Evaluation of different mobility models in equation-based infectious disease dynamics

Master thesis

submitted by **Carlotta Felicia Gerstein**

on February 6, 2025

at the Mathematical Institute of the University of Cologne
in cooperation with the German Aerospace Center (DLR)

Main supervision: **Prof. Dr.-Ing. Gregor Gassner**
Supervision (DLR): **Dr. Martin Joachim Kühn**

Contents

List of Figures

List of Tables

1 Introduction

Infectious diseases represent a pervasive global phenomenon, affecting millions of people each year [36]. Particularly in low-income countries, as defined by the World Bank [37], communicable diseases pose a significant threat, with eight of the ten leading causes of death in 2021 being classified as communicable [45]. While diseases such as malaria and tuberculosis record a significant decrease in mortality rates compared to the year 2000 [45], the recent pandemic of SARS-CoV-2 has demonstrated the potential for rapid disease outbreaks. During the first two years after the onset of the pandemic, the World Health Organization (WHO) has recorded over 14.9 million excess deaths attributed to COVID-19 [43]. In 2021, COVID-19 emerged as the second leading cause of death globally [45]. Although COVID-19 is no longer considered to be a *public health emergency of international concern*, the necessity of preparing for future events remains [46]. Continuously changing and evolving viruses and bacteria contribute to the fact that infectious diseases will emerge and reemerge in the future [13, 36]. The advent of globalization, in particular the growth in short-term and air travel, has led to a significant potential for rapid geographical spread of infectious diseases [34, 36], requiring governments to act swiftly in case of an outbreak.

These reasons make it necessary to comprehend the dynamics involved in the spread of infectious diseases and to provide useful tools to support decision-making in disease control. In this regard, mathematical models have proven to be powerful methods [36, 24]. As has been demonstrated in several disease outbreaks and pandemics, they represent an important tool to support public health policy in designing prevention and intervention strategies [13, 36].

Compartmental models based on ordinary differential equations (ODEs) are among the earliest and most widely used epidemiological models for describing disease dynamics within a population [9, 24]. Given the extensive research on these models, they represent well-established methods for analysis and benefit from their comparatively low computational cost.

However, they are based on the assumption that the population is "well-mixed", implying that all compartments mix in a homogeneous pattern. While individuals in different compartments can have different properties concerning the transmission of the disease, the "well-mixed" assumption describes the absence of geographical or social barriers within the population. This assumption is not only overly simplified [25], particularly on

a large scale, it also results in the loss of spatial information regarding the population. To address some of these limitations, we examine models that take the spatial heterogeneity of the regions into consideration. Spatially resolved models allow for the evaluation of regional interventions and interventions connected to the movement of individuals [19].

There are multiple modeling approaches for introducing spatial heterogeneity and mobility to epidemiological models. Models that account for spatial heterogeneity incorporate space in either continuous or discrete form [25]. Continuous modeling of space often describes the (continuous) motion of individuals and includes partial differential equations of reaction-diffusion type [7, Ch. 15.3]. Considering air travel or sparsely populated regions, these models often fail to capture the spatial spread of the disease on a larger scale [3]. For that case, space is often introduced through spatial discretization, of which metapopulation models are one of the most popular representatives [25]. Metapopulation models are characterized by the division of the population into multiple distinct subpopulations that interact on some level. When modeling spatial spread, the subpopulations are identified with spatially separated regions, connected through links that represent mobility patterns.

In this thesis, we will study two different approaches on integrating mobility to a model. The first approach is represented by a model based on ODEs that incorporates the effect of mobility directly in its equations [40, 23]. In the second approach [20], the authors proposed to connect decoupled ODE systems for the local regions through a graph. We evaluate these two approaches on metapopulation models, extend the first approach and analyze all three approaches with regard to practicability, complexity, and predictions.

The structure of this thesis is as follows. The foundation for studying metapopulation models is laid in Chapter 2 by introducing a first compartmental model for describing the dynamics of infectious diseases. In Chapter 3, we present the metapopulation models, following the aforementioned equation-based and graph-based approach. Further, we derive the next generation matrices for the equation-based models to validate them using the basic reproduction numbers. We examine the implementations of both approaches and perform a detailed complexity analysis in terms of the number of floating point operations required for the calculation of numerical solutions in Chapter 4. Subsequently, in Chapter 5, the models discussed are compared with regard to their performance based on numerical simulations. Using the example of North Rhine-Westphalia in Germany, we compare the predictions of the models for a fictional scenario. Finally, we draw a conclusion in Chapter 6 and provide a foundation for future research.

2 Infection dynamics models based on ordinary differential equations

In this chapter, we derive a simple compartmental model which describes the dynamics of an infectious disease spread within a population. Afterward, we present possible extensions to this model and generalize it to take the age structure of a population into account.

In the process of formulating compartmental models, a population is divided into a number of distinct classes, which are called *compartments*. Every compartment combines individuals with a specific property which, in the epidemiological context, are often different stages of the disease. Each individual is assigned to a single compartment and can transition between the different compartments [25].

In compartmental models, the flows between the compartments are described by differential equations [9, p. 807 f.]. The type of the differential equation depends on the purpose and the assumptions of the model, and can vary from ordinary to integro-, partial or stochastic differential equations [13, 2].

One of the earliest compartmental models to describe the dynamics of infectious diseases was proposed by Kermack and McKendrick in 1927 [15]. Following their proposal, the population is divided into the classes S, I, and R which are defined as follows:

- S refers to the group of *susceptible* individuals, i.e., those who can be infected with the disease.
- I contains the *infected/infectious* individuals who spread the disease to the susceptible individuals by contact.
- R is the group of *recovered/removed* individuals who are not infectious and cannot be infected by others.

As time progresses, individuals interact with each other and contacts of a susceptible with an infectious individual lead to a spread of the disease. As individuals get infected, they move to the I compartment and the population of susceptible individuals decreases by the same number as the infective population increases. After the infective period, individuals recover and move to the R compartment. Given that a compartment represents

Figure 2.1: **Schematic representation of the flows in a *SIR* model.** Individuals in the Susceptible compartment can become infected on contact with infectious individuals and move to the compartment Infected. After recovery, they move to the Recovered compartment. The only compartment in which individuals are infectious and able to spread the infection is the infected compartment, which is highlighted in red.

a share of the population rather than distinct individuals, the movement of individuals between compartments corresponds to the transition of shares of the population.

The meaning of the R compartment can vary. In the initial formulation by Kermack and McKendrick [15], these individuals were either individuals who had recovered or individuals who had died of the disease. However, there are also potential scenarios where removed individuals are those who have been isolated from the population [7, p. 23] or immunized, e.g., through vaccination [25, p. 216]. In the following, we will interpret individuals of the R compartment as fully recovered and omit the possibility of dying from the disease for the models considered in this thesis. A model which uses these compartments is called an *SIR* model. A schematic of these compartmental models is typically presented as a flowchart. The flowchart of the *SIR* model can be found in Fig. 2.1. It shows that individuals in the Recovered compartment cannot move to any other compartment and remain removed for the rest of the time modeled.

As models are a simplification of reality, it is necessary to make certain assumptions before formulating a model. The prefix *SIR* indicates that individuals progress through the infection course described above. Thus, the model used varies according to the disease being studied. For instance, an *SIR* model is based on the assumption that individuals are unable to reinfect themselves with the disease, at least within the time frame under consideration. In the case of diseases which do not confer immunity, an *SIS*-type model could be used and *SIRS*-type models are more suitable for diseases with temporary, fast waning immunity [7, p. 23]. Here, we will only study epidemic diseases, which are characterized by a sudden outbreak followed by a period of disappearance [7, p. 21]. While these diseases may recur, there is often a period of several years between outbreaks. Given the relatively short temporal scale of a single outbreak, it is reasonable to neglect birth and natural death rates, maintaining a constant population size N. In addition, given the short time period, it is acceptable to assume that, for diseases providing temporary immunity, the immunity does not wane during the specified period.

A further significant assumption is that the population is “well-mixed”, meaning that every individual has the same probability of contacting any other individual. However, this is highly simplified at multiple levels, as, e.g., it can be reasonably assumed that

individuals meet their colleagues or friends more frequently than strangers. Additionally, the population of a particular compartment is assumed to be homogeneous, i.e., every individual within a compartment behaves the same way and has the same properties concerning the transmission dynamics of diseases. One of the most significant aspects of heterogeneity in disease modeling is the age structure of populations [7, p. 14]. The age of individuals has an impact on parameters governing transmission dynamics, such as transmission rates, recovery rates or death rates [5, 39, 26]. Furthermore, age strongly impacts the contact patterns of the population, with individuals tending to meet preferably with individuals of a similar age [28]. After constructing a simple compartmental model for a homogeneous population, we will gradually look at an increasingly heterogeneous population, starting with a stratification by age in Section 2.3.

2.1 A simple ODE-SIR model

In the compartmental models in this thesis, the transitions between compartments are described by ordinary differential equations (ODEs) where S, I and R are functions of time $t \in \mathbb{R}_{\geq 0}$. The number of susceptible individuals at a time t is denoted by $S(t)$. Respectively, the number of infected and recovered individuals at time t are denoted by $I(t)$ and $R(t)$. The total population size N satisfies

$$N = S(t) + I(t) + R(t), \quad \forall t \in \mathbb{R}_{\geq 0}. \tag{2.1.1}$$

Over time, the infection spreads and shares of the population are transitioned among compartments. For reasons of readability, we will often drop t from the notation and write S, I, and R for the number of individuals in each compartment.

Let $\phi \in \mathbb{R}_{\geq 0}$ be the fixed, mean number of contacts an individual has during a unit time, e.g., a day. The fraction of infected individuals in the total population is I/N, which can be interpreted as the probability of meeting an infected individual, given there is a contact. Accordingly, the mean number of contacts with infected individuals for a susceptible individual in unit time, is $\phi I/N$. For instance, let us assume that an individual has about four contacts with other individuals per day and that half of the population is infected. In this scenario, we would assume that the individual has two contacts with infected individuals during that day, which implies that infected individuals do not tend to isolate themselves more often than susceptible individuals. Let $\rho \in [0, 1]$ be the fixed probability of infection on contact, i.e., the probability that the pathogen is transmitted if a contact occurs. This results in $\phi\rho I/N$ transmissions per susceptible individual and $\phi\rho SI/N$ new transmissions during unit time in total. From these considerations follows:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\phi\rho\frac{SI}{N}. \tag{2.1.2}$$

The population of susceptible individuals is therefore reduced by the number of new transmissions during unit time. In some sources, it is assumed that every contact is effective, i.e., if there is a contact with an infected individual, the pathogen is transmitted with certainty, see e.g., [7, p. 25]. In this case, the probability of infection is $\rho = 1$ and we get $\phi SI/N$ transmissions. In some other sources, the product of ϕ and ρ is combined to a single parameter $\beta = \phi\rho$ of mean number of effective contacts, which leads to

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta\frac{SI}{N} \tag{2.1.3}$$

transmissions in unit time. This form of transmission rate is called *standard incidence*. If the population size N is constant, we can simplify to

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\widehat{\beta}SI \tag{2.1.4}$$

by $\widehat{\beta} = \beta/N = \phi\rho/N$. This corresponds to another common assumption, called *mass action incidence*, stating that the incidence is proportional to the product of S and I [25, p. 37].

Both assumptions agree for a constant population size, but differ otherwise. Since $\beta = \widehat{\beta}N$ gives the mean number of effective contacts, with increasing population size the mass action incidence implicitly assumes a higher contact rate [13, p. 602 f.]. In comparison to this, the standard incidence assumes the number of effective contacts β and the contact probability is determined as β/N. According to [13], the standard incidence is the more accurate assumption, since the number of contacts and the contact patterns are often similar in large and small populations.

It is common to assume that the infected individuals leave the Infected compartment at a constant rate [7] which corresponds to the inverse of the mean infectious period $T_I \in \mathbb{R}_{>0}$. With equation (2.1.2) this leads us to the following system of ODEs:

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\phi\rho\frac{IS}{N}, \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \phi\rho\frac{IS}{N} - \frac{1}{T_I}I, \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \frac{1}{T_I}I.
\end{aligned} \tag{2.1.5}$$

If the compartment sizes at time zero

$$S(0) = S_0,\ I(0) = I_0,\ R(0) = R_0 \tag{2.1.6}$$

such that $N = S_0 + I_0 + R_0$ are known, then (2.1.5) and (2.1.6) formulate an initial value problem.

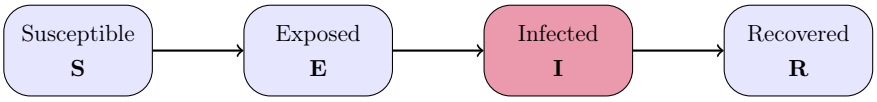


Figure 2.2: **Schematic representation of the flows in a *SEIR* model.** If an infectious individual transmits the pathogen to a susceptible individual, it moves to the Exposed compartment. After the latent period, the individual moves to the Infected compartment and eventually recovers, entering the Recovered compartment. The only compartment in which individuals are infectious and able to spread the infection is the Infected compartment, which is highlighted in red.

2.2 State extensions to the ODE-SIR model

As mentioned earlier, there are many adaptations and extensions to the *SIR* model presented above. The choice of additional compartments or features depends on the disease being studied, the research question, and the purpose of the model. Extending the model can make it more realistic and accurate, and can change the conclusions that we draw from the model [25, p. 91].

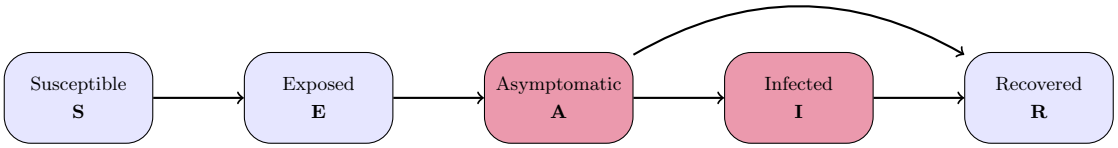
A first and commonly used extension is the consideration of the latent phase. It is defined as the time between the exposure to the disease and the onset of infectiousness [27]. Many diseases have a latent phase at the beginning of the course of infection [25, p. 91 f.]. During this phase, individuals are infected but cannot contribute to the spread of infection since they are not yet infectious. A *SEIR* model takes this into account by introducing a fourth compartment E (*Exposed*) which follows the susceptible stage. The flowchart for such a model can be found in Fig. 2.2. Since exposed individuals are not yet infectious, they do not contribute to the number of transmissions and hence to the change in the Susceptible compartment. As in the Infected compartment, individuals leave the Exposed compartment at a constant rate which we will denote by $1/T_E$ where $T_E \in \mathbb{R}_{>0}$ is the mean latent period.

This yields the *SEIR* model:

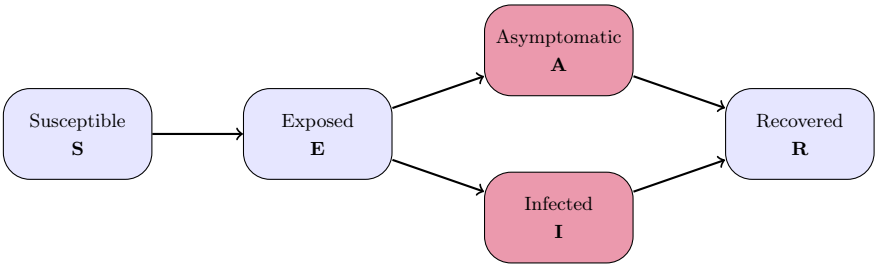
$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\phi\rho\frac{IS}{N}, \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \phi\rho\frac{IS}{N} - \frac{1}{T_E}E, \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{1}{T_E}E - \frac{1}{T_I}I, \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \frac{1}{T_I}I.
\end{aligned} \tag{2.2.1}$$

We can further extend this model by adding a compartment A (*Asymptomatic*) to account for asymptomatic cases. Individuals in this compartment are infected and also infectious, but have not (yet) developed symptoms. This allows to model, for instance,

that symptomatic individuals tend to isolate themselves more often than asymptomatic individuals, resulting in a different contribution to the spread of the disease. After this stage, individuals can either develop symptoms or recover directly. Therefore, this compartment is placed in between the Exposed and Infected compartments, and the flowchart of the model is shown in Fig. 2.3a.



(a) Schematic representation of the flows in a *SEAIR* model. After the latent period, individuals move to the Asymptomatic compartment. Some individuals develop symptoms, moving to the Infected compartments, others recover directly and move to the Recovered compartment.



(b) Schematic representation of the flows in a *SEAIR* model in [25, p. 92f.]. After the latent period, some individuals develop symptoms and move to the Infected compartments while the others move to the Asymptomatic compartment, following the asymptomatic course of disease.

Figure 2.3: **Schematic representations of the flows in different *SEAIR*-type models.** The compartments in which individuals are infectious and capable of transmitting the infection are the Asymptomatic and Infected compartments, which are highlighted in red.

Another but similar model is described in [25, p. 92f.], where the asymptomatic stage is not placed before the infected stage. Instead, a fraction of exposed individuals develop symptoms and all other individuals follow the asymptomatic course of disease. The flowchart for this model is presented in Fig. 2.3b.

Overall, there are no limits to the kind of extensions: In [17] the authors considered (partial) vaccinations using additional compartments. The authors of [7, p. 131-133] describe a model which introduces a compartment for individuals that receive treatment and recover faster.

