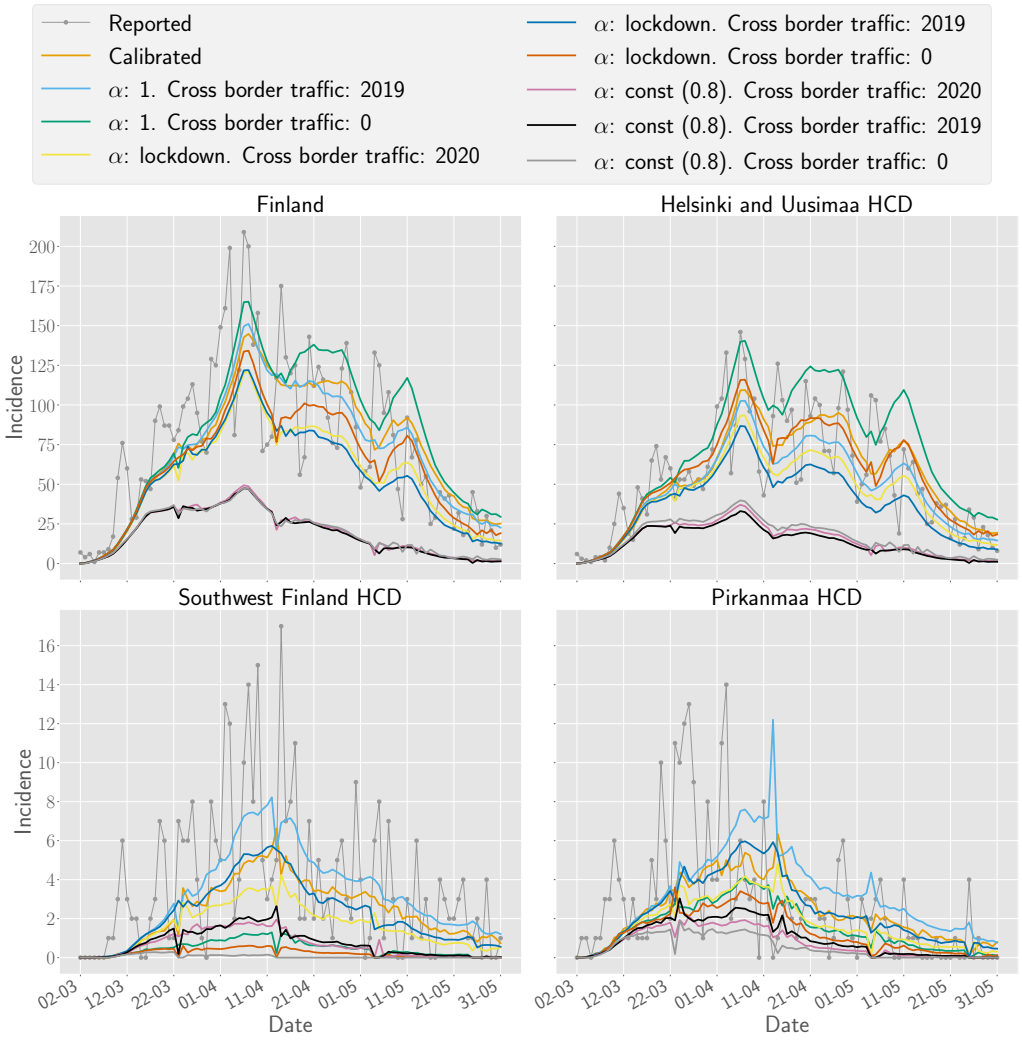

Figure 15: All counterfactual scenarios from 2 March 2020 to 31 May 2020, except for scenarios that use inner the HCD mobility parameter $\alpha^{(2019)}$. The orange line is the incidence of the calibrated model. The rest of the lines represent the incidence with each counterfactual scenario using $\alpha^{(\text{const})}$ and $\alpha^{(\text{lockdown})}$ in combination with all possible scenarios of cross-border traffic. The dash-dotted line represents the reported incidence by THL.

is when cross-border traffic is zero, no matter what values of $\alpha_k(t)$ is used. The result suggests that visitors caused most of the infections in the HCD during *spring 2020*. If there is a prohibition on traffic entrance to the HCD, there is an evident decline in the incidence. Following this logic, as expected, the highest incidence is seen when the cross-border traffic is similar to 2019. Since there were no mobility restrictions, the cross-border traffic of 2019 is greater than the one of 2020. Hence more visitors will arrive at Southwest Finland HCD and spread the disease.

The Pirkanmaa HCD follows a pattern similar to that of the Southwest Finland HCD. The scenario with the highest incidence is when the cross-border traffic is at its highest in 2019. In general, when comparing cross-border traffic with the

same $\alpha_k(t)$ for these two HCDs, the scenario with the most incidence is when the cross-border traffic is at its highest.

In summary, removing cross-border mobility causes the highest incidence in Finland and in the Helsinki and Uusimaa HCD, but it also causes the lowest incidence in the Southwest Finland and Pirkanmaa HCDs. Nevertheless, the opposite happens when cross-border mobility is highest in 2019; Finland and the Uusimaa HCD have the lowest incidence but the highest presents in the other HCDs.

Finally, the last counterfactual scenarios presented are those in which $\alpha^{(2019)}$ alters the contacts (Figure 16).