To monitor hospitalized cases, we could extend the model to include a compartment H (*Hospitalized*) for severe cases that have to be treated in the hospital. Infected individuals could either recover or develop a severe course of the disease. From there, it could be further possible to develop a critical course that demands intensive care, which can

be, e.g., handled with a compartment U (for *in intensive care **u**nits*) afterward. As seen in the case of COVID-19, there are diseases which, in the worst case, result in death [44]. This could be modeled using a compartment D (*Dead*), from which there is no further transition possible. A model using all the compartments above as well as a compartment for undetected cases is presented in [16].

Every additional compartment increases the complexity of the model. The focus of this thesis, however, is the consideration of different spatial resolutions and mobility realizations, such that we stick here to the simpler epidemiological *SEIR* model (2.2.1).

2.3 Age stratification (Model A)

The extensions seen so far consist of additional compartments, reflecting different stages of the disease. As mentioned before, age can have a significant impact on infectious disease transmission and the model. For several diseases, recovery rates or transmission parameters are highly age-dependent [39, 4, 26]. Moreover, it strongly impacts contact rates since age groups do not mix heterogeneously [28]. To incorporate the age structure in our model, we divide the population into $J \in \mathbb{N}$ distinct age groups. Afterward, every individual is assigned to one of the compartments Susceptible, Exposed, Infected or Recovered. We denote the group of susceptible individuals of age group i with S_i for $i \in \{1, \dots, J\}$. For the other compartments, we proceed accordingly. The total size of age group i is denoted by $N_i = S_i + E_i + I_i + R_i$ for $i \in \{1, \dots, J\}$. Due to the short time horizon of days or weeks considered in our simulations, individuals are assumed to not change between age groups.

Analogously to the in Section 2.1 defined parameters, we now define the age-dependent equivalents. The contact rate depends on the age groups of both individuals involved in a contact. Therefore, the contact rate ϕ becomes a contact rate matrix $\phi = (\phi_{ij})_{i,j \in \{1,\dots,J\}}$, where the entry $\phi_{ij} \in \mathbb{R}_{\geq 0}$ gives the mean number of contacts an individual of age group i has with individuals from age group j for $i, j \in \{1, \dots, J\}$. Note that the contact rate matrix is not necessarily symmetric. We furthermore generalize the transmission probability $\rho_i \in [0, 1]$ for $i \in \{1, \dots, J\}$, now depending on the particular age group of the susceptible individuals, in order to model age-dependent susceptibility. The change in the susceptible subpopulation of one age group i due to an age group j per unit time is $\rho_i S_i \phi_{ij} I_j / N_j$ for $i, j \in \{1, \dots, J\}$. The total number of transmissions per unit time in age group i is then determined by the sum over all age groups. Moreover, the age-dependent, mean stay times in the Exposed and Infected compartments are denoted by $T_{E_i} \in \mathbb{R}_{>0}$ and $T_{I_i} \in \mathbb{R}_{>0}$, respectively. As individuals cannot change between age groups, individuals of an age group i leave the corresponding Exposed compartment at a constant rate $1/T_{E_i}$, entering the Infected compartment of the age group i. For the transition of I_i to R_i, the transition rate is $1/T_{I_i}$, accordingly.

This yields the age resolved *SEIR* model:

$$\begin{aligned}
\frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\rho_i S_i \sum_{j=1}^{J} \phi_{ij} \frac{I_j}{N_j}, \\
\frac{\mathrm{d}E_i}{\mathrm{d}t} &= \rho_i S_i \sum_{j=1}^{J} \phi_{ij} \frac{I_j}{N_j} - \frac{1}{T_{E_i}} E_i, \\
\frac{\mathrm{d}I_i}{\mathrm{d}t} &= \frac{1}{T_{E_i}} E_i - \frac{1}{T_{I_i}} I_i, \\
\frac{\mathrm{d}R_i}{\mathrm{d}t} &= \frac{1}{T_{I_i}} I_i,
\end{aligned} \tag{2.3.1}$$

for $i \in \{1, \ldots, J\}$. For later comparisons, we will call this model *Model A* with $J \in \mathbb{N}$ age groups and denote it by $\mathcal{M}_J^{\mathrm{A}}$.

2.3.1 Basic reproduction number for Model A

The basic reproduction number is an important characteristic of an infectious disease. It is defined as the expected number of secondary infections produced by a single infected individual in a totally susceptible population, and is usually denoted by $\mathcal{R}_0$ [7, p. 6f.]. Aside from its quantitative value, it is often used to consider whether disease dynamics are expected to take up or to go down. For deterministic models, the threshold value 1 yields if a disease dies out or if an outbreak occurs. In simple compartmental models, it can be obtained by observing the number of infections in unit time of a single infective individual and multiplying it by the mean time spent in the infective compartments. In more complex models which contain sub-compartments with different susceptibility, this de facto linearization for the calculation of this number requires a more general approach. In this chapter, we derive the basic reproduction number for Model A with $J \in \mathbb{N}$ age groups (2.3.1) to present an approach we can later use for the basic reproduction number of the spatially resolved ODE models. The generalized approach consists of deriving the *next generation matrix*, whose (k, l) entry yields the number of secondary infections in compartment k caused by an individual of compartment l [7, p. 182]. The basic reproduction number is then given by the spectral radius of this matrix [38].

Before considering the *SEIR* model, we define the necessary quantities based on a generic compartmental model. We assume to have a model with $n \in \mathbb{N}$ compartments, where $x(t) = (x_1(t), \ldots, x_n(t))$ denotes the model state at time $t \in \mathbb{R}_{\geq 0}$. Here, $x_k(t) \geq 0$ denotes the number of individuals in compartment k. In the following, we will often drop t from the notation and simply write x_k for the number of individuals in this compartment k. To construct the next generation matrix, we distinguish between two kinds of compartments: infected and non-infected compartments. Individuals in an infected compartment do not necessarily have to be infectious, but they currently carry the pathogen

and are not yet recovered. For instance, the *SEIR* model has the infected compartments Exposed and Infected, although individuals in the Exposed compartment cannot contribute to the transmission of the pathogen directly. We denote the subset of indices corresponding to infected compartments with $\mathcal{J} \subseteq \{1, \dots, n\}$. Furthermore, we define the set of all disease-free states by $X_s = \{x \in \mathbb{R}^n_{\geq 0} \mid x_k = 0 \; \forall k \in \mathcal{J}\}$.

Next, we distinguish between inflows from a non-infected compartment to an infected compartment $k \in \mathcal{J}$ which we denote by $\mathcal{F}_k$, and all other inflows, denoted by $\mathcal{V}_k^+$. An outflow is denoted by $\mathcal{V}_k^-$ and we define $\mathcal{V}_k := \mathcal{V}_k^- - \mathcal{V}_k^+$. Now, we can write the system of differential equations as:

$$x_k'(t) = \mathcal{F}_k(t) - \mathcal{V}_k(t), \qquad k = 1, \dots, n. \tag{2.3.2}$$

In Lemma 1 of [38] two matrices $F(x^*), V(x^*) \in \mathbb{R}^{m \times m}$ where $m = |\mathcal{J}|$, are defined by

$$F(x^*) := \left(\frac{\partial \mathcal{F}_k}{\partial x_l}(x^*) \right)_{k,l \in \mathcal{J}} \quad \text{and} \quad V(x^*) := \left(\frac{\partial \mathcal{V}_k}{\partial x_l}(x^*) \right)_{k,l \in \mathcal{J}}, \tag{2.3.3}$$

for a disease-free equilibrium (DFE) $x^* \in X_s$.

The next generation matrix is then defined by $F(x^*)V(x^*)^{-1}$. The effective reproduction number is obtained by $\mathcal{R}_{\mathit{eff}} = \rho\left(FV^{-1}\right)$, where $\rho(\cdot)$ is the spectral radius. To obtain the basic reproduction number, we chose $x_0^* \in X_s$ to be the DFE of a fully susceptible population. We refer to [38] for the proof and further information concerning the derivation of this approach. In the following, we use this theory to compute the basic reproduction numbers for several ODE models.

We apply this approach to Model A (2.3.1) and compute the matrices $F(x^*)$ and $V(x^*)$, for a DFE $x^* \in X_s$. To accomplish this, we sort the compartments as follows:

$$S_1, \dots, S_J, E_1, \dots, E_J, I_1, \dots, I_J, R_1, \dots, R_J. \tag{2.3.4}$$

Note that the infected compartments are E_i and I_i for all $i \in \{1, \dots, J\}$. Therefore, the set of corresponding indices is

$$\mathcal{J} = \{k \mid x_k \in \{E_i, I_i \mid i = 1, \dots, J\}\}. \tag{2.3.5}$$

For a model state

$$x = (S_1, \dots, S_J, E_1, \dots, E_J, I_1, \dots, I_J, R_1, \dots, R_J), \tag{2.3.6}$$

we observe:

$$\mathcal{F}_k(x) = \begin{cases} 0, & x_k = S_i, \\ \rho_i S_i \sum_{j=1}^J \phi_{ij} \frac{I_j}{N_j}, & x_k = E_i, \\ 0, & x_k = I_i, \\ 0, & x_k = R_i, \end{cases} \quad \text{and} \quad \mathcal{V}_k(x) = \begin{cases} \rho_i S_i \sum_{j=1}^J \phi_{ij} \frac{I_j}{N_j}, & x_k = S_i, \\ \frac{1}{T_{E_i}} E_i, & x_k = E_i, \\ \frac{1}{T_{I_i}} I_i - \frac{1}{T_{E_i}} E_i, & x_k = I_i, \\ -\frac{1}{T_{I_i}} I_i, & x_k = R_i. \end{cases} \tag{2.3.7}$$

Let $x^* \in X_s$ be a DFE of (2.2.1). Then x^* is of the form

$$x^* = (S_0, \dots, S_J, 0, \dots, 0, 0, \dots, 0, N_0 - S_0, \dots, N_J - S_J). \tag{2.3.8}$$

To construct the matrices $F(x^*)$ and $V(x^*)$ we need to differentiate $\mathcal{F}_k(x)$ and $\mathcal{V}_k(x)$ with respect to x_l, where $k, l \in \mathcal{J}$, i.e., $x_k, x_l \in \{E_1, \dots, E_J, I_1, \dots, I_J\}$. We have:

$$f_{k,l}(x^*) = \frac{\partial \mathcal{F}_k}{\partial x_l}(x^*) = \begin{cases} 0, & x_k = E_i, x_l = E_j, \\ \rho_i S_i \phi_{ij} \frac{1}{N_j}, & x_k = E_i, x_l = I_j, \\ 0, & x_k = I_i, x_l = E_j, \\ 0, & x_k = I_i, x_l = I_j, \end{cases} \tag{2.3.9}$$

for the entries of $F(x^*)$. Consequently, $F(x^*)$ has the form

$$F(x^*) = \left(\begin{array}{c|ccc} \multirow{3}{*}{$\mathbf{0}$} & \rho_1 S_1 \phi_{11} \frac{1}{N_1} & \rho_1 S_1 \phi_{12} \frac{1}{N_2} & \cdots & \rho_1 S_1 \phi_{1J} \frac{1}{N_J} \\ & \rho_2 S_2 \phi_{21} \frac{1}{N_1} & \rho_2 S_2 \phi_{22} \frac{1}{N_2} & \cdots & \rho_2 S_2 \phi_{2J} \frac{1}{N_J} \\ & \vdots & \vdots & \ddots & \vdots \\ & \rho_J S_J \phi_{J1} \frac{1}{N_1} & \rho_J S_J \phi_{J2} \frac{1}{N_2} & \cdots & \rho_J S_J \phi_{JJ} \frac{1}{N_J} \\ \hline \mathbf{0} & \multicolumn{3}{c}{\mathbf{0}} \end{array}\right). \tag{2.3.10}$$

For $V(x^*)$, we derive:

$$v_{k,l}(x^*) = \frac{\partial \mathcal{V}_k}{\partial x_l}(x^*) = \begin{cases} \frac{1}{T_{E_j}} \delta_{ij}, & x_k = E_i, x_l = E_j, \\ 0, & x_k = E_i, x_l = I_j, \\ -\frac{1}{T_{E_j}} \delta_{ij}, & x_k = I_i, x_l = E_j, \\ \frac{1}{T_{I_j}} \delta_{ij}, & x_k = I_i, x_l = E_j, \end{cases} \tag{2.3.11}$$

where δ_{ij} is the Kronecker delta

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{2.3.12}$$

In matrix form, we get:

$$V(x^*) = \left(\begin{array}{cccc|cccc} \frac{1}{T_{E_1}} & 0 & \cdots & 0 & & & & \\ 0 & \frac{1}{T_{E_2}} & & \vdots & & & & \\ \vdots & & \ddots & 0 & & \mathbf{0} & & \\ 0 & \cdots & 0 & \frac{1}{T_{E_J}} & & & & \\ \hline -\frac{1}{T_{E_1}} & 0 & \cdots & 0 & \frac{1}{T_{I_1}} & 0 & \cdots & 0 \\ 0 & -\frac{1}{T_{E_2}} & & \vdots & 0 & \frac{1}{T_{I_2}} & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & -\frac{1}{T_{E_J}} & 0 & \cdots & 0 & \frac{1}{T_{I_J}} \end{array}\right). \tag{2.3.13}$$

We need to invert this matrix and then compute $F(x^*)V(x^*)^{-1}$. The inverse of a block matrix B with a zero upper right block, i.e.,

$$B = \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} \tag{2.3.14}$$

with invertible block matrices $A \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{m\times m}$, is

$$B^{-1} = \begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix}, \tag{2.3.15}$$

since

$$\begin{aligned} BB^{-1} &= \begin{pmatrix} A & 0 \\ C & D \end{pmatrix}\begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix} = \begin{pmatrix} \mathrm{Id}_n & 0 \\ CA^{-1} - CA^{-1} & \mathrm{Id}_m \end{pmatrix} = \mathrm{Id}_{n+m}, \\ B^{-1}B &= \begin{pmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{pmatrix}\begin{pmatrix} A & 0 \\ C & D \end{pmatrix} = \begin{pmatrix} \mathrm{Id_n} & 0 \\ -D^{-1}C + D^{-1}C & \mathrm{Id}_m \end{pmatrix} = \mathrm{Id}_{n+m}. \end{aligned} \tag{2.3.16}$$

Here, Id_d is the identity matrix of $\mathbb{R}^{d\times d}$ for $d \in \{n,\ m,\ n+m\}$.

As the blocks of $V(x^*)$ are diagonal matrices, the inverse of each block is easy to calculate. Using (2.3.15), we obtain the inverse of $V(x^*)$:

$$V(x^*)^{-1} = \left(\begin{array}{cccc|cccc} T_{E_1} & 0 & \cdots & 0 & & & & \\ 0 & T_{E_2} & & \vdots & & & & \\ \vdots & & \ddots & 0 & & \mathbf{0} & & \\ 0 & \cdots & 0 & T_{E_J} & & & & \\ \hline T_{I_1} & 0 & \cdots & 0 & T_{I_1} & 0 & \cdots & 0 \\ 0 & T_{I_2} & & \vdots & 0 & T_{I_2} & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & T_{I_J} & 0 & \cdots & 0 & T_{I_J} \end{array}\right). \tag{2.3.17}$$

Multiplying both matrices $F(x^*)$ and $V(x^*)^{-1}$ yields:

$$F(x^*)V(x^*)^{-1} = \left(\begin{array}{c|c} A(x^*) & A(x^*) \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \tag{2.3.18}$$

where

$$A(x^*) = \begin{pmatrix} \rho_1 S_1 \phi_{11} \frac{T_{I_1}}{N_1} & \rho_1 S_1 \phi_{12} \frac{T_{I_2}}{N_2} & \cdots & \rho_1 S_1 \phi_{1J} \frac{T_{I_J}}{N_J} \\ \rho_2 S_2 \phi_{21} \frac{T_{I_1}}{N_1} & \rho_2 S_2 \phi_{22} \frac{T_{I_2}}{N_2} & \cdots & \rho_2 S_2 \phi_{2J} \frac{T_{I_J}}{N_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_J S_J \phi_{J1} \frac{T_{I_1}}{N_1} & \rho_J S_J \phi_{J2} \frac{T_{I_2}}{N_2} & \cdots & \rho_J S_J \phi_{JJ} \frac{T_{I_J}}{N_J} \end{pmatrix}. \tag{2.3.19}$$

We need to compute the spectral radius, i.e., the largest eigenvalue in absolute value of this matrix. From [29, Theorem 2.1], we conclude that $\rho(F(x^*)V(x^*)^{-1}) = \rho(A(x^*))$ as the eigenvalues of the lower right block of $F(x^*)V(x^*)^{-1})$ are 0. For a single age group, i.e., $J = 1$, a DFE has the form $x^* = (S, 0, 0, N - S) \in X_s$ and we obtain the next generation matrix

$$F(x^*)V(x^*)^{-1} = \begin{pmatrix} \rho S \phi \frac{T_I}{N} & \rho S \phi \frac{T_I}{N} \\ 0 & 0 \end{pmatrix}, \tag{2.3.20}$$

with the eigenvalues $\lambda_1 = 0$ and $\lambda_2 = \rho S \phi \frac{T_I}{N}$. We therefore obtain the effective reproduction number for the *SEIR* model without age-resolution $\mathcal{M}_1^{\mathrm{A}}$ (2.2.1)

$$\mathcal{R}_{\mathit{eff}} = \rho S \phi \frac{T_I}{N}. \tag{2.3.21}$$

To obtain the basic reproduction number, we chose the DFE $x_0^* = (N, 0, 0, 0)$ and conclude

$$\mathcal{R}_0 = \rho N \phi \frac{T_I}{N} = \rho \phi T_I. \tag{2.3.22}$$

With increasing number of age groups $J \in \mathbb{N}$, the analytical solution of the eigenvalues becomes more and more complex, and we do not give these solutions here. For $J > 1$, we will compute these values numerically. However, for a totally susceptible population $x_0^* = (N_1, \ldots, N_J, 0, \ldots, 0)$ we observe that $A(x^*)$ simplifies to

$$A(x^*) = \begin{pmatrix} \rho_1 \phi_{11} T_{I_1} & \rho_1 \phi_{12} T_{I_2} & \cdots & \rho_1 \phi_{1J} T_{I_J} \\ \rho_2 \phi_{21} T_{I_1} & \rho_2 \phi_{22} T_{I_2} & \cdots & \rho_2 \phi_{2J} T_{I_J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_J \phi_{J1} T_{I_1} & \rho_J \phi_{J2} T_{I_2} & \cdots & \rho_J \phi_{JJ} T_{I_J} \end{pmatrix}, \tag{2.3.23}$$

i.e., its entries and hence the basic reproduction number is independent of the population size.

3 Metapopulation models for infectious disease dynamics

One assumption of the models presented so far is that the population is "well-mixed", i.e., all compartments mix homogeneously. This highly simplifying assumption is generally incorrect on all scales. The larger the scale, the higher the number of individuals with which a particular individual does not have any contact at all. Models that account for this heterogeneity are therefore more accurate and also more useful for evaluating control strategies which affect the movement of individuals [25].

There are essentially two forms of incorporating the spatial component - the discrete and the continuous form. For the continuous form, e.g., through the use of the diffusion equation, we refer to [7, Ch. 14.2] or [25, Ch. 15.3]. We will focus on models that use a discrete spatial discretization, which are called *metapopulation models*. Metapopulation models divide the total population into groups of subpopulations that interact at some level. As the total population is then made up of several spatially distributed subpopulations, we call the total population a metapopulation. Each subpopulation lives in a different area, which will be denoted a *patch* or *region*. Depending on the modeling scale, a patch can be a city, a county, or a country. Naturally existing mobility between these patches is then considered as links between the patches. This could be some train connections between cities or air traveling between countries.

We will only consider short-term migration, which means that individuals travel to another patch for some short time and then return. As we are considering short-term epidemic models, we will neglect the fact of long-term migration, where individuals move to another patch and settle there. In the following sections, we present two approaches for modeling the spatial spread in metapopulation models.

3.1 Spatial resolution and mobility in ODE models

We start by reformulating the ODE system (2.2.1) in order to account for the changed dynamics due to commuting and spatial heterogeneity. We follow the idea of [40] and extend their approach to an *SEIR* model, similar to [23]. To put the focus on the spatial stratification and for the sake of a simple presentation, we will drop the age stratification derived in Section 2.3 in the first steps and reintroduce it later.

3.1.1 A first metapopulation model (Model B)

We will now derive a first ODE-based metapopulation model motivated by [40] and [23]. The following explanations differ from these papers in several aspects. Foremost, both papers use a different assumption about transmission of infection compared to us. As described in Section 2.1, there are mainly the mass action with coefficient $\widehat{\beta}$ and the standard incidence assumption with coefficient β/N. They are very similar, but the standard incidence is normalized by the population size [25, p. 38]. We recall that $\widehat{\beta}$ implicitly also depends on the population size, so for every patch, we will obtain a different coefficient. However, the models of [40] and [23] assume independence of the patch. To avoid confusion, we use the standard incidence and consider patch sizes in the equations.

Additionally, the authors of [23] assume that exposed individuals are infectious and infectious individuals are quarantined immediately after being confirmed as infected. We will use the definition of exposed individuals being not yet infectious and assume that there is no quarantining. Therefore, every E in their equations will be exchanged by an I in our equations. Moreover, the second focus of [23], in which the model is extended to consider infections during commuting, is omitted for reasons of simplicity.

We divide the population under study into $M \in \mathbb{N}$ subpopulations which are spatially separated from each other. The size of the subpopulation at patch $\mathcal{P}_n$ will be denoted by P_n for each $n \in \{1, \dots, M\}$. In each patch, the subpopulation is assumed to be "well-mixed", i.e., when considering only a single patch, everybody has the same probability of contacting each other. The spread of the disease is modeled with an *SEIR* model, so each subpopulation is further divided into the epidemiological groups Susceptible, Exposed, Infected and Recovered. We denote the number of susceptible individuals at a patch $\mathcal{P}_n$ by S_n for all $n \in \{1, \dots, M\}$ and proceed analogously for the other compartments. The total size of the subpopulation at patch $\mathcal{P}_n$ at a time $t \in \mathbb{R}_{\geq 0}$ satisfies

$$P_n(t) = S_n(t) + E_n(t) + I_n(t) + R_n(t). \tag{3.1.1}$$

Two distinct patches $\mathcal{P}_n$ and $\mathcal{P}_m$ are connected by interaction strengths $h_{nm} \in \mathbb{R}_{\geq 0}$ and $h_{mn} \in \mathbb{R}_{\geq 0}$ for $n, m \in \{1, \dots, M\},\ n \neq m$. The interaction strengths are derived by

$$h_{nm} = \frac{H_{nm}}{P_n}, \tag{3.1.2}$$

where $H_{nm} \in \mathbb{R}_{\geq 0}$ is the average number of commuters from patch $\mathcal{P}_n$ to $\mathcal{P}_m$ per day. Therefore, h_{nm} is the share of individuals from patch $\mathcal{P}_n$ that travels to $\mathcal{P}_m$. We do not assume that mobility between patches is symmetric, i.e., we allow $h_{nm} \neq h_{mn}$ for two patches $\mathcal{P}_n$ and $\mathcal{P}_m$ with $n \neq m$.

As commuting is not performed explicitly, P_n remains constant throughout the simulation. Instead, infectious individuals from a patch$\mathcal{P}_n$ contribute to the spread of the

disease by considering their theoretical impact in other patches $\mathcal{P}_m$, which is proportional to either h_{nm} or h_{mn}. This theoretical impact will be concretized into explicit formulas in the following.

As time progresses, infections can occur from the following three sources. Consider two patches $\mathcal{P}_n$ and $\mathcal{P}_m$ with $n, m \in \{1, \dots, M\},\ n \neq m$ and an individual originating from $\mathcal{P}_n$. This individual can make contacts with infectious individuals

I. from patch $\mathcal{P}_n$, if it stays at this location,

II. from patch $\mathcal{P}_m$ that have commuted to patch $\mathcal{P}_n$, if it stays at this location or

III. from patch $\mathcal{P}_m$, if it travels from $\mathcal{P}_n$ to $\mathcal{P}_m$.

The first source corresponds to the case where every susceptible individual of the patch $\mathcal{P}_n$ can potentially meet infectious individuals from the same patch. Here, we apply the basic *SEIR* model (2.2.1). Note that N is exchanged by P_n, as this is the number of individuals at this location. For the first source, we obtain

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{I})} = -\rho\phi\frac{S_n I_n}{P_n} \tag{3.1.3}$$

transmissions in unit time.

Source II corresponds to the case where infectious individuals originating from patch $\mathcal{P}_m$ travel to patch $\mathcal{P}_n$. Since h_{mn} is the share of individuals traveling from $\mathcal{P}_m$ to $\mathcal{P}_n$, $h_{mn}I_m$ is the proportion of infectious individuals from patch $\mathcal{P}_m$ which commute to patch $\mathcal{P}_n$, to eventually spread the infection. Accordingly, for a susceptible individual from patch $\mathcal{P}_n$, the chance to meet an infectious individual from $\mathcal{P}_m$ at patch $\mathcal{P}_n$ is $h_{mn}I_m/P_n$. Therefore, Source II results in

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{II})} = -\rho\phi S_n \sum_{m\neq n} h_{mn}\frac{I_m}{P_n} \tag{3.1.4}$$

transmissions in unit time in the population originating from patch $\mathcal{P}_n$.

Similarly, $h_{nm}S_n$ is the proportion of susceptible individuals from patch $\mathcal{P}_n$ which commute to patch $\mathcal{P}_m$ where they can be exposed to infectious individuals at patch $\mathcal{P}_m$. At patch $\mathcal{P}_m$, the chance to meet an infectious individual from that patch is I_m/P_m. This means we have

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{III})} = -\rho\phi S_n \sum_{m\neq n} h_{nm}\frac{I_m}{P_m} \tag{3.1.5}$$

transmissions in unit time due to Source III.

Combining all these terms on new transmissions, via

$$\frac{\mathrm{d}S_n}{\mathrm{d}t} = \frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{I})} + \frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{II})} + \frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{III})}, \tag{3.1.6}$$

we get

$$\begin{aligned}\frac{\mathrm{d}S_n}{\mathrm{d}t} &= -\rho\phi\frac{S_n I_n}{P_n} - \rho\phi S_n \sum_{m\neq n} h_{mn}\frac{I_m}{P_n} - \rho\phi S_n \sum_{m\neq n} h_{nm}\frac{I_m}{P_m} \\ &= -\rho\phi S_n \frac{I_n}{P_n} - \rho\phi S_n \sum_{m\neq n}\left(\frac{h_{mn}I_m}{P_n} + \frac{h_{nm}I_m}{P_m}\right) \end{aligned} \tag{3.1.7}$$

transmissions per unit time in total. We define $h_{nn} = 1/2$ for reasons of convenience and get

$$\frac{\mathrm{d}S_n}{\mathrm{d}t} = -\rho\phi S_n \sum_{m=1}^{M}\left(\frac{h_{mn}}{P_n} + \frac{h_{nm}}{P_m}\right) I_m. \tag{3.1.8}$$

The transitions from Exposed to Infected and from Infected to Recovered do not change compared to the standard *SEIR* model (2.2.1). As before, individuals leave the Exposed and Infected compartment at a constant rate. We assume that this rate is independent of the patch and obtain:

$$\begin{aligned}\frac{\mathrm{d}S_n}{\mathrm{d}t} &= -\rho\phi S_n \sum_{m=1}^{M}\left(\frac{h_{mn}}{P_n} + \frac{h_{nm}}{P_m}\right) I_m, \\ \frac{\mathrm{d}E_n}{\mathrm{d}t} &= \rho\phi S_n \sum_{m=1}^{M}\left(\frac{h_{mn}}{P_n} + \frac{h_{nm}}{P_m}\right) I_m - \frac{1}{T_E}E_n, \\ \frac{\mathrm{d}I_n}{\mathrm{d}t} &= \frac{1}{T_E}E_n - \frac{1}{T_I}I_n, \\ \frac{\mathrm{d}R_n}{\mathrm{d}t} &= \frac{1}{T_I}I_n, \end{aligned} \tag{3.1.9}$$

for all $n \in \{1, \ldots, M\}$.

Except for the different transmission assumption and the additional Exposed compartment, this model corresponds to the model of [40]. Note that they use h_{nm} for the average volume of commuters in unit time between patch $\mathcal{P}_n$ and $\mathcal{P}_m$, which corresponds to H_{nm} in this thesis. We will refer to the model (3.1.9) as *Model B* and denote it by $\mathcal{M}^{\mathrm{B}}_{1,M}$, for $M \in \mathbb{N}$ patches and a single age group.

There are some imprecisions in this model. First, the model describes the scenario of Source II as the fraction of infectious individuals from other patches traveling to patch $\mathcal{P}_n$ where they can potentially make contact with any susceptible individual originating from this patch. At the same time, Source III corresponds to the scenario of a fraction of susceptible individuals from patch $\mathcal{P}_n$ traveling to other patches, where they can potentially meet every infectious individual originating from that patch. Summing up all sources means that infectious individuals contribute to the spread of the disease at different locations at the same time, which is not correct. Also, we scaled the probability

of contacting another individual by the total originating population, where some individuals may have left the patch $\mathcal{P}_n$ and others may have entered due to commuting. As we did not assume the commuting strengths to be symmetric, it is not accurate to assume that the number of individuals which are at the patch after out-commuters have left and in-commuters have entered, is the same as the number of individuals originating from there. Moreover, at their target patch, the susceptible individuals should be able to catch the pathogen not only from the infectious individuals originating from that patch, but also from infectious individuals which are at the patch after commuting. In the following, we want to revise these imprecisions and present an updated model.

3.1.2 A second model including correct normalization and in and out of home patch times (Model C)

We start with the same setting of patches and their subpopulations, and divide each subpopulation again into the four compartments of the *SEIR* model. As before, these patches are connected through the mean numbers of commuters H_{nm} and the corresponding commuting strengths h_{nm} (3.1.2). Before we study the infection dynamics between the patches, we have to make some more observations about the populations. It is important to see that there is a difference between the number of individuals originating from a certain patch and the number of individuals which actually are at a certain patch. We consider a fixed patch $\mathcal{P}_n$ with a subpopulation of size $P_n = S_n + E_n + I_n + R_n$ as before. Both P_n and the sizes of compartments are the numbers of individuals before commuting. Now after commuting, the number of individuals at patch $\mathcal{P}_n$, which we denote by N_n, is reduced by the number of out-commuters and increased by the number of in-commuters to patch $\mathcal{P}_n$:

$$N_n = P_n - \sum_{m \neq n} H_{nm} + \sum_{m \neq n} H_{mn}. \tag{3.1.10}$$

Again, commuting is not performed explicitly. Thus, the distinction between before and after commuting is purely synthetic, introduced for the sake of enhanced comprehension, and N_n remains constant.

We apply the basic *SEIR* model (2.2.1) to the cases resulting from the same three sources I, II, III as before. In order to address the aforementioned shortcomings, we determine the respective population sizes at each patch, considering the changes in populations due to commuting. The number of susceptible individuals at a patch $\mathcal{P}_n$ for $n \in \{1, \ldots, M\}$ differs from S_n by the number of individuals who have left the patch for commuting, as well as those who have entered through commuting.

For Source I, we have to determine the number of susceptible individuals which originate from patch $\mathcal{P}_n$ and stay at their home patch. We, thus, define the share of individuals

staying at patch $\mathcal{P}_n$ by:

$$h_{nn} := \left(1 - \sum_{m \neq n} h_{nm}\right). \tag{3.1.11}$$

After commuting, the number of susceptible individuals remaining at patch $\mathcal{P}_n$ is

$$S_n^r = S_n h_{nn}. \tag{3.1.12}$$

We define the number of remaining infected individuals I_n^r respectively.

Note that the probability of meeting an infected individual at the respective patch is scaled by N_n, as this is the number of individuals which are actually at this patch, while P_n represents the number of individuals originating in $\mathcal{P}_n$. While N_n does not include the number of individuals which commute to a different patch, it includes that number of individuals which originate from another patch and which commute to patch $\mathcal{P}_n$. Thus, the number of transmissions due to Source I in unit time is

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{I})} = -\rho\phi\frac{S_n^r I_n^r}{N_n}. \tag{3.1.13}$$

Source II corresponds to the case where infectious individuals originating from a different patch $\mathcal{P}_m$ with $m \in \{1, \ldots, M\}$, $m \neq n$ travel to patch $\mathcal{P}_n$. The number of infectious commuters arriving at the destination $\mathcal{P}_n$ is defined by

$$I_n^c = \sum_{m \neq n} h_{mn} I_m. \tag{3.1.14}$$

There, they can spread the infection among all those susceptible individuals which stayed in $\mathcal{P}_n$, i.e., S_n^r. This leads to

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{II})} = -\rho\phi\frac{S_n^r I_n^c}{N_n} \tag{3.1.15}$$

transmissions in unit time due to Source II.

The total number of infectious individuals at a certain patch $\mathcal{P}_m$ for $m \in \{1, \ldots, M\}$ is obtained through the number of infectious individuals originating from $\mathcal{P}_m$ and infectious individuals having commuted from $\mathcal{P}_l$ to $\mathcal{P}_m$, for $l \in \{1, \ldots, M\}, l \neq m$:

$$\begin{aligned} I_m^{tot} &= I_m^r + I_m^c \\ &= h_{mm} I_m + \sum_{l \neq m} h_{lm} I_l \\ &= \sum_{l=1}^{M} h_{lm} I_l. \end{aligned} \tag{3.1.16}$$

For simplicity, here, we assume that infectious individuals travel with the same rate as susceptible individuals, but this assumption could be easily generalized by introducing a corresponding reduction term.

In the case of Source III, individuals from patch $\mathcal{P}_n$ travel to patch $\mathcal{P}_m$, where the number of infectious individuals is I_m^{tot}. As out-commuting individuals then stay for a certain time in patch $\mathcal{P}_m$, their chance of meeting an infected individual has to be normalized by N_m instead of N_n. This yields

$$\frac{\mathrm{d}S_n}{\mathrm{d}t}^{(\mathrm{III})} = -\rho\phi S_n \sum_{m\neq n} \frac{h_{nm} I_m^{tot}}{N_m} \tag{3.1.17}$$

transmissions in unit time due to Source III.

We now merge the new transmissions from the three different sources into one single ODE using (3.1.6):

$$\begin{aligned}
\frac{\mathrm{d}S_n}{\mathrm{d}t} &= -\frac{\rho\phi}{N_n} S_n^r I_n^r - \frac{\rho\phi}{N_n} S_n^r I_n^c - \rho\phi S_n \sum_{m\neq n} \frac{h_{nm} I_m^{tot}}{N_m} \\
&= -\frac{\rho\phi}{N_n} S_n^r I_n^{tot} - \rho\phi S_n \sum_{m\neq n} \frac{h_{nm} I_m^{tot}}{N_m} \\
&= -\rho\phi S_n \left(\frac{1}{N_n} h_{nn} I_n^{tot} + \sum_{m\neq n} \frac{h_{nm} I_m^{tot}}{N_m} \right) \\
&= -\rho\phi S_n \sum_{m=1}^{M} \frac{h_{nm} I_m^{tot}}{N_m}.
\end{aligned} \tag{3.1.18}$$

The aim of this thesis is to compare pure equation-based models such as (3.1.18) to a semi-discrete hybrid graph-ODE-based model, which is presented in the next section. For that model, we will assume that individuals stay at home for half a day and, in case of commuting or traveling, are out of the home patch for the other half of the day. In the model above, this is not reflected, as individuals are considered to have out of home patch contacts for the whole simulation. To better reflect this discrete real-world effect, we scale equation (3.1.18) by 0.5 and add equation (2.1.2) applied to each patch when neglecting commuting, scaled by 0.5 as well.