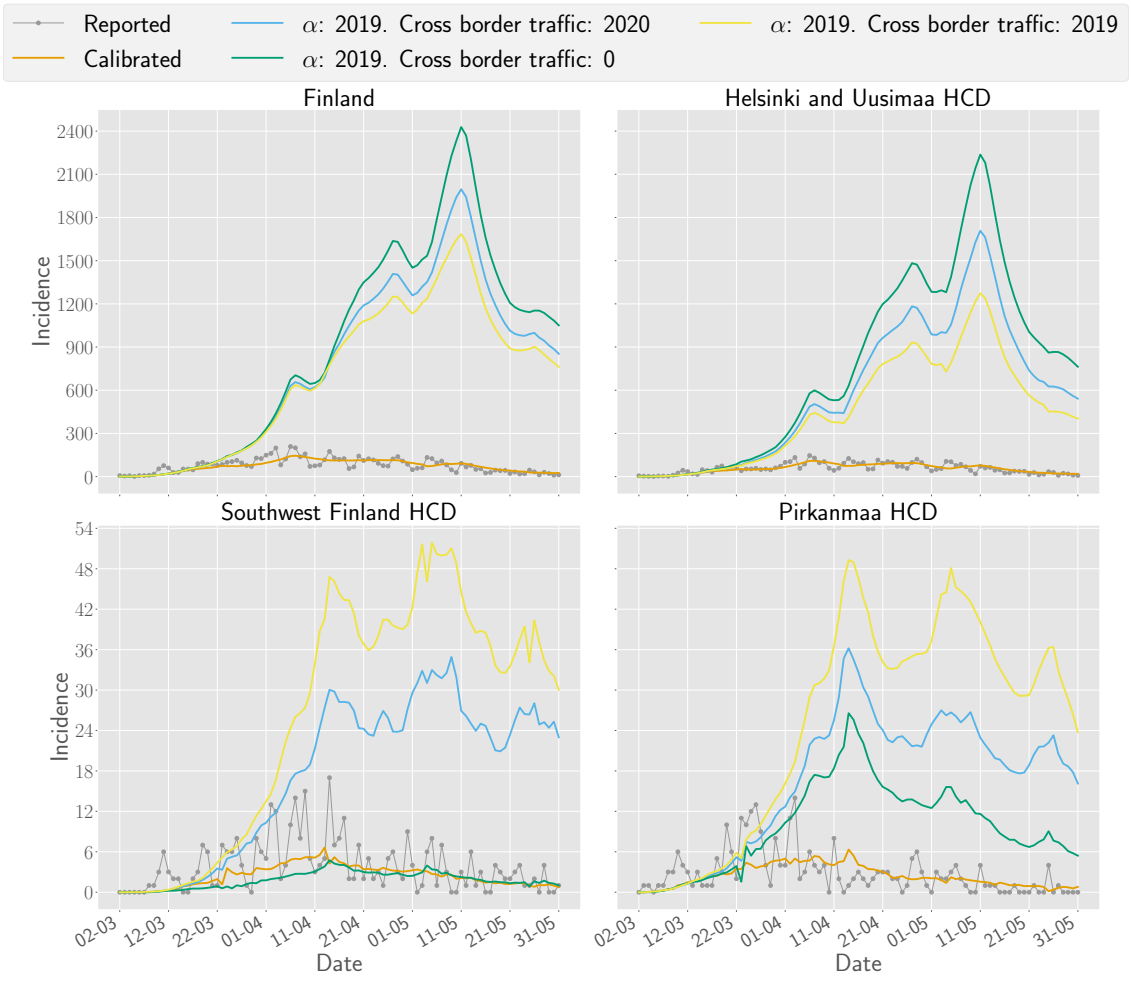

Figure 16: Counterfactual scenarios with inner HCD mobility parameter $\alpha^{(2019)}$ from 2 March 2020 to 31 May 2020. The orange line is the incidence of the calibrated model. The rest of the lines represent the incidence with each counterfactual scenario using $\alpha^{(2019)}$ in combination with all possible scenarios of cross-border traffic. The dash-dotted line represents the reported incidence by THL.

In Figure 16, there is a substantial increment in the incidence using $\alpha^{(2019)}$ with respect to the calibrated model results. At a national level, the incidence with $\alpha^{(2019)}$ is over ten times higher than that with the calibrated model. This increment of the

incidence happens with the three types of cross-border traffic. Suggesting that, no matter what the implemented control on the cross-border traffic, the incidence would have been outstandingly higher if NPIs had not taken place inside the HCDs.

However, although the incidence is higher with $\alpha^{(2019)}$, differences depend on the cross-border traffic used. The highest incidence occurs without cross-border traffic, followed by cross-border traffic of 2020 and finally traffic from 2019. The phenomenon is explained similarly as before; with more border traffic, more infectious people get out of the HCDs, but there are fewer infections inside and vice versa. Therefore, eliminating the cross-border traffic would increase the incidence in the Helsinki and Uusimaa HCD. Thus, the situation at a national level would be similar since the national situation is highly dependent on the Helsinki and Uusimaa HCD.

In Southwest Finland, without cross-border traffic the incidence is similar to the calibrated model. The result implies that even without implementing NPIs inside Southwest Finland, closing the border would have been enough to keep the incidence similar to the calibrated scenario. This result agrees with the earlier results that suggested that visitors caused most of the infections in the Southwest Finland HCD. Oppositely, incidence would be at its highest when the cross-border traffic also is, when the cross-border traffic is similar to 2019.

The Pirkamaa HCD follows a similar pattern; the incidence is highest when cross-border traffic is also at its highest. This confirms that, outside the Helsinki and Uusimaa HCD, the incidence would increase with more cross-border traffic. However, with less cross-border traffic, the incidence would increase in Helsinki and Uusimaa but decline in every other HCD. This phenomenon is valid for all values of $\alpha_k(t)$ studied here.

4.2.1 Effective reproduction numbers

Cross-border mobility does not directly influence the value of R_{eff}. However, altering $\beta_k(t)$, with $\alpha^{(\text{const})}$, $\alpha^{(\text{lockdown})}$ or $\alpha^{(2019)}$, will do so. R_{eff} is important because it communicates directly the state of the epidemic; it has been used by decision-makers through the epidemic. Therefore it is sensible to look at how R_{eff} changes with the different values of $\alpha_k(t)$ (see Figure 17).

The case in which $\alpha^{(\text{const})}$ alters the contact inside the HCDs confirms the expected drop in the values of R_{eff}. In fact, with this scenario, R_{eff} drops below unity at the national level just after the government declared a state of emergency on 16 March. Furthermore, after 16 March, R_{eff} stays below unity for most of *spring 2020*.

The R_{eff} in the Southwest Finland and Pirkanmaa HCDs with $\alpha^{(\text{const})}$ is similar to Finland's. However, it is not similar to the Helsinki and Uusimaa HCD. Although the value of R_{eff} in the Helsinki and Uusimaa HCD with $\alpha^{(\text{const})}$ is lower with respect to the value of R_{eff} of the calibrated model, it is still over unity for most of the beginning of *spring 2020*. It only drops below one after the week of 6 April. The result hints that an even larger reduction in contacts, well below 20%, would have been necessary to drive R_{eff} below unity and prevent more infections in the Helsinki and Uusimaa HCD.