We then obtain

$$\frac{\mathrm{d}S_n}{\mathrm{d}t} = - \underbrace{\frac{1}{2}\rho\phi \frac{S_n I_n}{P_n}}_{\text{No commuting, } \textit{SEIR} \text{ model}} \underbrace{- \frac{1}{2}\rho\phi S_n \sum_{m=1}^{M} \frac{h_{nm} I_m^{tot}}{N_m}}_{\text{Commuting}}. \tag{3.1.19}$$

All of these statements do not affect the equations for the other compartments, which is why the rest of the model is similar to model (3.1.9):

$$
\begin{aligned}
\frac{\mathrm{d}S_n}{\mathrm{d}t} &= -\frac{1}{2}\rho\phi S_n\left(\frac{I_n}{P_n} + \sum_{m=1}^{M}\frac{h_{nm}I_m^{tot}}{N_m}\right),\\
\frac{\mathrm{d}E_n}{\mathrm{d}t} &= \frac{1}{2}\rho\phi S_n\left(\frac{I_n}{P_n} + \sum_{m=1}^{M}\frac{h_{nm}I_m^{tot}}{N_m}\right) - \frac{1}{T_E}E_n,\\
\frac{\mathrm{d}I_n}{\mathrm{d}t} &= \frac{1}{T_E}E_n - \frac{1}{T_I}I_n,\\
\frac{\mathrm{d}R_n}{\mathrm{d}t} &= \frac{1}{T_I}I_n,
\end{aligned}
\tag{3.1.20}
$$

for all $n \in \{1, \dots, M\}$. We call this model *Model C* and denote it by $\mathcal{M}^{\mathrm{C}}_{1,M}$ for $M \in \mathbb{N}$ patches and a single age group.

We will now extend Model C (3.1.20) by an age resolution. In the numerical experiments, we will then also consider how the stratification by age affects the runtime in the different models. To derive the age-resolved equations of Model C, we proceed analogously to Section 2.3. For every patch $\mathcal{P}_n$, $n \in \{1, \dots, M\}$, we divide its subpopulation into $J \in \mathbb{N}$ age groups. Every compartment is now indicated by two indices i and n where $i = 1, \dots, J$ enumerates the age groups and $n = 1, \dots, M$ enumerates the patches. The number of susceptible individuals originating from patch $\mathcal{P}_n$ and belonging to age group i is then given by $S_{i,n}$. We define $E_{i,n}$, $I_{i,n}$ and $R_{i,n}$ accordingly. Also, we observe that

$$
P_{i,n} = S_{i,n} + E_{i,n} + I_{i,n} + R_{i,n} \tag{3.1.21}
$$

is the number of individuals which belong to the age group i and originate from patch $\mathcal{P}_n$. For simplicity and due to a lack of more resolved data, we assume the mobility to be independent of age. As a result, we cannot use equation (3.1.10) to derive the number of individuals which are actually at a patch and belong to a specific age group. Thus, we compute the commuting strengths h_{nm} as

$$
h_{nm} = \frac{H_{nm}}{\sum_{i=1}^{J} P_{i,n}}, \tag{3.1.22}
$$

and use

$$
N_{i,n} = P_{i,n} - \sum_{m \neq n} h_{nm}P_{i,n} + \sum_{m \neq n} h_{mn}P_{i,m} \tag{3.1.23}
$$

to obtain the total number of individuals in age group i at patch $\mathcal{P}_n$ after commuting has taken place. Note that with more resolved data, a model generalization is easily possible. Lastly, we need to set the total number of infected individuals at a patch $\mathcal{P}_m$ which belong to age group i, using (3.1.16):

$$
I_{i,m}^{tot} = \sum_{l=1}^{M} h_{lm}I_{i,l}. \tag{3.1.24}
$$

Using these refinements, we can set up the age-resolved metapopulation model:

$$\begin{aligned}
\frac{\mathrm{d}S_{i,n}}{\mathrm{d}t} &= -\frac{1}{2}\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \left(\frac{I_{j,n}}{P_{j,n}} + \sum_{m=1}^{M} \frac{h_{nm} I_{j,m}^{tot}}{N_{j,m}} \right), \\
\frac{\mathrm{d}E_{i,n}}{\mathrm{d}t} &= \frac{1}{2}\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \left(\frac{I_{j,n}}{P_{j,n}} + \sum_{m=1}^{M} \frac{h_{nm} I_{j,m}^{tot}}{N_{j,m}} \right) - \frac{1}{T_{E_i}} E_{i,n}, \\
\frac{\mathrm{d}I_{i,n}}{\mathrm{d}t} &= \frac{1}{T_{E_i}} E_{i,n} - \frac{1}{T_{I_i}} I_{i,n}, \\
\frac{\mathrm{d}R_{i,n}}{\mathrm{d}t} &= \frac{1}{T_{I_i}} I_{i,n},
\end{aligned} \tag{3.1.25}$$

for all $i \in \{1, \dots, J\}$ and $n \in \{1, \dots, M\}$. As this is the age-resolved version of Model C we denote model (3.1.25) by $\mathcal{M}_{J,M}^{\mathrm{C}}$.

In both models Model B (3.1.9) and Model C (3.1.25), the epidemiological parameters and the contact patterns do not depend on the patches. This could be easily resolved by exchanging the contact matrix ϕ by a region-specific contact matrix ϕ_n, which would even allow modeling different non-pharmaceutical interventions in different patches. The same way, we could implement location-specific parameters governing the transmission of the disease or the recovery, for example if some locations have better medical care than others. Since we will only study dynamics in a part of Germany, we will assume both the contact patterns and the epidemiological parameters are independent of the patch.

3.1.3 Basic reproduction numbers for Model B and Model C

We now provide the basic reproduction numbers for an age-resolved version of Model B (3.1.9) and Model C (3.1.25). Here, we make use of the approach based on the next generation matrix we have already seen for Model A in Section 2.3.

To compute the basic reproduction number for the age-resolved version of Model B (3.1.9), we use the age-resolved model:

$$\begin{aligned}
\frac{\mathrm{d}S_{i,n}}{\mathrm{d}t} &= -\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \sum_{m=1}^{M} \left(\frac{h_{nm}}{P_{j,m}} + \frac{h_{mn}}{P_{j,n}} \right) I_{j,m}, \\
\frac{\mathrm{d}E_{i,n}}{\mathrm{d}t} &= \rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \sum_{m=1}^{M} \left(\frac{h_{nm}}{P_{j,m}} + \frac{h_{mn}}{P_{j,n}} \right) I_{j,m} - \frac{1}{T_{E_i}} E_{i,n}, \\
\frac{\mathrm{d}I_{i,n}}{\mathrm{d}t} &= \frac{1}{T_{E_i}} E_{i,n} - \frac{1}{T_{I_i}} I_{i,n}, \\
\frac{\mathrm{d}R_{i,n}}{\mathrm{d}t} &= \frac{1}{T_{I_i}} I_{i,n},
\end{aligned} \tag{3.1.26}$$

for all $i \in \{1, \dots, J\}$ and $n \in \{1, \dots, M\}$. As the derivation of (3.1.26) from (3.1.9) is straightforward and analogous to the previous developments, we refrain from providing

more details. The stratification into $J \in \mathbb{N}$ age groups is indicated by an index J, i.e., (3.1.26) is denoted by $\mathcal{M}^{\mathrm{B}}_{J,M}$.

We start by considering Model B with $J \in \mathbb{N}$ age groups and $M \in \mathbb{N}$ patches and order the compartments first chronologically with respect to (potential) disease progression, then by age group and then by patch, i.e., a model state x has the form

$$x = (S_{1,1}, \dots, S_{1,M}, S_{2,1}, \dots, S_{J,M}, E_{1,1}, \dots, E_{J,M}, I_{1,1}, \dots, I_{J,M}, R_{1,1}, \dots, R_{J,M}). \tag{3.1.27}$$

The compartments corresponding to an active infection (not necessarily infectious) in this model are all compartments $E_{i,n}$ and $I_{i,n}$ for $i \in \{1, \dots, J\}$, $n \in \{1, \dots, M\}$. Therefore, the set of indices corresponding to infected compartments is

$$\mathcal{J} = \{k \mid x_k \in \{E_{i,n}, I_{i,n} \mid i = 1, \dots, J,\ n = 1, \dots, M\}\}. \tag{3.1.28}$$

In the following considerations, the superindex represents the association with the model. For a model state x, we conclude the flows:

$$\mathcal{F}^{\mathrm{B}}_k(x) = \begin{cases} 0, & x_k = S_{i,n}, \\ \rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \sum_{m=1}^{M} \left(\frac{h_{nm}}{P_{j,m}} + \frac{h_{mn}}{P_{j,n}} \right) I_{j,m}, & x_k = E_{i,n}, \\ 0, & x_k = I_{i,n}, \\ 0, & x_k = R_{i,n}, \end{cases} \tag{3.1.29}$$

and

$$\mathcal{V}^{\mathrm{B}}_k(x) = \begin{cases} \rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \sum_{m=1}^{M} \left(\frac{h_{nm}}{P_{j,m}} + \frac{h_{mn}}{P_{j,n}} \right) I_{j,m}, & x_k = S_{i,n}, \\ \frac{1}{T_{E_i}} E_{i,n}, & x_k = E_{i,n}, \\ \frac{1}{T_{I_i}} I_{i,n} - \frac{1}{T_{E_i}} E_{i,n}, & x_k = I_{i,n}, \\ -\frac{1}{T_{I_i}} I_{i,n}, & x_k = R_{i,n}. \end{cases} \tag{3.1.30}$$

Let $x^* \in X_s$ be a DFE. Then it has the form

$$x^* = (S_{1,1}, \dots, S_{J,M}, 0, \dots, 0, 0, \dots, 0, P_{1,1} - S_{1,1}, \dots, P_{J,M} - S_{J,M}). \tag{3.1.31}$$

Next, we differentiate $\mathcal{F}^{\mathrm{B}}_k(x)$ with respect to x_l, where $k, l \in \mathcal{J}$. We have

$$f^{\mathrm{B}}_{k,l}(x^*) = \frac{\partial \mathcal{F}^{\mathrm{B}}_k}{\partial x_l}(x^*) = \begin{cases} 0, & x_k = E_{i,n}, x_l = E_{j,m}, \\ \rho_i S_{i,n} \phi_{ij} \left(\frac{h_{nm}}{P_{j,m}} + \frac{h_{mn}}{P_{j,n}} \right), & x_k = E_{i,n}, x_l = I_{j,m}, \\ 0, & x_k = I_{i,n}, x_l = E_{j,m}, \\ 0, & x_k = I_{i,n}, x_l = I_{j,m}. \end{cases} \tag{3.1.32}$$

In matrix form, $F^{\mathrm{B}}(x^*)$ writes:

$$F^{\mathrm{B}}(x^*) = \left(\begin{array}{c|ccc} & F^{\mathrm{B}}_{1,1}(x^*) & F^{\mathrm{B}}_{1,2}(x^*) & \cdots & F^{\mathrm{B}}_{1,J}(x^*) \\ & F^{\mathrm{B}}_{2,1}(x^*) & F^{\mathrm{B}}_{2,2}(x^*) & \cdots & F^{\mathrm{B}}_{2,J}(x^*) \\ \mathbf{0} & \vdots & \vdots & \ddots & \vdots \\ & F^{\mathrm{B}}_{J,1}(x^*) & F^{\mathrm{B}}_{J,2}(x^*) & \cdots & F^{\mathrm{B}}_{J,J}(x^*) \\ \hline \mathbf{0} & \multicolumn{3}{c}{\mathbf{0}} \end{array}\right) \tag{3.1.33}$$

where

$$F^{\mathrm{B}}_{i,j}(x^*) = \rho_i \phi_{ij} \begin{pmatrix} S_{i,1}\left(\frac{h_{11}}{P_{j,1}} + \frac{h_{11}}{P_{j,1}}\right) & S_{i,1}\left(\frac{h_{12}}{P_{j,2}} + \frac{h_{21}}{P_{j,1}}\right) & \cdots & S_{i,1}\left(\frac{h_{1M}}{P_{j,M}} + \frac{h_{M1}}{P_{j,1}}\right) \\ S_{i,2}\left(\frac{h_{21}}{P_{j,1}} + \frac{h_{12}}{P_{j,2}}\right) & S_{i,2}\left(\frac{h_{22}}{P_{j,2}} + \frac{h_{22}}{P_{j,2}}\right) & \cdots & S_{i,2}\left(\frac{h_{2M}}{P_{j,M}} + \frac{h_{M2}}{P_{j,2}}\right) \\ \vdots & \vdots & \ddots & \vdots \\ S_{i,M}\left(\frac{h_{M1}}{P_{j,1}} + \frac{h_{1M}}{P_{j,M}}\right) & S_{i,M}\left(\frac{h_{M2}}{P_{j,2}} + \frac{h_{2M}}{P_{j,M}}\right) & \cdots & S_{i,M}\left(\frac{h_{MM}}{P_{j,M}} + \frac{h_{MM}}{P_{j,M}}\right) \end{pmatrix}. \tag{3.1.34}$$

For the matrix $V^{\mathrm{B}}(x^*)$ we observe:

$$v^{\mathrm{B}}_{k,l}(x^*) = \frac{\partial \mathcal{V}^{\mathrm{B}}_k}{\partial x_l}(x^*) = \begin{cases} \frac{1}{T_{E_j}}\delta_{ij}\delta_{nm}, & x_k = E_{i,n}, x_l = E_{j,m}, \\ 0, & x_k = E_{i,n}, x_l = I_{j,m}, \\ -\frac{1}{T_{E_j}}\delta_{ij}\delta_{nm}, & x_k = I_{i,n}, x_l = E_{j,m}, \\ \frac{1}{T_{I_j}}\delta_{ij}\delta_{nm}, & x_k = I_{i,n}, x_l = I_{j,m}, \end{cases} \tag{3.1.35}$$

where δ_{ij} is the Kronecker delta.

Therefore, the matrix $V^{\mathrm{B}}(x^*)$ is

$$V^{\mathrm{B}}(x^*) = \left(\begin{array}{cccc|cccc} \frac{1}{T_{E_1}}\mathrm{Id}_M & 0 & \cdots & 0 & & & & \\ 0 & \frac{1}{T_{E_2}}\mathrm{Id}_M & & \vdots & & & & \\ \vdots & & \ddots & 0 & & & \mathbf{0} & \\ 0 & \cdots & 0 & \frac{1}{T_{E_J}}\mathrm{Id}_M & & & & \\ \hline -\frac{1}{T_{E_1}}\mathrm{Id}_M & 0 & \cdots & 0 & \frac{1}{T_{I_1}}\mathrm{Id}_M & 0 & \cdots & 0 \\ 0 & -\frac{1}{T_{E_2}}\mathrm{Id}_M & & \vdots & 0 & \frac{1}{T_{I_2}}\mathrm{Id}_M & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & -\frac{1}{T_{E_J}}\mathrm{Id}_M & 0 & \cdots & 0 & \frac{1}{T_{I_J}}\mathrm{Id}_M \end{array}\right), \tag{3.1.36}$$

where Id_M is the identity matrix of $\mathbb{R}^{M\times M}$.

Using (2.3.15) we get the inverse of $V^{\mathrm{B}}(x^*)$

$$\left(V^{\mathrm{B}}(x^*)\right)^{-1} = \left(\begin{array}{cccc|cccc} T_{E_1}\mathrm{Id}_M & 0 & \cdots & 0 & & & & \\ 0 & T_{E_2}\mathrm{Id}_M & & \vdots & & & & \\ \vdots & & \ddots & 0 & & \multicolumn{2}{c}{\mathbf{0}} & \\ 0 & \cdots & 0 & T_{E_J}\mathrm{Id}_M & & & & \\ \hline T_{I_1}\mathrm{Id}_M & 0 & \cdots & 0 & T_{I_1}\mathrm{Id}_M & 0 & \cdots & 0 \\ 0 & T_{I_2}\mathrm{Id}_M & & \vdots & 0 & T_{I_2}\mathrm{Id}_M & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & T_{I_J}\mathrm{Id}_M & 0 & \cdots & 0 & T_{I_J}\mathrm{Id}_M \end{array}\right). \tag{3.1.37}$$

The next generation matrix of the age-resolved version of Model B (3.1.26) is

$$F^{\mathrm{B}}(x^*)\left(V^{\mathrm{B}}(x^*)\right)^{-1} = \left(\begin{array}{c|c} A^{\mathrm{B}}(x^*) & A^{\mathrm{B}}(x^*) \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \tag{3.1.38}$$

where

$$\begin{aligned} A^{\mathrm{B}}(x^*) &= \begin{pmatrix} F^{\mathrm{B}}_{1,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{B}}_{1,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{B}}_{1,J}(x^*)(T_{I_J}\mathrm{Id}_M) \\ F^{\mathrm{B}}_{2,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{B}}_{2,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{B}}_{2,J}(x^*)(T_{I_J}\mathrm{Id}_M) \\ \vdots & \vdots & \ddots & \vdots \\ F^{\mathrm{B}}_{J,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{B}}_{J,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{B}}_{J,J}(x^*)(T_{I_J}\mathrm{Id}_M) \end{pmatrix} \\ &= \begin{pmatrix} T_{I_1}F^{\mathrm{B}}_{1,1}(x^*) & T_{I_2}F^{\mathrm{B}}_{1,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{B}}_{1,J}(x^*) \\ T_{I_1}F^{\mathrm{B}}_{2,1}(x^*) & T_{I_2}F^{\mathrm{B}}_{2,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{B}}_{2,J}(x^*) \\ \vdots & \vdots & \ddots & \vdots \\ T_{I_1}F^{\mathrm{B}}_{J,1}(x^*) & T_{I_2}F^{\mathrm{B}}_{J,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{B}}_{J,J}(x^*) \end{pmatrix}. \end{aligned} \tag{3.1.39}$$

As in the case of Model A (2.3.1), we see that it suffices to compute the spectral radius of $A^{\mathrm{B}}(x^*)$. Choosing the DFE $x_0^* = (P_{1,1},\ldots,P_{M,M},0,\ldots,0,0,\ldots,0,0,\ldots,0)$, we obtain the basic reproduction number of Model B

$$\mathcal{R}_0^{\mathrm{B}} = \rho\left(A^{\mathrm{B}}(x^*)\right). \tag{3.1.40}$$

Let us now study the basic reproduction number of Model C. When comparing Model B (3.1.26) and Model C (3.1.25), we observe that $\mathcal{V}_k^{\mathrm{B}} = \mathcal{V}_k^{\mathrm{C}}$ for all $x_k \in \mathcal{J}$ and thus $V^{\mathrm{B}}(x^*) = V^{\mathrm{C}}(x^*)$. Hence, we only have to construct $F^{\mathrm{C}}(x^*)$ for Model C (3.1.25).

For a model state x, we observe:

$$\mathcal{F}_k^{\mathrm{C}}(x) = \begin{cases} 0, & x_k = S_{i,n}, \\ \frac{1}{2}\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \left(\frac{I_{j,n}}{P_{j,n}} + \sum_{m=1}^{M} \frac{h_{nm} I_{j,m}^{tot}}{N_{j,m}} \right), & x_k = E_{i,n}, \\ 0, & x_k = I_{i,n}, \\ 0, & x_k = R_{i,n}, \end{cases} \tag{3.1.41}$$

and differentiate this with respect to x_l for $k, l \in \mathcal{J}$. Recall that $I_{j,m}^{tot}$ depends on the infectious population in each patch and age group via (3.1.24). Let $i \in \{1, \ldots, J\}$, $n \in \{1, \ldots, M\}$, k such that $x_k = E_{i,n}$. We reformulate $\mathcal{F}_k$ as follows:

$$\begin{aligned} \mathcal{F}_k^{\mathrm{C}}(x) &= \frac{1}{2}\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \left(\frac{I_{j,n}}{P_{j,n}} + \sum_{m=1}^{M} \frac{h_{nm} I_{j,m}^{tot}}{N_{j,m}} \right) \\ &= \frac{1}{2}\rho_i S_{i,n} \sum_{j=1}^{J} \phi_{ij} \left(\frac{I_{j,n}}{P_{j,n}} + \sum_{m=1}^{M} \frac{h_{nm} \sum_{\widehat{m}=1}^{M} h_{\widehat{m}m} I_{j,\widehat{m}}}{N_{j,m}} \right). \end{aligned} \tag{3.1.42}$$

A DFE $x^* \in X_s$ for Model C has the same form as for Model B, i.e., the form (3.1.31). We can now differentiate $\mathcal{F}_k$ with respect to x_l where $k, l \in \mathcal{J}$:

$$f_{k,l}^{\mathrm{C}}(x^*) = \frac{\partial \mathcal{F}_k^{\mathrm{C}}}{\partial x_l}(x^*) = \begin{cases} 0, & x_k = E_{i,n}, x_l = E_{j,m}, \\ \frac{1}{2}\rho_i S_{i,n} \phi_{ij} \left(\frac{\delta_{nm}}{P_{j,n}} + \sum_{\widehat{m}=1}^{M} \frac{h_{n\widehat{m}} h_{m\widehat{m}}}{N_{j,\widehat{m}}} \right), & x_k = E_{i,n}, x_l = I_{j,m}, \\ 0, & x_k = I_{i,n}, x_l = E_{j,m}, \\ 0, & x_k = I_{i,n}, x_l = I_{j,m}. \end{cases} \tag{3.1.43}$$

The matrix form of $F^{\mathrm{C}}(x^*)$ is similar to $F^{\mathrm{B}}(x^*)$:

$$F^{\mathrm{C}}(x^*) = \left(\begin{array}{c|ccc} & F_{1,1}^{\mathrm{C}}(x^*) & \cdots & F_{1,J}^{\mathrm{C}}(x^*) \\ \mathbf{0} & \vdots & \ddots & \vdots \\ & F_{J,1}^{\mathrm{C}}(x^*) & \cdots & F_{J,J}^{\mathrm{C}}(x^*) \\ \hline \mathbf{0} & \multicolumn{3}{c}{\mathbf{0}} \end{array} \right), \tag{3.1.44}$$

where

$$F^{\mathrm{C}}_{i,j}(x^*) = \frac{1}{2}\rho_i\phi_{ij}\begin{pmatrix} S_{i,1}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{1\widehat{m}}h_{1\widehat{m}}}{N_{j,\widehat{m}}} & S_{i,1}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{1\widehat{m}}h_{2\widehat{m}}}{N_{j,\widehat{m}}} & \cdots & S_{i,1}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{1\widehat{m}}h_{M\widehat{m}}}{N_{j,\widehat{m}}} \\ S_{i,2}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{2\widehat{m}}h_{1\widehat{m}}}{N_{j,\widehat{m}}} & S_{i,2}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{2\widehat{m}}h_{2\widehat{m}}}{N_{j,\widehat{m}}} & \cdots & S_{i,2}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{2\widehat{m}}h_{M\widehat{m}}}{N_{j,\widehat{m}}} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i,M}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{M\widehat{m}}h_{1\widehat{m}}}{N_{j,\widehat{m}}} & S_{i,M}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{M\widehat{m}}h_{2\widehat{m}}}{N_{j,\widehat{m}}} & \cdots & S_{i,M}\sum\limits_{\widehat{m}=1}^{M}\frac{h_{M\widehat{m}}h_{M\widehat{m}}}{N_{j,\widehat{m}}} \end{pmatrix} + \begin{pmatrix} \frac{1}{P_{j,1}} & 0 & \dots & 0 \\ 0 & \frac{1}{P_{j,2}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{P_{j,M}} \end{pmatrix}. \tag{3.1.45}$$

Multiplying $F^{\mathrm{C}}(x^*)$ with $\left(V^{\mathrm{C}}(x^*)\right)^{-1} = \left(V^{\mathrm{B}}(x^*)\right)^{-1}$ yields

$$F^{\mathrm{C}}(x^*)\left(V^{\mathrm{C}}(x^*)\right)^{-1} = \left(\begin{array}{c|c} A^{\mathrm{C}}(x^*) & A^{\mathrm{C}}(x^*) \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right), \tag{3.1.46}$$

where

$$\begin{aligned} A^{\mathrm{C}}(x^*) &= \begin{pmatrix} F^{\mathrm{C}}_{1,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{C}}_{1,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{C}}_{1,J}(x^*)(T_{I_J}\mathrm{Id}_M) \\ F^{\mathrm{C}}_{2,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{C}}_{2,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{C}}_{2,J}(x^*)(T_{I_J}\mathrm{Id}_M) \\ \vdots & \vdots & \ddots & \vdots \\ F^{\mathrm{C}}_{J,1}(x^*)(T_{I_1}\mathrm{Id}_M) & F^{\mathrm{C}}_{J,2}(x^*)(T_{I_2}\mathrm{Id}_M) & \cdots & F^{\mathrm{C}}_{J,J}(x^*)(T_{I_J}\mathrm{Id}_M) \end{pmatrix} \\ &= \begin{pmatrix} T_{I_1}F^{\mathrm{C}}_{1,1}(x^*) & T_{I_2}F^{\mathrm{C}}_{1,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{C}}_{1,J}(x^*) \\ T_{I_1}F^{\mathrm{C}}_{2,1}(x^*) & T_{I_2}F^{\mathrm{C}}_{2,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{C}}_{2,J}(x^*) \\ \vdots & \vdots & \ddots & \vdots \\ T_{I_1}F^{\mathrm{C}}_{J,1}(x^*) & T_{I_2}F^{\mathrm{C}}_{J,2}(x^*) & \cdots & T_{I_J}F^{\mathrm{C}}_{J,J}(x^*) \end{pmatrix}. \end{aligned} \tag{3.1.47}$$

For $x_0^* = (P_{1,1}, \dots, P_{M,M}, 0, \dots, 0, 0, \dots, 0, 0, \dots, 0)$, the basic reproduction number is given by

$$\mathcal{R}_0^{\mathrm{C}} = \rho\left(A^{\mathrm{C}}(x^*)\right). \tag{3.1.48}$$

As in Section 2.3 for Model A, we will not give an analytical solution here but compute the basic reproduction numbers numerically in Section 5.3.

3.2 A semi-discrete hybrid graph-based ODE model (Model D)

Another approach to integrate spatial dynamics in *SEIR*-type disease models was given by [20]. In this approach, decoupled ODE systems have been connected by a graph, obtaining a hybrid graph-ODE-based approach. In a first step, we, again, divide the total population in $M \in \mathbb{N}$ smaller subpopulations that live in spatially separated areas, which we call patches. Every subpopulation is further separated into $J \in \mathbb{N}$ age groups, as well as into the four epidemiological groups Susceptible, Exposed, Infected and Recovered. In line with the equation-based metapopulation models Model B (3.1.26) and Model C (3.1.25), we denote the number of susceptible individuals of age group i at patch $\mathcal{P}_n$ by $S_{i,n}$ for $i \in \{1, \dots, J\}$ and $m \in \{1, \dots, M\}$. For the other compartments we proceed accordingly, obtaining $E_{i,n}$, $I_{i,n}$ and $R_{i,n}$. The spatially separated patches can be connected through migration of one patch to another. As before, only short-term migration is considered. More precisely, an assumption of this model is that individuals travel after half a day and return home at the end of the day. In addition, this model allows the commuting rates to depend on infection states and on the age of an individual. For instance, it can be modeled that infectious individuals travel less than healthy individuals as they could be isolated. In order to introduce the hybrid graph-ODE-based approach, we need to introduce some basic graph structures.

A *graph* G is a pair $G = (V, E)$, where V is the finite set of *vertices* and E the set of *edges* [18]. Every edge connects two vertices, so E is a set of 2-element subsets of V. For a *directed graph*, every edge $e \in E$ is associated with an ordered pair (v_1, v_2), where $v_1, v_2 \in V$, and called a directed edge. A directed graph containing multiple directed edges which connect the same vertices is called a *directed multi-graph* [33]. In the following, the set of all directed edges (v_1, v_2) connecting two vertices $v_1, v_2 \in V$ is called the *multi-edge* $e_{v_1 v_2}$.

Comparing these definitions to the model setting, it seems natural to identify the nodes of a graph with the patches and the edges with the mobility between them. Since we have not assumed that the mobility is symmetric, i.e., that the number of individuals traveling from patch $\mathcal{P}_n$ to $\mathcal{P}_m$ is equal to the number of individuals commuting from $\mathcal{P}_m$ to $\mathcal{P}_n$, this network corresponds to a directed graph. As the commuting rates between the patches may depend on the infection states and age of individuals, the network is a directed multi-graph, containing multi-edges which combine edges for each pair of infection state and age group. The multi-edge connecting two patches $\mathcal{P}_n$ and $\mathcal{P}_m$ is denoted by e_{nm} and consists of J edges per compartment, i.e., of $4J$ edges in total, where $J \in \mathbb{N}$ is the number of age groups. In the simplest case, in which the commuting rates do not depend on the infection states or the age, every single edge of e_{nm} corresponds to h_{nm} (3.1.2).

The difference in this graph-based approach [20] is that we assign a separate *SEIR*

model to each patch, i.e., the equations of different patches are not coupled to each other. Instead of one large system of ODEs for all patches, we have many smaller ODE systems, precisely one for each patch. In a model with $M \in \mathbb{N}$ patches, the approach presented in Section 3.1 leads to a system of $4M$ coupled ODEs while the graph-based approach leads to M systems with 4 ODEs each.

In contrast to the equation-based metapopulation models, in which commuting is considered at any time point throughout the simulation, the graph-based model regards commuting as a form of human contact behavior, with trips for work at specific time points. At the discrete time points, individuals are explicitly exchanged between patches by transitioning shares of the population between the corresponding ODE systems. In between the discrete time points, every ODE model can be considered independent and can be solved with a common ODE solver. Given the independence of local models, the solution between the discrete time points can be done in parallel, providing great parallelization properties.

We now illustrate the numerical procedure for solving the model using an example with two patches $\mathcal{P}_1$ and $\mathcal{P}_2$ and the simple case where mobility rates do not depend on age or infection states. Hence, each single edge of e_{12} corresponds to the commuting strength h_{12}, i.e., the share of individuals from patch $\mathcal{P}_1$ commuting to $\mathcal{P}_2$, and e_{21} vice versa. The numerical procedure is visualized in Fig. 3.1. From a given start date t_0 and corresponding initial values

$$S_{i,n}(t_0), E_{i,n}(t_0), I_{i,n}(t_0), R_{i,n}(t_0), \quad n = 1, 2, \text{ and } i = 1, \ldots, J, \tag{3.2.1}$$

an arbitrary ODE solver is advanced (in multiple steps) with user-defined or adaptive time stepping from start date t_0 to preliminary end time $t_0 + 1/2$. Then, the commuting is modeled by exchanging the respective number of individuals in each age group and each infection state. At patch $\mathcal{P}_1$, the number of susceptible individuals of an age group i is reduced by $h_{12}S_{i,1}(t + 0.5)$, since this is the mean number of susceptible individuals commuting to patch $\mathcal{P}_2$. Concurrently, the number of susceptible individuals at patch $\mathcal{P}_2$ is increased by the same number. This is done for each compartment and age group, as well as for the reverse direction. At the same time, the shares of commuting individuals are also stored separately with the current infection states to later extrapolate the infection states of individuals commuting home. Next, the solver is advanced from the time point of commuting $t_0 + 1/2$ to the end of the day at time $t_0 + 1$.

At $t_0 + 1$, the commuters are expected to return to their home patches. The nature of the equation-based models, which do not trace individuals, poses a challenge to adequately find the developed infection states of the commuters from $t_0 + 1/2$. When adding individuals to another patch and advancing the solver, information on commuting individuals is lost in the entire population of the target patch. This effect can be problematic, as the objective is to ensure that the individuals added are those who return at the end

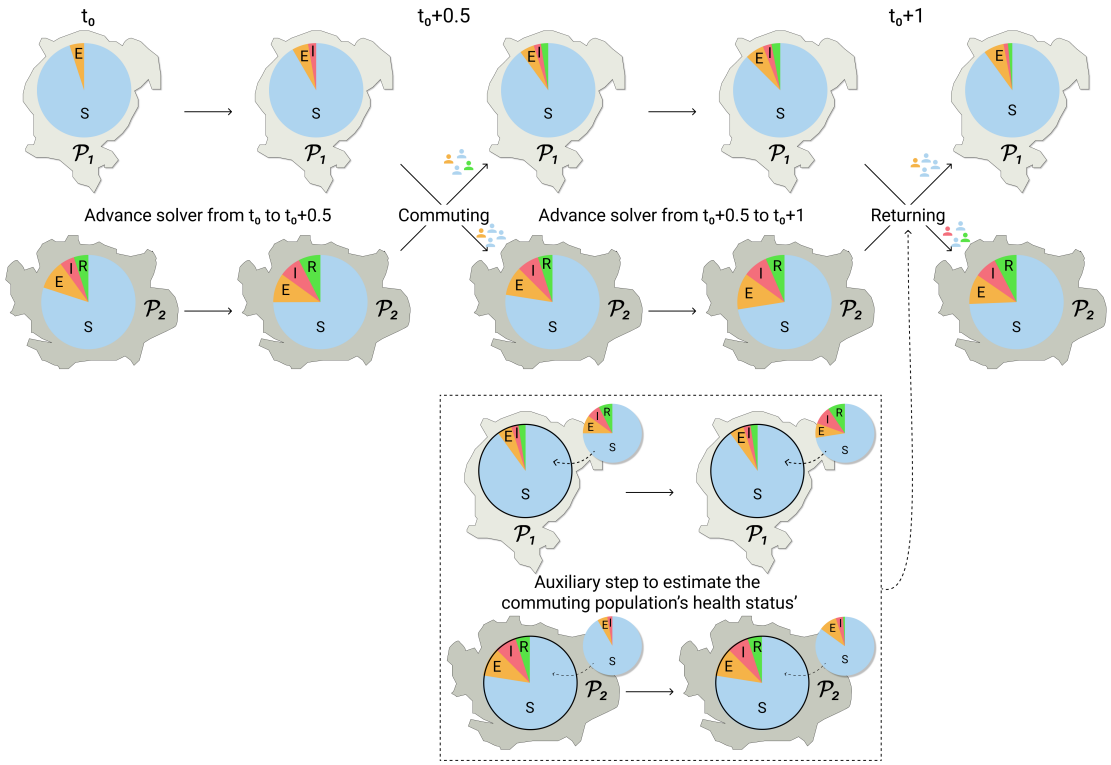


Figure 3.1: **Schematic representation of the numerical solution for the graph-based metapopulation model.** Two patches $\mathcal{P}_1$ and $\mathcal{P}_2$ are represented by separate *SEIR* models. After advancing a solver for a half day, commuters are exchanged between patches. Next, the solver is advanced for another half day. Before returning the commuters, their infection states are estimated using one step of a low-order single-step method where the populations at the patches are used as contact population only. Hence, the populations at the patches remain constant during the auxiliary step, as indicated by the presence of a black border. Once their health status' are determined, the individuals are returned. This figure is adopted from [20, Fig. 2].

of the day. We illustrate this issue using the following simplified example.