The value of R_{eff} with $\alpha^{(\text{lockdown})}$ almost overlaps with the R_{eff} of the calibrated

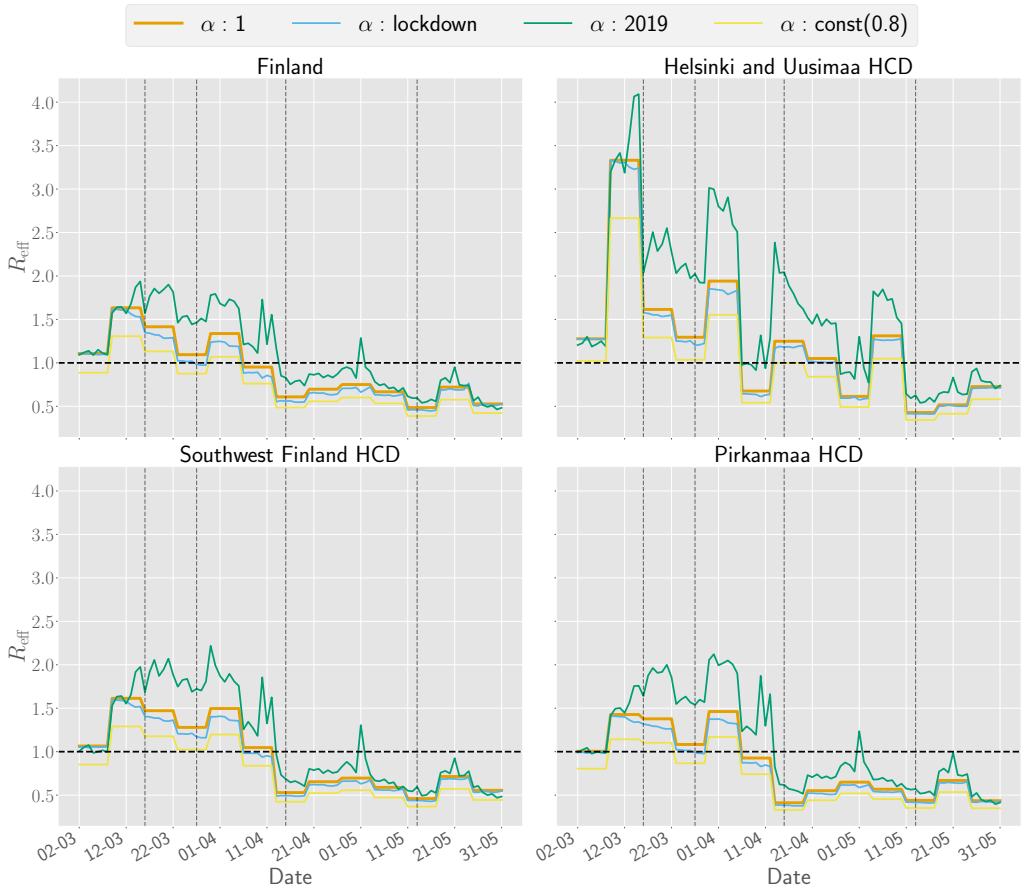

Figure 17: Comparison between R_{eff} values using $\alpha^{(\text{const})}$, $\alpha^{(\text{lockdown})}$, $\alpha^{(2019)}$, and the calibrated model from 2 March 2020 to 31 May 2020. The discontinuous horizontal black line indicates $R_{\text{eff}} = 1$, the equilibrium of the epidemic. The discontinuous vertical lines mark dates of meaningful mobility changes. The first line marks when the state of emergency was declared in Finland on 16 March 2020. The second and third lines mark when Uusimaa border was closed, from 27 March to 15 April. The last black line is 13 May, when the gradual reopening of schools started along with the ease of other mobility restrictions.

model. The overlap suggests that not enough people stayed in lockdown to change the value of R_{eff} in a significant way. However, enough people stayed at home to prevent some infections (Figure 15).

Contrarily to $\alpha^{(\text{const})}$, $\alpha^{(2019)}$ increases the value of R_{eff}. The superior number of contacts as compared to those in 2019 than in 2020 justifies the increment. Without NPIs that reduce contacts in Finland, $R_{\text{eff}} > 1.5$ for most of the beginning of *spring 2020*. It would drop below unity until 15 April, but the R_{eff} of $\alpha^{(2019)}$ would still be closer to it than the values of R_{eff} of the calibrated model. By the end of *spring 2020*, the values of R_{eff} with $\alpha^{(2019)}$ and of the calibrated model are very similar. This suggest that, by the end of *spring 2020*, the population had already regained the patterns and habits of 2019.

As with other results, the observed behavior of R_{eff} in Finland looks similar to

the behavior in Southwest Finland and Pirkanmaa with $\alpha^{(2019)}$. Again the exception is the Helsinki and Uusimaa HCD. As shown in Figure 17, the NPIs that reduced mobility in the Helsinki and Uusimaa HCD were crucial to reducing R_{eff} during 2020. With contact patterns similar to 2019, R_{eff} is well above unity for the majority of *spring 2020*. It has its peak the week of 9 March, when at the last two days of the week, $R_{\text{eff}} > 4$. It is until the last three weeks of *spring 2020*, that the values of R_{eff} of 2019 are similar to the values of R_{eff} of 2020.

4.2.2 Deaths with the counterfactual scenarios

A final way to observe the differences between counterfactual scenarios is by estimating the number of deaths under each of them. For this purpose, the deaths per reported case are calculated; they are obtained by dividing the reported number of deaths by the prevalence of the calibrated model over *spring 2020*. Finally, the obtained number of deaths per reported case is used to multiply the prevalence of each counterfactual scenario to obtain the number of deaths for each scenario.

$\alpha_k(t)$	0	2020	2019
2019	**3426.61**	2925.57	2606.36
1	36.52	0	-7.79
lockdown	-42.89	-73.08	-78.85
const (0.8)	-221.22	-221.53	**-224.25**

Table 7: Difference between the estimated deaths in each counterfactual scenario with respect to the estimated deaths of the calibrated model. Each table entry represents one counterfactual scenario, a combination of one form of $\alpha_k(t)$ with one type of cross-border traffic. The columns are organized from the scenario with less cross-border traffic (0) to most (2019). Similarly, the rows are organized from $\alpha^{(2019)}$ that will upscale the most $\beta_k(t)$ (top row), to $\alpha^{(\text{const})}$ that will downscale it the most (bottom row). In bold font are the scenarios that produce the larges and the smallest number of deaths.

Table 7 exhibits that the counterfactual scenario with the smallest number of deaths is the combination of $\alpha^{(\text{const})}$ with cross-border traffic from 2019. The scenario with the most deaths is the combination of $\alpha^{(2019)}$ with no cross-border traffic.

In general, the scenarios with the least inner mobility are the ones with fewer deaths and the scenarios with more inner mobility with most. Furthermore, comparing only cross-border traffic between the same $\alpha_k(t)$, the scenario with fewer deaths is always with cross-border traffic of 2019.

Throughout this chapter, I have presented the main results of the Thesis. Next, in Chapter 5, I reflect on the implication of the results, their interpretation, and the limitations of the model and data.

5 Discussion

NPIs include social distancing measures adopted by society as well as government-mandated restrictions. Therefore, they were the most critical measure to control the spread of coronavirus disease at the beginning of the epidemic in the spring of 2020. This Thesis has analyzed the effect of NPIs in Finland on the outcome of the epidemic, taking as the primary indicator the incidence of reported cases.