Let us consider a patch $\mathcal{P}_n$ where everybody is susceptible, and a patch $\mathcal{P}_m$ where everybody is infectious. For simplicity, we assume $e_{nm} = 0$, i.e., commuting is only performed in the direction $\mathcal{P}_m \to \mathcal{P}_n$. During commuting, some infectious individuals from patch $\mathcal{P}_m$ travel to patch $\mathcal{P}_n$. After advancing the solver, we expect several susceptible individuals from patch $\mathcal{P}_m$ to get exposed to the pathogen and upon transmission getting into the Exposed state. Additionally, we expect some infectious individuals to recover. If returning is performed in the same way as the initial commuting, this would result in a returning population consisting of the same distribution as the population at patch $\mathcal{P}_n$. In particular, we would transfer individuals with a susceptible infection state. Since the initial commuting population consisted only of infectious individuals, these individuals

can only be either in the Infected or the Recovered infection states. We visualize this example in Fig. 3.2. For simplicity, the visualization ignores the fact that the infection progresses also in the population at patch $\mathcal{P}_m$.

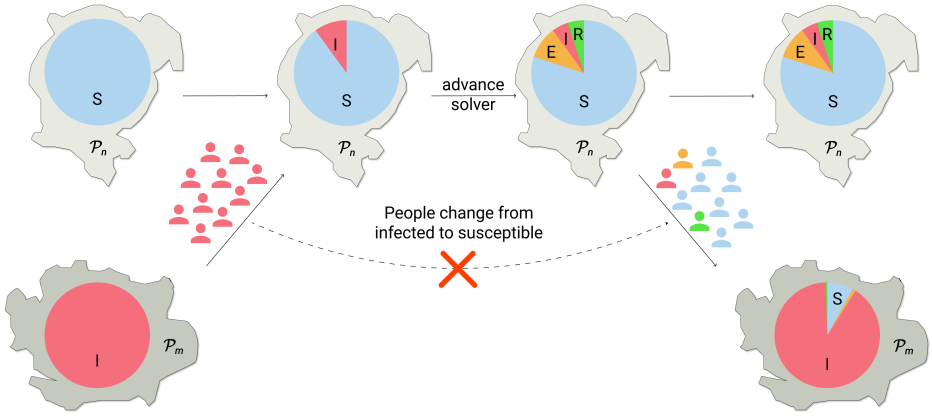


Figure 3.2: **Visualization of the necessity of the auxiliary step for handling returning commuters.** We have one population of only susceptible individuals (top) and one with only infected individuals (bottom). During commuting, individuals from one population travel to the other, where the infection advances. At the end of the day, the same fraction of individuals returns. Without further considerations, this approach leads to the problem that former infectious individuals are now susceptible, which is not possible in the ODE models under study.

To better extrapolate the returning population's health status', the model performs an additional step with the commuting population. Using a single-step method, e.g., the explicit Euler method, with a step size of $h = 0.5$ days, we estimate the most likely infection states of the commuters at the end of the day, given the commuting population at time $t_0+1/2$. To achieve this, we apply the *SEIR* model to this commuting population, but use the population of the target patch at time $t_0 + 1/2$ as a contact population. The single-step method approximates the change in infection states among the commuting population during their stay at the target location. Consequently, we determine the numbers of returning individuals in each infection state and update each subpopulation according to these numbers. After all individuals have returned, the next simulation day begins with the advancing of all ODE systems representing the patches.

In the following chapters, we will refer to the model derived using the hybrid graph-ODE-based approach with *Model D* and denote it by $\mathcal{M}^{\mathrm{D}}_{J,M}$.

The here presented graph-based model offers a wide range of options for further extensions and improvements. For instance, as demonstrated by [19], the incorporation of commuter testing can enhance the model's realism. A different extension concerns the accounting of transmissions during commuting, as has been proposed by [47]. However,

since it also introduces greater complexity, we will not consider it for the purposes of this thesis.

Theoretically, the graph-based approach is not limited to ODE-based models in the patches. For instance, the patches could also use integro-differential equation- (IDE) [42] or agent-based (ABM) [14] models. Moreover, patches with ODE models and patches with ABMs can be combined in hybrid models [6]. However, as the focus of this thesis is the introduction of mobility to ODE models, we will not go into this any further, but continue with the graph of *SEIR* models.

While the numerical procedure of the graph-based model has proven to be more challenging, this approach has also shown better parallelization properties than the equation-based metapopulation models. We will compare the models in the following chapters on a theoretical level and based on numerical simulations.

4 Implementation

After having derived the models from a theoretical point of view, we now describe how the models are put into practice. We will begin with describing the software framework used by the models and which served as a basis for this thesis. Then, we will explain how the models are implemented. While the framework offers more than 10 different model types, we will restrict ourselves to the implementations of the ODE metapopulation Model C (3.1.25) and the graph-based Model D. The complexities of both implementations will be compared on a theoretical level using the number of floating point operations (FLOP) as an indicator.

Both models have been integrated into the software MEmilio [21] which is still under development. The implementation of the graph-based model approach described in Section 3.2 was already contained in the software and will only be explained briefly. The implementation of the equation-based metapopulation approach is part of this thesis and will be described in the following. Our implementation makes wide use of the infrastructure provided by the software framework which we will explain beforehand, using Model A, i.e., the simple ODE *SEIR* model (2.3.1), as an example. The framework for the models and their simulations is implemented in C++.

4.1 Existing prerequisites in the software

We start by describing the implementation of Model A (2.3.1) which was already part of the software. Afterward, we outline which changes had to be made for the extension to the metapopulation model. Each model has its own folder and namespace `mio::model_space::` containing the classes specific to that model, where `model_space` is replaced by a model specific name. However, these classes often inherit from other classes which allow a wide range of options and specifications through their generic implementation. For many matrix and vector operations, we use the library *Eigen* which provides fast and reliable algorithms for linear algebra in C++ [12].

A model is defined by a class `Model` that inherits from a class `FlowModel` which is itself a child class of `CompartmentalModel`. Both `FlowModel` and `CompartmentalModel` are generic classes from which various compartmental models are derived. The class `Model` is specific to the model, which here is Model A. Listing 4.1 shows the headers of `Model` and its base classes, including the template parameters, as well as the most

important members in each class. It is also possible to derive the model class from the `CompartmentalModel` class directly. In the following, we briefly visualize the difference between a `FlowModel` and a `CompartmentalModel` using the transitions from and to the Infected compartment as an example. While the `CompartmentalModel` computes the derivative in the Infected compartment as

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{E}{T_E} - \frac{I}{T_I}, \tag{4.1.1}$$

the `FlowModel` will compute

$$\begin{aligned} F_{E\to I} &= E/T_E, \\ F_{I\to R} &= I/T_I, \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= F_{E\to I} - F_{I\to R}. \end{aligned} \tag{4.1.2}$$

While this may seem more inconvenient, it allows more insight into the explicit flows, as the aggregated solution to the first computation loses track of the particular in- and outflows of the compartments. Moreover, for more complex flows, the `FlowModel` requires fewer computations, as the `CompartmentalModel` computes the flow for both compartments involved.

`Model` holds a template parameter `FP`, which determines the floating point precision used by the model. In the simulations considered in this thesis, `FP` is specified by `double`. This template parameter is also passed to the base classes `FlowModel` and `CompartmentalModel`. Additionally, `FlowModel` has the template parameters `Comp, Pop, Params` and `Flows` where the first three are specified and passed to `CompartmentalModel`. We will describe the specifications for the template parameters in the following. The template parameter `Flows` is used to define which transitions occur in the model. For the *SEIR* model, these are the transitions from Susceptible to Exposed, from Exposed to Infected and from Infected to Recovered.

The template parameter `Comp` determines the compartments of interest, which for Model A are the compartments *Susceptible (S)*, *Exposed (E)*, *Infected (I)* and *Recovered (R)*, accordingly. They are given by an enum class `InfectionState` that enumerates these compartments contiguously from 0, ending with a non-compartment `Count` which represents the number of compartments.

`Pop` determines how the model population is stored. It is usually specified by a class `Populations` which allows splitting a population into smaller categories, such as age groups or gender, and stores a corresponding value of type `FP`, given as a template parameter. To achieve that an arbitrary number of categories can be implemented, `Populations` is a variadic class template which allows using any number of template parameters and which takes a parameter pack `Categories` as a template parameter. The class inherits from a class `CustomIndexArray` that allows multidimensional typesafe

```cpp
template <typename FP>
class Model : public FlowModel<FP, InfectionState, Populations<FP,
       AgeGroup, InfectionState>, Parameters<FP>, Flows>
{
    void get_flows(...) const override;
};

template <typename FP, class Comp, class Pop, class Params, class Flows>
class FlowModel : public CompartmentalModel<FP, Comp, Pop, Params>
{
    virtual void get_flows(...) const;
    void get_derivatives(...) const override;
};

template <typename FP, class Comp, class Pop, class Params>
class CompartmentalModel
{
    virtual void get_derivatives(...) const;
    void eval_right_hand_side(...) const;
    Populations populations{};
    ParameterSet parameters{};
};
```

Listing 4.1: **Code snippet of the SEIR Model class and its parent classes.** Class headers including templatization and the most important members. The `Model` class is specific to the *SEIR* model and inherits from the `FlowModel` class, which is derived from the `CompartmentalModel` class.

indexing into a one-dimensional array. This is achieved with a class `Index`, which connects a `size_t` with a tag of any type and combines several `Index` objects to obtain indices of multiple dimensions. Tags are used to "stratify" the population, i.e., for a population divided into the compartments, the `InfectionState` enum class serves as a tag. We can use such a custom index array to store a population divided into, e.g., age groups and compartments, together with the respective number of individuals. The `Index` class makes it (near) impossible to confuse indices or index order, as long as the tags are unique.

The *SEIR* model describes a population divided into the particular compartments and age groups. Thus, `Pop` is specified by `Populations<FP, AgeGroup, InfectionState>`, where `InfectionState` is the enum from before. `AgeGroup` is a struct which is derived from the `Index` class to provide a dynamically sized tag and to shorten the notation. Later, we will also use this concept to model regions and then go into more detail.

The last template parameter of the `FlowModel` is `Params`, which is specified by a class `Parameters` that is specific to the model and inherits from a parent class `ParameterSet`.

`ParameterSet` holds the parameters of the model, such as the transmission probability or the mean transition times from Exposed to Infected or from Infected to Recovered, defined through structs passed to it via variadic template parameters. Every parameter struct has member functions that return the name of the parameter or a default value. By adding new structs, we can later implement additional parameters very easily and can use get and set functions without additional effort.

As we will use the same structure later on, we will take a look at the parameter `ContactPatterns` and its underlying type `UncertainContactMatrix`. The class provides a (contact) matrix of size $J \times J$ where J is the number of age groups. It gives the contact frequency of individuals depending on their age, i.e., the mean number of contacts between individuals of age group i with age group j, for all age groups $i, j \in \{1, \dots, J\}$. Additionally, the class can hold information to adjust the frequencies during the simulation, for example to model non-pharmaceutical interventions like social distancing.

As an object of a derived class of `CompartmentalModel`, the model holds an attribute `populations` of type `Population` that represents the initial population and an attribute `parameters` of type `Parameters` with the model specific parameters, compare line 19 and 20 of Listing 4.1. Both the constructor of `Population` and `Parameters` depend on the number of age groups, hence the constructor of the model is called with a parameter `size_t number_age_groups`. The model constructor first instantiates both `Population` and `Parameters`, then passes them to the constructor of its parent class, `FlowModel`.

The differential equations of the model are formulated through a member function `get_flows()`. It implements the values of the flows between compartments for a current model state, which then are combined by a function `get_derivatives()` to the values of the derivative, as in (4.1.2). In Section 3.2, we explained the necessity for the additional step, in which the commuting population's infection states are estimated when returning. As the population at the target region serves only as a contact population, the implementation of `get_flows()` distinguishes between contact and evolving population. For all other cases without mobility, both populations are the same.

The initial population together with the differential equations pose an initial value problem (IVP) which we solve numerically. For full control on the step size and to keep the complexity analysis in the following section simple, we will mainly focus on a low-order, fixed step-size method, i.e., the explicit Euler method. This method approximates the integration over time of the IVP iteratively in discrete steps via

$$y^{k+1} = y^k + hf(t_k, y^k). \tag{4.1.3}$$

Here, f is the right-hand side of the ODE system, y^k is the approximation of the solution to the IVP at time $t_k = t_0 + kh$ and h is the step size.

In the software, a numerical integration scheme is specified by an implementation of an interface `IntegratorCore`, defining the respective step function. For the Euler

```
1  class EulerIntegratorCore : public IntegratorCore<FP>
2  {
3      public:
4          EulerIntegratorCore() : IntegratorCore<FP>(FP{}, FP{})
5          {
6          }
7
8          bool step(const DerivFunction<FP>& f,
9              Eigen::Ref<const Vector<FP>> yt, FP& t, FP& dt,
10             Eigen::Ref<Vector<FP>> ytp1) const override
11         {
12             f(yt, t, ytp1);
13             ytp1 = yt + dt * ytp1;
14             t += dt;
15             return true;
16         }
17 };
```

Listing 4.2: **Code snippet of the `EulerIntegratorCore` class**. The `EulerIntegratorCore` is derived from the `IntegratorCore` class with empty declarations for the minimal and maximal step size, as the Euler method uses a fixed step size. The `step()` function is used to perform a single integration step for a function to be integrated `f`.

method, the respective implementation `EulerIntegratorCore` is shown in Listing 4.2. `IntegratorCore` manages the minimal and maximal step sizes for adaptive methods and requires setting these in its constructor. Being an explicit method with fixed step size, the `EulerIntegratorCore` inherits from the `IntegratorCore` with empty declarations of minimal and maximal step size, compare line 4. The step function performs a single integration step for a function `f` to be integrated. Here, the value of `f` for a current approximation `yt` at time point `t` is computed in line 12 and this value temporarily stored in the vector of the next result `ytp1`. Afterward, the approximation at the next time point is computed according to (4.1.3) and the current time point is updated by the step size. In a recent change to the software, `Vector` was exchanged by `Eigen::VectorX`. However, this change is not considered in this thesis.

The `step()` method allows implementing very general integration methods, such as Runge-Kutta methods. For more realistic comparisons with respect to the time to solution, we use the adaptive Runge-Kutta Cash-Karp 5(4) method. For this, we use the implemented function `runge_kutta_cash_karp54` from *Boost* [1] in combination with the implementation `ControlledStepperWrapper` of `IntegratorCore`, which allows using pre-implemented integration schemes.

An IVP is solved by an instance of the `OdeIntegrator` class, which holds a shared pointer to an implementation of `IntegratorCore`. This `OdeIntegrator` advances from

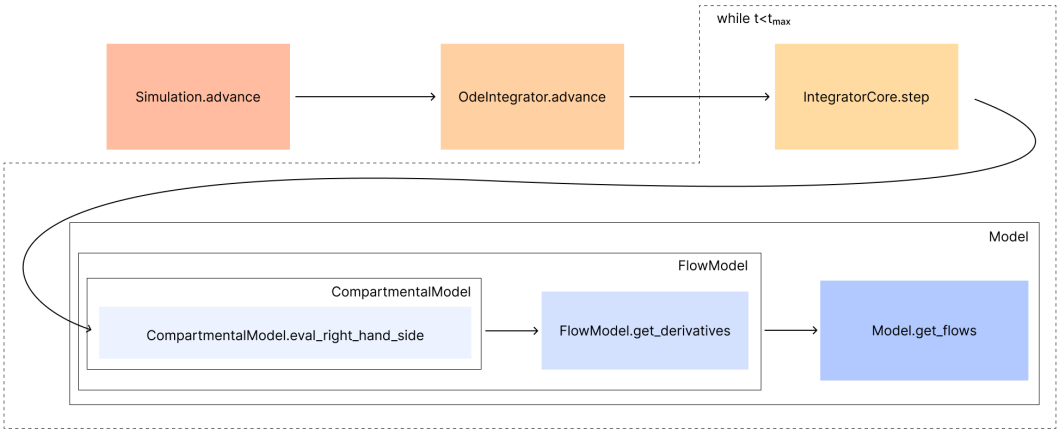


Figure 4.1: **Simplified call graph for the simulation process.** Every call from one function to another is indicated by an arrow, where the arrow is pointed from the caller to the callee. To run a simulation, the `advance()` function of the `Simulation` is called, which then calls the `advance()` function of the integrator. The integrator calls the `step()` function of the core, which computes the result of the next time point, using the right-hand side of the ODEs returned by `eval_right_hand_side()`. To calculate the right-hand side of the model, `eval_right_hand_side()` calls `get_derivatives()`, which combines the flows returned by `get_flows()`. The integrator lets the core perform steps until t_{max} is reached, which is indicated by the dashed box. The boxes with solid borders represent the inheritance of the model classes.

a time t_0 to t_{max}, letting the core perform steps until t_{max}. The model and the numerical method with which it is solved is connected through a class `Simulation`. A `Simulation` object uses the model to define the initial values of the problem returned by `get_initial_values()` which returns the initial population. Also, it defines the right-hand side of the system via a member function `get_flows()` of `Model`. This forms the IVP which can be solved by the `OdeIntegrator`.