The change in mobility from 2019 to 2020 reflected the effect of NPIs; it showed how the restrictions imposed by the government in symbiosis with the social distancing measures adopted by individuals altered the mobility patterns from one year to the other. This Thesis analyzed the effect of different NPIs on the outcome of the epidemic. The analysis was conducted through counterfactual scenarios in mobility, comparing the change in incidence and deaths under different mobility patterns during spring 2020. The purpose was to determine which mobility aspects are more influential when preventing the spread of the disease.

I first calibrated a compartmental metapopulation model that describes the evolution of coronavirus in Finland. The model was seeded by mobility data from a major telecommunications operator in Finland, and ABC was used to calibrate it to resemble the reported incidence by THL during spring 2020. The mobility of this calibrated metapopulation model was altered to form the counterfactual scenarios.

5.1 Calibration of the compartmental metapopulation model

Chapter 3 presented the incidence of the calibrated model and its comparison with the reported incidence by THL. The calibrated model did resemble the incidence reported by THL. Furthermore, the effective reproduction number (R_{eff}) that resulted from the calibration was similar to the one obtained by other studies in Finland [61]. Thus, the results obtained by the model were trustworthy.

The variation over time of R_{eff} reflected the effect of the applied NPIs. After the first implemented NPIs on 16 March 2020, the value of R_{eff} started to drop. These first NPIs included a generalized lockdown that reduced the number of contacts; the drop in the transmission rate of the disease associated with lockdown measures coincided with the observed in Refs. [52], [54], [59] for other countries. When the government reopened the borders of Uusimaa, R_{eff} was below unity, and it stayed here until the end of spring 2020. The result suggested two points: the progression of the epidemic, reflected in R_{eff}, was indeed influenced by NPIs, and the model could capture that influence.

However, there were shortcomings both in the modeling phase as well as in the calibration phase. In the modeling phase, the assumption of homogeneous mixing in a region was done. A closer-to-reality approach would be to consider the age of individuals inside each region and the mixing patterns of each age group. With the introduction of age groups, it would also be possible to model the associated risk of the elderly to become infectious, which is more significant than younger-age groups.

When calibrating the model an informative and often used metric is the hospital

occupation. The reported incidence underestimates the real number of infected people because of the high number of asymptomatic individuals. Moreover, the reported incidence is also tightly related to the number of performed tests. With the number of hospitalized individuals, one could estimate the number of unreported cases since hospital occupation tends to be a more stable and reliable metric. Unfortunately, in the analysis that was done in this Thesis, the number of hospitalized individuals could not be used because of the way THL reports it. THL reports the hospital occupation at a coarser level than it reports the incidence; it divides Finland into five sub-regions instead of twenty-one as for the incidence. I could have compromised and used a coarser geographical level to use hospitalizations, but there would not be as many geographical regions. Therefore, I chose to have more sub-regions to analyze the mobility between regions more in depth.

The number of performed tests, as the number of hospitalizations, was left out when calibrating the model. There are nowcast models that account for the delay in the reporting and the underreporting [58]. The approach taken in this Thesis could benefit from nowcast, and it would be possible to calibrate the model to the real number of infected people instead of the reported one. There are two direct advantages when doing this: (1) to calculate the number of unreported cases, (2) to have a more accurate model of the epidemic.

Another aspect worth exploring is the use of a more elaborate ABC algorithm to infer parameters. This Thesis used one of the simplest algorithms from the ABC family. Thus, a more elaborate algorithm, such as Sequential Monte Carlo ABC and post-processing the output of ABC [30], [37], may speed up the inference process and improve the calibration quality.

Finally, it is worth mentioning that the model is general and can be applied to any country or epidemic. If used in a different country, the model needs mobility and population data from the country in question. In addition, the epidemic parameters values, such as the latent and recovery periods, would be needed to use the model for another disease.

5.2 Effectiveness of NPIs

From the counterfactual scenarios presented in Chapter 4, the main conclusion was that limiting the mobility and number of contacts inside the regions had a more significant impact on the incidence, and the progression of the epidemic, than reducing the cross-border traffic between them. This result coincided with the results in Ref. [45]. NPIs that prevent the gathering of people, especially indoors where the coronavirus has proven to spread easier, are examples of limiting the contacts inside the regions.

The model showed that closing the borders of a highly infectious region negatively impacted the overall number of infections in a country. By keeping the borders of the region closed, both infectious and susceptible people would get locked in; this could promote the contacts inside the region, causing a rise in incidence. On the other hand, opening the borders of the region would let susceptible individuals travel and prevent getting infected in a less infectious region.

Furthermore, there are other two notable factors to consider when considering closing the borders of a region. The first one is that ensuring a completely closed region, in which no infectious individuals would go out, is unfeasible. As a result, there may be a leakage of undetected infectious individuals to other regions. Such leakage would not completely prevent the arrival of the disease to other regions but may help delay it, as suggested by Refs. [45], [64]. The second factor to consider is the number of infectious individuals in other regions. If there are many infectious individuals in other regions, the border restrictions to prevent the arrival of more may not be decisive to prevent further infections. Nevertheless, border restrictions would make a difference when there are no infectious people in other regions. Hence there is a need for an adequate surveillance system.

The observations made so far have analyzed the case of Finland. Finland has the characteristic that the capital region concentrates one-third of the population. The consequence of this characteristic is that the national incidence numbers will be closely related to the ones of the capital region. On the other hand, national incidence numbers, and other epidemic indicators, could not have such a strong dependence on a region in a country with a more evenly distributed population. Still, based on results, prioritizing the reduction of contacts inside regions would bring the best results when trying to control the spread of the epidemic. This is because reducing the contacts inside regions prevents infections independently of the population density.

The work performed in the Thesis was highly dependent on the mobility data available. A more refined analysis would require a more specific type of data. For example, knowing the exact number of visitors to relevant locations, such as restaurants or convenience stores, or directly having the anonymized location of the users of the telecommunications company would enable more in depth analysis.

Facebook [48] and Google [55] data sets have been used to try to overcome the lack of refined mobility data. However, the dataset provided by Facebook only indicates the relative change in mobility with respect to a baseline week; it does not link this relative change to a specific business or location. On the other hand, Google also reports the relative change in mobility and does link the change to a specific type of location. The caveat of their data is that there is no way to determine the original baseline visitors; this is why the relative change reported by Google is often used as a multiplicator for baseline mobility data [60], [75], or directly used to predict the incidence of new cases [62], [74]. Therefore, more specific information is still needed to permit a more precise approach like agent-based modeling [39].