To understand the implementation of the simulation process, we can follow the simplified call graph presented in Fig. 4.1. To start a simulation, a function `advance()` of the simulation class is called, which then calls the function `advance()` of the integrator. This function repeatedly calls the step function of the `IntegratorCore` until the simulation end time is reached, using the right-hand side of the ODE system returned by `eval_right_hand_side()` of the `CompartmentalModel` class. We omit some technical details here. In fact, the integrator receives a function object to compute the right-hand side of the ODE system in order to decouple the model implementation from the numerical method. The right-hand side is inferred by `eval_right_hand_side()` via the `get_derivatives()` and `get_flows()` functions, as mentioned above. The results are stored as a `TimeSeries`, which allows tracing the sizes of compartments over time.

4.2 Implemented changes for the equation-based metapopulation Model C

The previous section described the implementation of equation-based models in the software MEmilio [21] using the *SEIR* model as an example. The basic structure of the implemented equation-based metapopulation model proposed with this thesis is very similar. It is derived from a `FlowModel` the same as the *SEIR* model and gets the same class `InfectionState` for the compartments.

To implement commuting, we can make use of the templatization of `Population` and `ParameterSet`. We remember the `Population` class from before and highlight the fact that it allows splitting a population into more than one category. We can interpret the home region of an individual as a characteristic, similar to the age of an individual. Therefore, similar to the struct `AgeGroup`, we implement a struct `Region` derived from the `Index` class, using the curiously recurring template pattern (CRTP) [41], compare Listing 4.3. This way, we can call any custom index array of regions with `Region(val)` and can save the extensive formulation with the index class. Note that a region is only a `size_t` that is used to index the population and the results.

```
struct Region : public Index<Region> {
    Region(size_t val)
        : Index<Region>(val)
    {
    }
};
```

Listing 4.3: **Code snippet of the struct `Region`.** `Region` is a struct that is derived from the `Index` class in order to index the population and the results.

We can now specify the template parameter `Pop` of the base class `FlowModel` with `Populations<FP, Region, AgeGroup, InfectionState>`. The model's constructor is called with the number of regions and the number of age groups. Both parameters are used to instantiate a `Population` of the intended stratification and a `ParameterSet`. Comparing the *SEIR* model (2.2.1) and Model C (3.1.20), we observe that most of the parameters remain the same. The metapopulation model implements two additional parameters: $N = (N_{i,n})_{i\in\{1,\dots,J\},\, n\in\{1,\dots,M\}}$, where $N_{i,n}$ represents the number of individuals of age group i which actually are at a region after commuting and the commuting strengths and $H = (h_{nm})_{n,m\in\{1,\dots,M\}}$, where h_{nm} is the share of individuals from region $\mathcal{P}_n$ which commute to region $\mathcal{P}_m$, analogously to what we derived in Section 3.1.2. We note that we do not need to implement a new parameter for $P_{i,n}$ which denotes the number of individuals of age group i originating from region $\mathcal{P}_n$. As this corresponds to the initial population, this is already represented by `m_population` of the model.

The commuting strengths H are implemented as an `UncertainContactMatrix` of `Region`s and added to the `ParameterSet` of the model. We compute the commuting strengths according to equation (3.1.22) and note that they do not depend on age groups. Therefore, the matrix H has dimension $M \times M$ where $M \in \mathbb{N}$ is the number of regions. The share of the individuals staying at a location is obtained by subtracting all fractions of leaving individuals from the value 1, see (3.1.11). This value is calculated for every region and saved on the diagonal of the matrix.

The population sizes N are represented by another instance of `Population` indicated by `Region` and `AgeGroup` and are computed and set at the beginning of a simulation according to formula (3.1.23).

The biggest modification to the model is the implementation of the equations. This is done in the function `get_flows()` and is outlined in lines 3 to 14 of Algorithm 1. In the beginning, the number of infected individuals per region are computed according to equation (3.1.16) for every region. Afterward, the right-hand side of the model is computed according to (3.1.25). In the next section, we will look in more depth at the algorithm and study the number of floating point operations.

We remark that we can use a similar implementation for the other equation-based metapopulation model, Model B, by implementing the respective equations of (3.1.9), but we will not go into detail here.

4.3 Theoretical complexity analysis

Before comparing results and timings of both models, we will, in this section, study the computational complexity, considering the number of floating point operations for a numerical solver with fixed step size. We will analyze the implementations of both models in terms of floating point operations (FLOP) and follow the approach of [22]. As already described there, this approach is simplified, since the number of operations does not correspond to the number of cycles on a real machine, but still gives an idea of the complexity of both algorithms. For a model $\mathcal{M}$, we denote its complexity by $\mathfrak{F}(\mathcal{M})$. Frequently, the term FLOPs is used to denote the number of floating point operations per second. Therefore, in order to avoid confusion, we will use the term FLOP to denote the plural form of floating point operation.

Complexity of Model C Algorithm 1 shows the pseudocode for solving Model C with the forward Euler method and provides the number of FLOP in each line, which we will determine in the following. For reasons of readability, we will only consider a single age group, in which case the model simplifies to (3.1.20) which we denoted by $\mathcal{M}^{\mathrm{C}}_{1,M}$. Fig. 4.2 shows the resulting flows between the compartments for a region $\mathcal{P}_n$, $n \in \{1, \dots, M\}$.

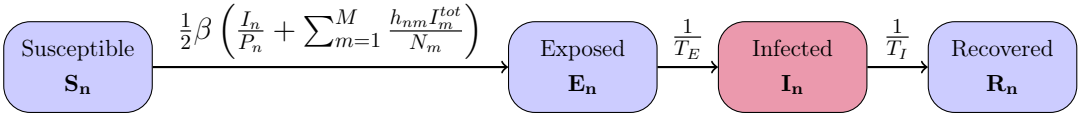


Figure 4.2: **Schematic representation of the flows in Model C for a region $\mathcal{P}_\mathbf{n}$.** The values of the flows for each transition are shown above the respective arrow. The corresponding model equations are found in (3.1.20). The compartment in which individuals are infectious is highlighted in red.

In order to provide a simplified description of the implementation, Algorithm 1 simplifies certain components which become more complicated for multiple age groups. However, it summarizes all computations that are performed for solving the model when considering only one age group. The algorithm uses several parameters, most of which are part of the parameter set of the model. For the sake of a simplified presentation, we present those parameters that are used for the computation part of the simulation as input parameters. The regions are given by the struct `Region` which serves as a shorthand notation to index the population and all `CustomIndexArray`s. The commuting strength matrix H and the population sizes after commuting N have been described above. The parameters governing the transmission ρ and ϕ are also part of the parameter set. The intermediate steps and the results are saved in a `TimeSeries`. In the algorithm, we will denote the vector of compartment sizes in the k-th time step with y^k. This vector contains the sizes of each compartment S, E, I, R for every region and is of the form

$$y^k = \left(S_1^k,\ \ldots,\ S_M^k,\ E_1^k,\ \ldots,\ E_M^k,\ I_1^k,\ \ldots,\ I_M^k,\ R_1^k,\ \ldots,\ R_M^k\right)^T. \tag{4.3.1}$$

With S^k we further refer to the subvector of y^k which contains the sizes of Susceptible compartments for all regions, i.e.:

$$S^k = \left(S_1^k,\ \ldots,\ S_M^k\right)^T. \tag{4.3.2}$$

The vectors E^k, I^k and R^k are defined accordingly. The initial conditions for all regions are given by y^0. For the simulation, we denote the start time t_0 and end time t_{max}, as well as a step size h for the Euler method. We assume w.l.o.g. that $(t_{max} - t_0)/h \in \mathbb{N}$.

The flows between compartments are stored in a vector of size $3M$ which is of the form

$$F = (F_{S_1\to E_1},\ \ldots,\ F_{S_M\to E_M},\ F_{E_1\to I_1},\ \ldots,\ F_{E_M\to I_M},\ F_{I_1\to R_1},\ \ldots,\ F_{I_M\to R_M})^T \tag{4.3.3}$$

and, similar to the vector of the compartment sizes, $F_{S\to E}$ denotes the subvector with entries $F_{S_n\to E_n}$, $n = 1,\ldots,M$. Once all flows are computed, we can combine them to obtain the derivative at the current time step in each compartment.

In the following, we look at the numerical procedure and count the FLOP that are needed for the computations. Due to the simplified representation, some small helper

Algorithm 1: Numerical solution of the equation-based metapopulation model.
`(2M²+8M) mult., 4M div., (2M²+7M) add., 3M sub.  = 4M²+22M/time step`

Input: $(\mathcal{P}_n)_{n\in\{1,\dots,M\}}$, H, N, ρ, ϕ, y^0, t_0, t_{max}, h

1	$t \leftarrow t_0$	
2	$k \leftarrow 0$	
3	**while** $t < t_{max}$ **do**	
4	$I_n^{tot} \leftarrow 0$	
5	**for** *each region* $\mathcal{P}_n$ **do**	
6	**for** *each region* $\mathcal{P}_m$ **do**	
7	$I_n^{tot} \leftarrow I_n^{tot} + h_{mn} I_m^k$	`1 mult., 1 add.`
8	$I_n^{tot} \leftarrow I_n^{tot}/N_n$	`1 div.`
9	$\widetilde{I} \leftarrow H \cdot I^{tot}$	`M² mult.  + M(M-1) add.`
10	**for** *each region* $\mathcal{P}_n$ **do**	
11	$F_{S_n \to E_n} \leftarrow 0.5\rho\phi S_n (I_n^k/P_n + \widetilde{I}_n)$	`4 mult., 1 div., 1 add.`
12	$F_{E_n \to I_n} \leftarrow E_n^k/T_E$	`1 div.`
13	$F_{I_n \to R_n} \leftarrow I_n^k/T_I$	`1 div.`
14	$y^{k+1} \leftarrow 0$	
15	$S^{k+1} \leftarrow S^{k+1} - F_{S\to E}$	`M sub.`
16	$E^{k+1} \leftarrow E^{k+1} + F_{S\to E} - F_{E\to I}$	`M add., M sub.`
17	$I^{k+1} \leftarrow I^{k+1} + F_{E\to I} - F_{I\to R}$	`M add., M sub.`
18	$R^{k+1} \leftarrow R^{k+1} + F_{I\to R}$	`M add.`
19	$y^{k+1} \leftarrow y^k + h\, y^{k+1}$	`4M mult., 4M add.`
20	$t \leftarrow t + h$	`[1 add.]`
21	$k \leftarrow k + 1$	`[1 add.]`

calculations appear in the pseudocode which do not appear in the implementation. We will therefore refrain from counting these operations, which are mostly FLOP of order $\mathcal{O}(1)$.

In line 4 to 13 of Algorithm 1, we compute the flows between the compartments. For this purpose, in line 4-7, we compute the total number of infectious individuals in each region, meaning all infected individuals of the respective region which stay there and the number of infected individuals which commute to this region. This corresponds to equation (3.1.16) of Section 3.1. In line 7 we count 1 multiplication and 1 addition which are called M^2 times each time step h. In line 8, we divide the number of infected individuals by the population size N_n of region $\mathcal{P}_n$ after commuting, resulting in the probability for meeting an infectious individual, given there is a contact, in this region. Since this is done for every region, we get M divisions. Line 9 computes the product of the matrix H with the vector I^{tot}, whose result is stored in a vector $\widetilde{I}$. As a matrix-vector product, line 9 corresponds to M^2 multiplications and $M(M-1)$ additions. Lines 10 to 18 compute the values of the differential equations at the current time t, using the intermediate results from before. As mentioned, we use a type of model that computes the flows before computing the sizes of each compartment at the current time. We begin with computing the flow from Susceptible to Exposed. For the n-th entry of $\widetilde{I}$, we observe:

$$\widetilde{I}_n = \sum_{m=1}^{M} \frac{h_{nm} I_m^{tot}}{N_m}. \tag{4.3.4}$$

Comparing (4.3.4) and the flow from Susceptible to Exposed in Model C (3.1.20), compare also Fig. 4.2, we obtain line 11. We count 1 division and 1 addition inside the parentheses, and 4 multiplications outside the parentheses. The computation of the other flows from the Exposed to Infected and Infected to Recovered compartments cost 1 division each (lines 12-13). Lines 11-13 are called for every region which leads to $4M$ multiplications, $3M$ divisions, and M additions. Next, we can compute the derivatives in the current time step in line 14-18, combining the values of the flows. We use the vector for the next intermediate step y^{k+1} as a temporary space to store the derivative and initialize it with 0. Now every flow is subtracted from a compartment subvector if it acts as an outflow to this compartment and added if it acts as an inflow. The Susceptible compartment has only one outflow, i.e., $F_{S\to E}$ and no inflow. Therefore, we only subtract the M-dimensional flow vector from S^{k+1}, which gives M subtractions. For the Exposed and Infected compartment we have an inflow and outflow each, yielding $2M$ additions and $2M$ subtractions. Furthermore, we get M additions from adding the flow from Infected to Recovered to the subvector of the Recovered compartment. The result for the next time step is computed according to the Euler method in line 18 which causes $4M$ multiplications and $4M$ additions.

For a single iteration of the while-loop, ie., a single time step this yields:

- $2M^2 + 8M$ multiplications,
- $4M$ divisions,
- $2M^2 + 7M$ additions,
- $3M$ subtractions.

Solving the model with a constant step size h requires $(t_{max} - t_0)/h$ steps. In total, the algorithm requires

$$\mathfrak{F}\left(\mathcal{M}^{\mathrm{C}}_{1,M}\right) = (4M^2 + 22M)\frac{(t_{max} - t_0)}{h} \tag{4.3.5}$$

FLOP.

Complexity of Model D We now compare the previously computed complexity to the complexity of the graph-based Model D with $J \in \mathbb{N}$ age groups and $M \in \mathbb{N}$ regions. In order to compare the complexity of both models, we also need to take a look at the implementation of the latter.

The implementation of a graph in the software MEmilio [21] is straightforward: A `Graph` is a templatized class that consists of two vectors of `Node`s and `Edge`s. Overall, the implementation is highly templatized, and we will only explain the specifications for our applications. Every node consists of an ID and a simulation of type `Simulation` which we have already seen in the context of Model A in Section 4.1 and which connects the model and the numerical method that is used. This matches the idea of the model that contains one (sub)model per region. The (sub)model is specified by the `Model` class of the *SEIR* model which was described in Section 4.1. While it supports stratification by age, we will compute and provide the detailed number of FLOP for a model with only one age group. The edges of the graph consist of a start and end node represented by the IDs and a property of type `MobilityEdge` which provides the infrastructure for the multi-edges described in Section 3.2. Every multi-edge describes how many individuals of each age group and in each compartment commute from the start to the end node per day. The software stores this number after commuting, in order to be able to estimate their infection state when returning, inside the edge.

The numerical procedure for solving the model is outlined as pseudocode in Algorithm 2. Similar to Model C, most of the parameters used are member variables in various classes. For the sake of a simple presentation, they are visualized as input parameters to the algorithm. The graph that models the metapopulation in a graph as described in Section 3.2 consists of M nodes. Since we did not assume that all regions are connected to another, the graph has $L \leq M^2$ multi-edges. Similar to the other models, the results are stored in a `TimeSeries` contained in the `Simulation` of each node. In

the algorithm, we denote the current result, i.e., the current compartment sizes of region $\mathcal{P}_n$ by y_n^k. Therefore, it is of the form

$$y_n^k = \begin{pmatrix} S_n^k \\ E_n^k \\ I_n^k \\ R_n^k \end{pmatrix}. \tag{4.3.6}$$

Additionally, the `Simulation` object is equipped with a start time t_0 and an end time t_{max}. In Section 3.2 we already described the process of solving the model roughly. Every model is advanced for a period of time when commuting is performed which leads to two levels of time steps: One that determines the intervals between commuting and returning which we denote by h_c, and one that is used for the integrator of each model denoted by h_s. We fix $h_c = 0.5$ such that individuals travel once a day for half a day, as described before. We assume w.l.o.g. that $h_c/h_s \in \mathbb{N}$ and $(t_{max} - t_0)/h_c \in \mathbb{N}$.

At the beginning of each iteration, the models are advanced from current time t to $t + h_c$ independently. This is line 5 of Algorithm 2 where we use an Euler method applied to the system of differential equations of the *SEIR* model. For a current state $y_n^k = \left(S_n^k, E_n^k, I_n^k, R_n^k\right)$ it computes the next states up to $t + h_c$. The state at time $t + h_c$ is denoted by y_n^{k+1}. The detailed description of the computations is presented in Algorithm 3.

The rough structure of Algorithm 3 is very similar to that of the Model C, only the calculation of the equations is much simpler, as every model handles only one region. In line 4 we compute the flow from the Susceptible to Exposed compartment, where the total population size is inferred by the sum over all compartments. Line 4 results in 3 multiplications, 1 division and 3 additions. For the other flows in line 5 and 6 we count 1 division each. Since the flows are only one-dimensional, we get 3 subtractions and 3 additions for the combination of the flows in total, compare lines 8-11. The Euler step in line 12 costs 4 multiplications and 4 additions due to the four compartments. All intermediate results are stored in the result `TimeSeries` for a detailed analysis afterward. However, we omit them in Algorithm 2 for the sake of a simplified representation. In total, we need $(t_{max} - t_0)/h_s$ time steps, which leads us to

- $7(t_{max} - t_0)/h_s$ multiplications,
- $3(t_{max} - t_0)/h_s$ divisions,
- $10(t_{max} - t_0)/h_s$ additions,
- $3(t_{max} - t_0)/h_s$ subtractions

for the Euler function. In Algorithm 2 this function is called in line 5 and in line 11, each time with $t_0 = t$ and $t_{max} = t + h_c$. We now continue in Algorithm 2. After the time

Algorithm 2: Numerical solution of the hybrid graph model
$11L$ + $14h_ch_s^{-1}M$ mult., $3L$ + $6h_ch_s^{-1}M$ div., $18L$ + $20h_ch_s^{-1}M$ add., $11L$ + $6h_ch_s^{-1}M$ sub.