Another inherent caveat to the data in this Thesis was that, as previously acknowledged, some municipalities on specific days had noisy measurements (see Chapter 3). Therefore, days with similar mobility to the ones with noisy measurements were used to replace it. Nevertheless, it is impossible to know how close or far the replaced data were from reality.

The results of this Thesis, and the previous studies, can be used as a stepping stone to a broader analysis on what happened in Finland and the world to inform the decisions of future epidemics. This Thesis has focused on the mobility restrictions between regions. Further analysis could be performed on the effect of specific

restrictions inside regions or the travel restrictions to arrive and leave the country. Moreover, as noted previously, the results of this Thesis were analyzing the case of Finland; future research may analyze if the effects observed here apply to another country or even in general for any region.

6 Conclusions

The Thesis has presented a compartmental metapopulation model, informed by mobile phone data, that analyzes the main NPIs applied in Finland from March 2020 to the end of May 2020, the spring of 2020. The period of spring 2020 is characterized by two facts. The first one is that during this period, COVID-19 reached Finland and caused a sudden appearance of new cases. Secondly, spring 2020 is the period when the government implemented the NPIs with the highest mobility restrictions.

Using the metapopulation model and counterfactual scenarios based on data, I evaluated the effectiveness of different NPIs. The counterfactual scenarios suggest that, when trying to prevent the spread of coronavirus, NPIs which aim to reduce the contacts that people make inside a region are more beneficial than NPIs that restrict the traffic between regions. Some examples of NPIs that reduce contacts are the closure of public places and restrictions on the number of people allowed to gather in closed spaces.

As mentioned, NPIs that restrict cross-border traffic between regions may be counter-effective. If closing the borders of a region promotes contacts inside it and the disease has already been established, then the closure would promote the spread of the disease. Furthermore, if the region is densely populated, the disadvantages of restricting cross-border traffic may outweigh the advantages. The number of infected people in the populated closed region may further increase the national incidence when closing its borders than if kept open. An example of border closure happened in Finland when the borders of Uusimaa were closed from 27 March to 15 April 2020.

Although this analysis comes after a considerable amount of time after the first outbreak of coronavirus cases, the results are not only applicable in the context of coronavirus in Finland. Therefore, decision-makers may apply the results of this Thesis, and other relevant studies, to other epidemics in which pharmaceutical interventions, such as a vaccine or medical treatments, are not available or not feasible to apply.

Over the past year, the scientific community has already performed several studies around COVID-19. Nonetheless, a posteriori research on it, which takes advantage of the diverse released data sources and information, would still provide some benefit in controlling future epidemics. In this respect, it is essential to understand the most effective measures to control coronavirus, have models ready to describe the disease, and provide adequate information for such models.

This Thesis has made it clear that public and available information is vital for all the performed studies and any a posteriori study that may happen. The coronavirus epidemic has brought new light in the field of epidemic and its importance. Therefore, it is in the best interest to keep releasing information to inform future research in the field.

References

[1] R. Ross, "An application of the theory of probabilities to the study of a priori pathometry.—part I," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 92, no. 638, pp. 204–230, Feb. 1916. DOI: 10.1098/rspa.1916.0007. [Online]. Available: https://doi.org/10.1098%5C%2Frspa.1916.0007.

[2] R. Ross and H. P. Hudson, "An application of the theory of probabilities to the study of a priori pathometry.—part II," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 93, no. 650, pp. 212–225, May 1917. DOI: 10.1098/rspa.1917.0014. [Online]. Available: https://doi.org/10.1098%5C%2Frspa.1917.0014.

[3] ——, "An application of the theory of probabilities to the study of a priori pathometry.—part III," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 93, no. 650, pp. 225–240, May 1917. DOI: 10.1098/rspa.1917.0015. [Online]. Available: https://doi.org/10.1098%5C%2Frspa.1917.0015.

[4] "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, Aug. 1927. DOI: 10.1098/rspa.1927.0118. [Online]. Available: https://doi.org/10.1098%5C%2Frspa.1927.0118.

[5] M. Greenwood, "On the statistical measure of infectiousness," *Journal of Hygiene*, vol. 31, no. 3, pp. 336–351, Jul. 1931. DOI: 10.1017/s002217240001086x. [Online]. Available: https://doi.org/10.1017%5C%2Fs002217240001086x.

[6] G. K. Zipf, "The p1 p2/d hypothesis: On the intercity movement of persons," *American Sociological Review*, vol. 11, no. 6, pp. 677–686, 1946, ISSN: 00031224. [Online]. Available: http://www.jstor.org/stable/2087063.

[7] H. Abbey, "An examination of the reed-frost theory of epidemics," *Human Biology*, vol. 24, no. 3, pp. 201–233, 1952, ISSN: 00187143, 15346617. [Online]. Available: http://www.jstor.org/stable/41449060.

[8] N. T. J. Bailey, "The total size of a general stochastic epidemic," *Biometrika*, vol. 40, no. 1/2, p. 177, Jun. 1953. DOI: 10.2307/2333107. [Online]. Available: https://doi.org/10.2307%5C%2F2333107.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, Jun. 1953. DOI: 10.1063/1.1699114. [Online]. Available: https://doi.org/10.1063%5C%2F1.1699114.

[10] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970. DOI: 10.1093/biomet/57.1.97. [Online]. Available: https://doi.org/10.1093%5C%2Fbiomet%5C%2F57.1.97.

[11] R. M. ANDERSON and R. M. MAY, "Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes," *Mathematical Medicine and Biology*, vol. 1, no. 3, pp. 233–266, 1984. DOI: 10.1093/imammb/1.3.233. [Online]. Available: https://doi.org/10.1093%5C%2Fimammb%5C%2F1.3.233.

[12] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984. DOI: 10.1109/tpami.1984.4767596. [Online]. Available: https://doi.org/10.1109%5C%2Ftpami.1984.4767596.

[13] D. B. Rubin, "Bayesianly justifiable and relevant frequency calculations for the applied statistician," *The Annals of Statistics*, vol. 12, no. 4, Dec. 1984. DOI: 10.1214/aos/1176346785. [Online]. Available: https://doi.org/10.1214%5C%2Faos%5C%2F1176346785.

[14] L. A. Rvachev and I. M. Longini, "A mathematical model for the global spread of influenza," *Mathematical Biosciences*, vol. 75, no. 1, pp. 3–22, Jul. 1985. DOI: 10.1016/0025-5564(85)90064-1. [Online]. Available: https://doi.org/10.1016%5C%2F0025-5564%5C%2885%5C%2990064-1.

[15] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, Sep. 1987. DOI: 10.1016/0370-2693(87)91197-x. [Online]. Available: https://doi.org/10.1016%5C%2F0370-2693%5C%2887%5C%2991197-x.

[16] O. Diekmann, J. Heesterbeek, and J. Metz, "On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations," *Journal of Mathematical Biology*, vol. 28, no. 4, Jun. 1990. DOI: 10.1007/bf00178324. [Online]. Available: https://doi.org/10.1007%5C%2Fbf00178324.

[17] S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly, "Inferring coalescence times from dna sequence data," *Genetics*, vol. 145, no. 2, pp. 505–518, 1997, ISSN: 0016-6731. eprint: https://www.genetics.org/content/145/2/505.full.pdf. [Online]. Available: https://www.genetics.org/content/145/2/505.

[18] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, "Population growth of human y chromosomes: A study of y chromosome microsatellites," *Molecular Biology and Evolution*, vol. 16, no. 12, pp. 1791–1798, Dec. 1999. DOI: 10.1093/oxfordjournals.molbev.a026091. [Online]. Available: https://doi.org/10.1093%5C%2Foxfordjournals.molbev.a026091.

[19] M. L. Cohen, "Changing patterns of infectious disease," *Nature*, vol. 406, no. 6797, pp. 762–767, Aug. 2000. DOI: 10.1038/35021206. [Online]. Available: https://doi.org/10.1038%5C%2F35021206.

[20] C. of Neglected Tropical Diseases, "Who report on global surveillance of epidemic-prone infectious diseases," World Health Organization, Tech. Rep. WHO/CDS/CSR/ISR/2000.1, Feb. 2000, p. 131. [Online]. Available: https://www.who.int/publications/i/item/WHO-CDS-CSR-ISR-2000.1.

[21] D. Bernoulli and S. Blower, "An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it," *Reviews in Medical Virology*, vol. 14, no. 5, pp. 275–288, Aug. 2004. DOI: 10.1002/rmv.443. [Online]. Available: https://doi.org/10.1002%5C%2Frmv.443.

[22] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "The modeling of global epidemics: Stochastic dynamics and predictability," *Bulletin of Mathematical Biology*, vol. 68, no. 8, pp. 1893–1921, Jun. 2006. DOI: 10.1007/s11538-006-9077-9. [Online]. Available: https://doi.org/10.1007%5C%2Fs11538-006-9077-9.

[23] C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell, "Synchrony, waves, and spatial hierarchies in the spread of influenza," *Science*, vol. 312, no. 5772, pp. 447–451, Mar. 2006. DOI: 10.1126/science.1125237. [Online]. Available: https://doi.org/10.1126%5C%2Fscience.1125237.

[24] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, "Modeling the worldwide spread of pandemic influenza: Baseline case and containment interventions," *PLoS Medicine*, vol. 4, no. 1, A. P. Galvani, Ed., e13, Jan. 2007. DOI: 10.1371/journal.pmed.0040013. [Online]. Available: https://doi.org/10.1371%5C%2Fjournal.pmed.0040013.

[25] L. J. S. Allen, "An introduction to stochastic epidemic models," in *Mathematical Epidemiology*, Springer Berlin Heidelberg, 2008, pp. 81–130. DOI: 10.1007/978-3-540-78911-6_3. [Online]. Available: https://doi.org/10.1007%5C%2F978-3-540-78911-6_3.

[26] N. Bharti, Y. Xia, O. N. Bjornstad, and B. T. Grenfell, "Measles on the edge: Coastal heterogeneities and infection dynamics," *PLoS ONE*, vol. 3, no. 4, S. I. Hay, Ed., e1941, Apr. 2008. DOI: 10.1371/journal.pone.0001941. [Online]. Available: https://doi.org/10.1371%5C%2Fjournal.pone.0001941.

[27] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences*, vol. 106, no. 51, pp. 21 484–21 489, Dec. 2009. DOI: 10.1073/pnas.0906910106. [Online]. Available: https://doi.org/10.1073%5C%2Fpnas.0906910106.

[28] K. Csilléry, M. G. Blum, O. E. Gaggiotti, and O. François, "Approximate bayesian computation (ABC) in practice," *Trends in Ecology & Evolution*, vol. 25, no. 7, pp. 410–418, Jul. 2010. DOI: 10.1016/j.tree.2010.04.001. [Online]. Available: https://doi.org/10.1016%5C%2Fj.tree.2010.04.001.

[29] O. Diekmann, *Mathematical tools for understanding infectious diseases dynamics*, eng, ser. Princeton series in theoretical and computational biology. Princeton: Princeton University Press, 2012.

[30] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, "Approximate bayesian computational methods," *Statistics and Computing*, vol. 22, no. 6, pp. 1167–1180, 2012. DOI: 10.1007/s11222-011-9288-2. [Online]. Available: https://doi.org/10.1007/s11222-011-9288-2.

[31] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, no. 7392, pp. 96–100, 2012. DOI: 10.1038/nature10856. [Online]. Available: https://doi.org/10.1038/nature10856.

[32] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, Nov. 2013. DOI: 10.1201/b16018. [Online]. Available: https://doi.org/10.1201%5C%2Fb16018.

[33] S. Charaudeau, K. Pakdaman, and P.-Y. Boëlle, "Commuter mobility and the spread of infectious diseases: Application to influenza in france," *PLoS ONE*, vol. 9, no. 1, A. Vespignani, Ed., e83002, Jan. 2014. DOI: 10.1371/journal.pone.0083002. [Online]. Available: https://doi.org/10.1371%5C%2Fjournal.pone.0083002.

[34] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, "On the use of human mobility proxies for modeling epidemics," *PLOS Computational Biology*, vol. 10, no. 7, pp. 1–15, Jul. 2014. DOI: 10.1371/journal.pcbi.1003716. [Online]. Available: https://doi.org/10.1371/journal.pcbi.1003716.

[35] *Analyzing and modeling spatial and temporal dynamics of infectious diseases*, eng. Hoboken, New Jersey: John Wiley & Sons, Inc., 2015, ISBN: 1-118-62991-4.

[36] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," vol. 87, no. 3, pp. 925–979, Aug. 2015. DOI: 10.1103/revmodphys.87.925. [Online]. Available: https://doi.org/10.1103%5C%2Frevmodphys.87.925.

[37] J. Lintusaari, M. U. Gutmann, R. Dutta, S. Kaski, and J. Corander, "Fundamentals and Recent Developments in Approximate Bayesian Computation," *Systematic Biology*, vol. 66, no. 1, e66–e82, Sep. 2016, ISSN: 1063-5157. DOI: 10.1093/sysbio/syw077. eprint: https://academic.oup.com/sysbio/article-pdf/66/1/e66/24194739/syw077.pdf. [Online]. Available: https://doi.org/10.1093/sysbio/syw077.

[38] D. Adam, "A guide to r — the pandemic's misunderstood metric," *Nature*, vol. 583, no. 7816, pp. 346–348, Jul. 2020. DOI: 10.1038/d41586-020-02009-w. [Online]. Available: https://doi.org/10.1038%5C%2Fd41586-020-02009-w.

[39] A. Aleta, D. Martín-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. L. Jr, S. Merler, A. Pentland, A. Vespignani, E. Moro, and Y. Moreno, "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19," vol. 4, no. 9, pp. 964–971, Aug. 2020. DOI: 10.1038/s41562-020-0931-9. [Online]. Available: https://doi.org/10.1038%5C%2Fs41562-020-0931-9.

[40] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang, "Presumed asymptomatic carrier transmission of COVID-19," *JAMA*, vol. 323, no. 14, p. 1406, Apr. 2020. DOI: 10.1001/jama.2020.2565. [Online]. Available: https://doi.org/10.1001%5C%2Fjama.2020.2565.

[41] BBC. (2020). "Coronavirus: Europe now epicentre of the pandemic, says WHO," [Online]. Available: https://www.bbc.com/news/world-europe-51876784 (visited on 09/16/2021).

[42] S. Bialek, E. Boundy, V. Bowen, N. Chow, A. Cohn, N. Dowling, S. Ellington, R. Gierke, A. Hall, J. MacNeil, P. Patel, G. Peacock, T. Pilishvili, H. Razzaghi, N. Reed, M. Ritchey, and E. S.-S. and, "Severe outcomes among patients with coronavirus disease 2019 (COVID-19) — united states, february 12–march 16, 2020," *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, no. 12, pp. 343–346, Mar. 2020. DOI: 10.15585/mmwr.mm6912e2. [Online]. Available: https://doi.org/10.15585%5C%2Fmmwr.mm6912e2.

[43] CDC. (2020). "Nonpharmaceutical interventions (NPIs)," [Online]. Available: https://www.cdc.gov/nonpharmaceutical-interventions/index.html (visited on 09/16/2021).

[44] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of COVID-19 explain inequities and inform reopening," *Nature*, vol. 589, no. 7840, pp. 82–87, Nov. 2020. DOI: 10.1038/s41586-020-2923-3. [Online]. Available: https://doi.org/10.1038%5C%2Fs41586-020-2923-3.

[45] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, Mar. 2020. DOI: 10.1126/science.aba9757. [Online]. Available: https://doi.org/10.1126%5C%2Fscience.aba9757.

[46] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *The Lancet Infectious Diseases*, vol. 20, no. 5, pp. 533–534, May 2020. DOI: 10.1016/s1473-3099(20)30120-1. [Online]. Available: https://doi.org/10.1016%5C%2Fs1473-3099%5C%2820%5C%2930120-1.

[47] S. Engebretsen, K. Engø-Monsen, M. A. Aleem, E. S. Gurley, A. Frigessi, and B. F. de Blasio, "Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: The case of bangladesh," *Journal of the Royal Society, Interface*, vol. 17, no. 167, p. 20 190 809, Jun. 2020, ISSN: 1742-5689.

DOI: 10.1098/rsif.2019.0809. [Online]. Available: https://europepmc.org/articles/PMC7328378.

[48] Facebook. (2020). "Protecting privacy in facebook mobility data during the covid-19 response," [Online]. Available: https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/ (visited on 09/17/2021).

[49] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, eabb6936, Mar. 2020. DOI: 10.1126/science.abb6936.

[50] FHI. (2020). "Statistikk om koronavirus og covid-19," [Online]. Available: https://www.fhi.no/sv/smittsomme-sykdommer/corona/dags--og-ukerapporter/dags--og-ukerapporter-om-koronavirus/ (visited on 09/17/2021).

[51] Finnish National Agency for Education. (2020). "COVID-related restrictions on education gradually lifted in finland from 14.5.2020," [Online]. Available: https://www.oph.fi/en/news/2020/covid-related-restrictions-education-gradually-lifted-finland-1452020 (visited on 09/28/2021).

[52] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, P. N. Perez-Guzman, N. Schmit, L. Cilloni, K. E. C. Ainslie, M. Baguelin, A. Boonyasiri, O. Boyd, L. Cattarino, L. V. Cooper, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, B. Djaafara, I. Dorigatti, S. L. van Elsland, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, N. C. Grassly, W. D. Green, T. Hallett, A. Hamlet, W. Hinsley, B. Jeffrey, E. Knock, D. J. Laydon, G. Nedjati-Gilani, P. Nouvellet, K. V. Parag, I. Siveroni, H. A. Thompson, R. Verity, E. Volz, C. E. Walters, H. Wang, Y. Wang, O. J. Watson, P. Winskill, X. Xi, P. G. T. Walker, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, and S. B. and, "Estimating the effects of non-pharmaceutical interventions on COVID-19 in europe," *Nature*, vol. 584, no. 7820, pp. 257–261, Jun. 2020. DOI: 10.1038/s41586-020-2405-7. [Online]. Available: https://doi.org/10.1038%5C%2Fs41586-020-2405-7.

[53] Folkhälsomyndigheten. (2020). "Bekräftade fall i sverige – daglig uppdatering," [Online]. Available: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analyser/bekraftade-fall-i-sverige/ (visited on 09/17/2021).

[54] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, "Spread and dynamics of the COVID-19 epidemic in italy: Effects of emergency containment measures," vol. 117, no. 19, pp. 10 484–10 491, Apr. 2020. DOI: 10.1073/pnas.2004978117. [Online]. Available: https://doi.org/10.1073%5C%2Fpnas.2004978117.

[55] Google LLC. (2020). "Google covid-19 community mobility reports," [Online]. Available: https://www.google.com/covid19/mobility/ (visited on 09/17/2021).

[56] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020. DOI: 10.1016/s0140-6736(20)30183-5. [Online]. Available: https://doi.org/10.1016%5C%2Fs0140-6736%5C%2820%5C%2930183-5.

[57] Kuntaliitto. (2020). "Specialist medical care," [Online]. Available: https://www.kuntaliitto.fi/sosiaali-ja-terveysasiat/terveydenhuolto/erikoissairaanhoito (visited on 09/28/2021).

[58] S. F. McGough, M. A. Johansson, M. Lipsitch, and N. A. Menzies, "Nowcasting by bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking," vol. 16, no. 4, B. Lewis, Ed., e1007735, Apr. 2020. DOI: 10.1371/journal.pcbi.1007735. [Online]. Available: https://doi.org/10.1371%5C%2Fjournal.pcbi.1007735.

[59] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, M. Jit, P. Klepac, S. Flasche, S. Clifford, C. A. B. Pearson, J. D. Munday, S. Abbott, H. Gibbs, A. Rosello, B. J. Quilty, T. Jombart, F. Sun, C. Diamond, A. Gimma, K. van Zandvoort, S. Funk, C. I. Jarvis, W. J. Edmunds, N. I. Bosse, and J. Hellewell, "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in wuhan, china: A modelling study," *The Lancet Public Health*, vol. 5, no. 5, e261–e270, May 2020. DOI: 10.1016/s2468-2667(20)30073-6. [Online]. Available: https://doi.org/10.1016%5C%2Fs2468-2667%5C%2820%5C%2930073-6.

[60] N. W. Ruktanonchai, J. R. Floyd, S. Lai, C. W. Ruktanonchai, A. Sadilek, P. Rente-Lourenco, X. Ben, A. Carioli, J. Gwinn, J. E. Steele, O. Prosper, A. Schneider, A. Oplinger, P. Eastham, and A. J. Tatem, "Assessing the impact of coordinated COVID-19 exit strategies across europe," vol. 369, no. 6510, pp. 1465–1470, Jul. 2020. DOI: 10.1126/science.abc5096. [Online]. Available: https://doi.org/10.1126%5C%2Fscience.abc5096.

[61] J. Solanpää. (2020). "COVID-19 outbreak in Finland," [Online]. Available: https://covid19.solanpaa.fi/ (visited on 10/22/2021).

[62] M. Sulyok and M. Walker, "Community movement and COVID-19: A global study using google's community mobility reports," vol. 148, 2020. DOI: 10.1017/s0950268820002757. [Online]. Available: https://doi.org/10.1017%5C%2Fs0950268820002757.

[63] The Guardian. (2020). "Italy records its deadliest day of coronavirus outbreak with 475 deaths," [Online]. Available: https://www.theguardian.com/world/2020/mar/18/coronavirus-lockdown-eu-belgium-germany-adopt-measures (visited on 09/16/2021).

[64] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. G. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, B. Wang, P. Yang, Y. Cui, Y. Song, P. Zheng, Q. Wang, O. N. Bjornstad, R. Yang, B. T. Grenfell, O. G. Pybus, and C. Dye, "An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china," *Science*, vol. 368, no. 6491, pp. 638–642, 2020. DOI: 10.1126/science.abb6105. eprint: https://www.science.org/doi/pdf/10.1126/science.abb6105. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.abb6105.

[65] Z.-D. Tong, A. Tang, K.-F. Li, P. Li, H.-L. Wang, J.-P. Yi, Y.-L. Zhang, and J.-B. Yan, "Potential presymptomatic transmission of SARS-CoV-2, zhejiang province, china, 2020," *Emerging Infectious Diseases*, vol. 26, no. 5, pp. 1052–1054, May 2020. DOI: 10.3201/eid2605.200198. [Online]. Available: https://doi.org/10.3201%5C%2Feid2605.200198.

[66] WHO. (2020). "COVID-19 - china," [Online]. Available: https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229 (visited on 09/16/2021).

[67] ——, (2020). "WHO director-general's opening remarks at the media briefing on COVID-19 - 11 march 2020," [Online]. Available: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (visited on 09/16/2021).

[68] ——, (2020). "WHO statement on novel coronavirus in thailand," [Online]. Available: https://www.who.int/news/item/13-01-2020-who-statement-on-novel-coronavirus-in-thailand (visited on 09/16/2021).

[69] YLE. (2020). "Finland closes schools, declares state of emergency over coronavirus," [Online]. Available: https://yle.fi/uutiset/osasto/news/finland_closes_schools_declares_state_of_emergency_over_coronavirus/11260062 (visited on 09/17/2021).

[70] ——, (2020). "Gov't eases curbs on cross-border travel, public events," [Online]. Available: https://yle.fi/uutiset/osasto/news/govt_eases_curbs_on_cross-border_travel_public_events/11396707 (visited on 09/17/2021).

[71] ——, (2020). "Seurasimme historiallista hetkeä rajalla - poliisikomentaja radiopuhelimessa: 'Täällä johto yksi, Uudenmaan sulku puretaan välittömästi'," [Online]. Available: https://yle.fi/uutiset/3-11306225 (visited on 09/17/2021).

[72] ——, (2020). "Uusimaa closes borders after late-night vote in parliament," [Online]. Available: https://yle.fi/uutiset/osasto/news/uusimaa_closes_borders_after_late-night_vote_in_parliament/11280719 (visited on 09/17/2021).

[73] CDC. (2021). "Scientific brief: Sars-cov-2 transmission," [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html (visited on 09/16/2021).

[74] C. Cot, G. Cacciapaglia, and F. Sannino, "Mining google and apple mobility data: Temporal anatomy for COVID-19 social distancing," vol. 11, no. 1, Feb. 2021. DOI: 10.1038/s41598-021-83441-4. [Online]. Available: https://doi.org/10.1038%5C%2Fs41598-021-83441-4.

[75] R. Dutta, S. N. Gomes, D. Kalise, and L. Pacchiardi, "Using mobility data in the design of optimal lockdown strategies for the COVID-19 pandemic," vol. 17, no. 8, M. (Ferrari, Ed., e1009236, Aug. 2021. DOI: 10.1371/journal.pcbi.1009236. [Online]. Available: https://doi.org/10.1371%5C%2Fjournal.pcbi.1009236.

[76] IATA. (2021). "Air passenger traffic and sales data," [Online]. Available: https://www.iata.org/en/services/statistics/passenger-traffic-data/ (visited on 11/12/2021).

[77] N. Perra, "Non-pharmaceutical interventions during the covid-19 pandemic: A review," *Physics Reports*, vol. 913, pp. 1–52, May 2021, ISSN: 0370-1573. DOI: 10.1016/j.physrep.2021.02.001. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2021.02.001.

[78] Statistics Finland. (2021). "Statistics finland's free-of-charge statistical databases," [Online]. Available: https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/ (visited on 09/20/2021).

[79] THL. (2021). "Confirmed corona cases in finland (COVID-19)," [Online]. Available: https://thl.fi/en/web/thlfi-en/statistics-and-data/data-and-services/open-data/confirmed-corona-cases-in-finland-covid-19- (visited on 09/17/2021).

[80] ——, (2021). "Hybridistrategian seurantaraportit," [Online]. Available: https://thl.fi/fi/web/infektiotaudit-ja-rokotukset/ajankohtaista/ajankohtaista-koronaviruksesta-covid-19/tilannekatsaus-koronaviruksesta/koronaviruksen-seuranta (visited on 11/14/2021).