Input: Graph = (nodes ($\mathcal{P}_n$), multi-edges (e_{nm})), t_0, t_{max}, h_s, $h_c = 0.5$

$t \leftarrow t_0$
$k \leftarrow 0$
while $t < t_{max}$ **do**
  for *each node* $\mathcal{P}_n$ **do**
    Advance the models to $t + h_c$:
    $y_n^{k+1} \leftarrow \text{Euler}(y_n^k, h_s, t, t + h_c)$
    $7h_ch_s^{-1}$ mult., $3h_ch_s^{-1}$ div., $10h_ch_s^{-1}$ add., $3h_ch_s^{-1}$ sub.
  $t \leftarrow t + h_c$ [1 add.]
  $k \leftarrow k + 1$ [1 add.]
  for *each multi-edge* e_{nm} **do**
    Exchange commuters according to commuting factors:
    $y_n^k, y_m^k, y_{nm}^k \leftarrow \text{Commute}(y_n^k, y_m^k, e_{nm})$ 4 mult., 4 add., 4 sub.
  for *each node* $\mathcal{P}_n$ **do**
    Advance the models to $t + h_c$:
    $y_n^{k+1} \leftarrow \text{Euler}(y_n^k, h_s, t, t + h_c)$
    $7h_ch_s^{-1}$ mult., $3h_ch_s^{-1}$ div., $10h_ch_s^{-1}$ add., $3h_ch_s^{-1}$ sub.
  $t \leftarrow t + h_c$ [1 add.]
  $k \leftarrow k + 1$ [1 add.]
  for *each multi-edge* e_{nm} **do**
    Euler step to estimate the most likely infection states of returning commuters:
    $y_n^k, y_m^k \leftarrow \text{Returns}(y_{nm}^{k-1}, y_m^{k-1}, y_n^k, y_m^k)$
    7 mult., 3 div., 14 add., 7 sub.

Algorithm 3: Euler-SIR(y^0, h_s, t_0, t_{max})
$7h_ch_s^{-1}$ `mult.`, $3h_ch_s^{-1}$ `div.`, $10h_ch_s^{-1}$ `add.`, $3h_ch_s^{-1}$ `sub.`

Input: y^0, h_s, t_0, $t_{max} = t_0 + h_c$

$t \leftarrow t_0$
$k \leftarrow 0$
while $t < t_{max}$ **do**
  $F_{S\to E} \leftarrow \rho\phi S^k I^k/(S^k + E^k + I^k + R^k)$ — `3 mult., 1 div., 3 add.`
  $F_{E\to I} \leftarrow E^k/T_E$ — `1 div.`
  $F_{I\to R} \leftarrow I^k/T_I$ — `1 div.`
  $y^{k+1} \leftarrow 0$
  $S^{k+1} \leftarrow S^{k+1} - F_{S\to E}$ — `1 sub.`
  $E^{k+1} \leftarrow E^{k+1} + F_{S\to E} - F_{E\to I}$ — `1 add., 1 sub.`
  $I^{k+1} \leftarrow I^{k+1} + F_{E\to I} - F_{I\to R}$ — `1 add., 1 sub.`
  $R^{k+1} \leftarrow R^{k+1} + F_{I\to R}$ — `1 add.`
  $y^{k+1} \leftarrow y^k + h_s\, y^{k+1}$ — `4 mult., 4 add.`
  $t \leftarrow t + h_s$ — `[1 add.]`
  $k \leftarrow k + 1$ — `[1 add.]`

Output: y^k

step is updated, commuting is applied for every multi-edge of the graph. The process of commuting is outlined in Algorithm 4.

The algorithm uses the populations in each compartment of two regions y_n^k and y_m^k and the respective multi-edge between those two e_{nm}. Every multi-edge consists of four single edges, one for each compartment, as we consider only a single age group. We interpret e_{nm} as a four-dimensional vector, where each entry gives the share of commuting individuals from region $\mathcal{P}_n$ to $\mathcal{P}_m$ for the corresponding compartment. First, the number

Algorithm 4: Commute(y_n^k, y_m^k, e_{nm})
`4 mult., 4 add., 4 sub.`

Input: y_n^k, y_m^k, e_{nm}

1 $y_{nm}^k \leftarrow e_{nm} y_n^k$ — `4 mult.`
2 $y_n^k \leftarrow y_n^k - y_{nm}^k$ — `4 sub.`
3 $y_m^k \leftarrow y_m^k + y_{nm}^k$ — `4 add.`

Output: y_n^k, y_m^k, y_{nm}^k

of commuters from region $\mathcal{P}_n$ is determined according to the population and the share of commuting individuals e_{nm}. The multiplication in line 1 of Algorithm 4 is performed componentwise, so we determine the number of commuting individuals per compartment

which is denoted by y_{nm}^k and will be needed later to determine the infection states of the returning individuals. Therefore, y_{nm}^k is returned by the function. Once determined, we subtract the numbers of commuters from the population of region $\mathcal{P}_n^k$ and add it to the population of region $\mathcal{P}_m^k$. Overall, we have 1 componentwise multiplication, 1 subtraction and 1 addition of four-dimensional arrays, which leads to

- 4 multiplications,
- 4 additions,
- 4 subtractions

for the commuting in line 7 of Algorithm 2.

After commuting, the models are advanced for another $h_c = 0.5$ days, for which we have already counted the FLOP. Note that one iteration of the while-loop in Algorithm 2 describes two time steps, since individuals commute after the first and return after the next. In Section 3.2 we already outlined the necessity of the auxiliary step when returning the commuters. We perform an Euler step where the commuting population from the commuting step before y_{nm}^k contacts the population of its target location right after commuting, and update the populations of source and target location afterward. This is detailed in Algorithm 5.

Algorithm 5: Returns(y_{nm}^{k-1}, y_m^{k-1}, y_n^k, y_m^k)

		7 mult., 3 div., 14 add., 7 sub.
	Input: y_{nm}^{k-1}, y_m^{k-1}, y_n^k, y_m^k	
1	$\beta \leftarrow \rho\phi/(S_m^{k-1} + E_m^{k-1} + I_m^{k-1} + R_m^{k-1})$	1 mult., 1 div., 3 add.
2	$F_{S_n \to E_n} \leftarrow \beta S_{nm}^{k-1} I_m^{k-1}$	2 mult.
3	$F_{E_n \to I_n} \leftarrow E_{nm}^{k-1}/T_E$	1 div.
4	$F_{I_n \to R_n} \leftarrow I_{nm}^{k-1}/T_I$	1 div.
5	$y_{nm}^k \leftarrow 0$	
6	$S_{nm}^k \leftarrow S_{nm}^k - F_{S_n \to E_n}$	1 sub.
7	$E_{nm}^k \leftarrow E_{nm}^k + F_{S_n \to E_n} - F_{E_n \to I_n}$	1 add., 1 sub.
8	$I_{nm}^k \leftarrow I_{nm}^k + F_{E_n \to I_n} - F_{I_n \to R_n}$	1 add., 1 sub.
9	$R_{nm}^k \leftarrow R_{nm}^k + F_{I_n \to R_n}$	1 add.
10	$y_{nm}^k \leftarrow y_{nm}^{k-1} + h_c\, y_{nm}^k$	4 mult., 4 add.
11	$y_n^k \leftarrow y_n^k + y_{nm}^k$	4 add.
12	$y_m^k \leftarrow y_m^k - y_{nm}^k$	4 sub.
	Output: y_n^k, y_m^k	

The factor β of the first flow is calculated by multiplying the transmission probability with the contact rate and dividing it by the size of the population at the target location

right after commuting. This population size is determined by the sum over all compartments, which requires 3 additions. The flow from Susceptible to Exposed is then computed by multiplying the factor β from the line before with the susceptible commuters and the number of infectious individuals after commuting, i.e., I_m^{k-1}. Lines 3-10 describe the remaining part of the Euler step but do not depend on the contact population. As this is just repetition of what we have already seen in Algorithm 3, we will not go into more detail here. During this Euler step, the returning population in each compartment y_{nm}^k is determined. Since the auxiliary step is just an approximation of the infection states of the commuters, it is possible that the calculated numbers of returning individuals in a compartment exceed the population of this compartment at the target location. To ensure non-negative compartments in that case, supernumerary individuals are subtracted from the biggest compartment. However, given that this does not occur in every step of the process, it is omitted in our complexity analysis. Next, the populations at the prior starting and target locations are updated through adding the returning individuals to the starting location and subtracting them from the target location. Since we have four-dimensional vectors, this requires 4 additions and 4 subtractions. For the return process, including the estimation of the infection states of the commuters, we have

- 7 multiplications,
- 3 divisions,
- 14 additions,
- 7 subtractions.

To determine the number of FLOP of Algorithm 2 we have to combine the number of FLOP of the individual components. Lines 5 and 11 are called for each region, which is why we multiply these FLOP with the number of regions M. The commuting and returning is applied to every multi-edge of the graph, so we multiply those FLOP with the number of edges L. For Algorithm 2, including all components, we have:

- $11L + 14h_c h_s^{-1} M$ multiplications,
- $3L + 6h_c h_s^{-1} M$ divisions,
- $18L + 20h_c h_s^{-1} M$ additions,
- $11L + 6h_c h_s^{-1} M$ subtractions.

Algorithm 2 makes $(t_{max} - t_0)/(2h_c)$ time steps, as the time is updated twice per iteration of the while-loop. Using $h_c = 0.5$, this yields:

$$\mathfrak{F}\left(\mathcal{M}_{1,M}^{\mathrm{D}}\right) = \left(43L + 23h_s^{-1}M\right)(t_{max} - t_0). \tag{4.3.7}$$

In the worst case, all regions are connected to every other region, which results in $L = M^2$ edges and thus the number of FLOP is:

$$\mathfrak{F}\left(\mathcal{M}^{\mathrm{D}}_{1,M}\right) = \left(43M^2 + 23h_s^{-1}M\right)(t_{max} - t_0). \tag{4.3.8}$$

Comparison of both complexities To compare both models based on the number of FLOP they require, we assume that the step sizes for the Euler methods in both models agree, i.e., $h_s = h$. We fix the step size $h = 0.1$ and a simulation time $(t_{max} - t_0)$ of one day. Furthermore, we assume $L = M^2$, i.e., all regions are connected to each other. In

M	$\mathfrak{F}\left(\mathcal{M}^{\mathrm{C}}_{1,M}\right)$ $10(4M^2 + 22M)$	$\mathfrak{F}\left(\mathcal{M}^{\mathrm{D}}_{1,M}\right)$ $\left(43M^2 + 230M\right)$
2	620	674
5	2 120	2 327
10	6 220	6 802
50	111 020	120 002
100	422 020	455 002
400	6 488 020	6 980 002

Table 4.1: **Theoretical number of FLOP for Model C and Model D.** Exemplary number of FLOP for different numbers of regions, assuming $t_{max} - t_0 = 1$ and a step size of $h = 0.1$ for the Euler method. For Model D it is additionally assumed that all regions are connected to another.

Table 4.1 we give the calculated values for the number of FLOP for a selection of numbers of regions. We see that the equation-based Model C requires slightly, i.e., 7-10% fewer operations than the graph-based Model D for the chosen numbers of regions. For more insights, we provide Fig. 4.3. Here, we visualize the computed number of FLOP as a function of the number of regions for different values of step sizes h. We see that the step size h has no significant effect on the number of FLOP of Model D, while the opposite is true for Model C. With a step size of $h = 0.05$, Model C requires even more FLOP compared to Model D.

Another observation we make is about the implementation of the age groups which we omitted in the calculation of the number of FLOP in both cases. We refrain from counting the exact number of FLOP for the case of multiple age groups, but provide the order of the algorithms based on their computational complexity. For a population with a single age group, the calculation of the total number of infectious individuals at a region $\mathcal{P}_m$ according to (3.1.16) requires a number of FLOP of order $\mathcal{O}(M^2)$, compare lines 5-8 of Algorithm 1. Now consider Model C (3.1.25) for a population divided into $J \in \mathbb{N}$ age groups and $M \in \mathbb{N}$ regions. The number of infectious individuals in each region has to

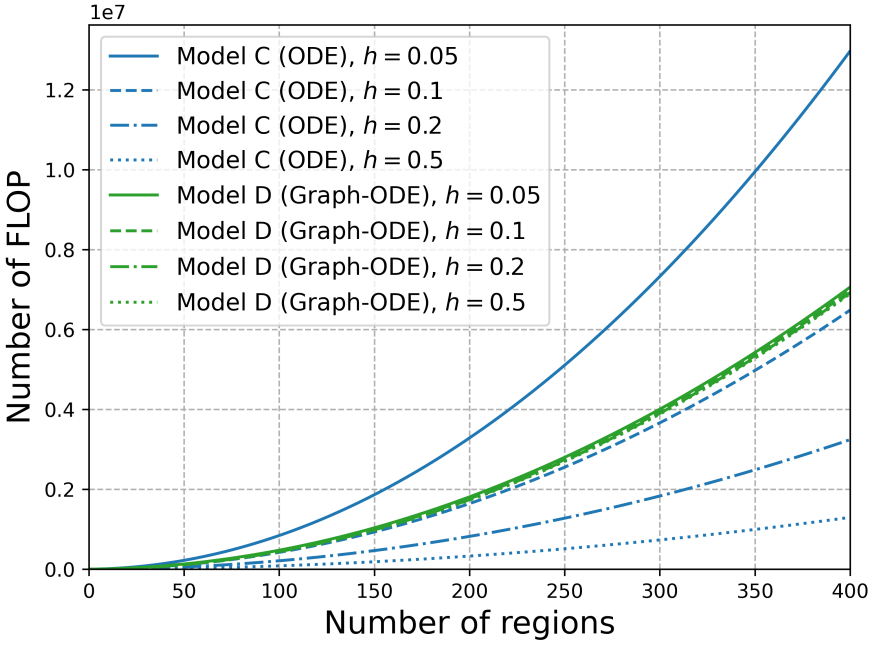


Figure 4.3: **Theoretical number of FLOP for Model C and Model D.** Calculated number of FLOP with respect to the number of regions for different step sizes h, assuming $t_{max} - t_0 = 1$. For Model D it is additionally assumed that all regions are connected to another.

be calculated for every age group. Thus, lines 5-8 are calculated J times and we obtain a complexity of order $\mathcal{O}(JM^2)$. To calculate the number of transmissions in unit time in an age group i, we have to consider the number of contacts an individual from the age group i has with individuals from this or all other age groups j, $j \in \{1, \dots, J\}$, obtaining a complexity of $\mathcal{O}(J^2)$. As the flow has to be calculated for every region, we therefore obtain a computational complexity of order $\mathcal{O}(J^2M)$ for the calculation of the flow from Susceptible to Exposed in the implementation of the age-resolved Model C. In total, we conclude a computational complexity of order

$$\mathfrak{F}\left(\mathcal{M}^{\mathrm{C}}_{J,M}\right) = \mathcal{O}\left(JM^2 + J^2M\right). \tag{4.3.9}$$

Recall that in Model D for a population which is divided into $J \in \mathbb{N}$ age groups, every multi-edge holds a single edge for each combination of infection state and age group. Hence, the commuting and returning of individuals has to be performed for each age group, resulting in a complexity of $\mathcal{O}\left(JM^2\right)$. Similar to Model C, the computation of the flow from Susceptible to Exposed in a region $\mathcal{P}_n$ requires the consideration of all combinations of age groups. As this flow has to be computed for every region, we obtain a complexity of $\mathcal{O}\left(J^2M\right)$ for the calculation of the flows. For the solution of the

age-resolved version of Model D, we conclude a computational complexity of

$$\mathcal{F}\left(\mathcal{M}^{\mathrm{D}}_{J,M}\right) = \mathcal{O}\left(JM^2 + J^2M\right). \tag{4.3.10}$$

Consequently, for $J \in \mathbb{N}$ age groups, Model C and Model D require a number of FLOP of the same order.

The graph-based model offers the advantage of eliminating multi-edges with minimal weight, and thus reducing the computational effort. While this is not provided in the current implementation of the equation-based model, a similar effect could be achieved by implementing a sparse matrix for the commuting strengths.

This chapter's presentation of the implementations of both models has established the foundation for a comparative analysis of the models through numerical simulations.

5 Numerical simulations

We now compare the models derived in previous chapters based on numerical simulations. After analyzing the computational complexities from a theoretical point of view in the previous chapter, we examine the models with regard to their runtime and behavior for an adaptive solver. We also examine whether the modifications for Model C have yielded improvements over the basis Model B.

To simulate a scenario of an infectious disease outbreak, we have to equip the models with meaningful parameters. The parameter values presented in this thesis are based on the outbreak of SARS-CoV-2 in Germany in 2020. However, for the comparison of the modeling approaches of mobility, these parameter values are not a primary factor. We remark that both approaches use the same social, epidemiological and mobility parameters, so we do not need to distinguish between the approaches regarding the parameters. The simulations that study an age-resolved population use the six age groups which are defined by the data from the Robert Koch Institute (RKI) [32].

For epidemiological parameters we follow [30], which provides age-resolved values for the mean time spent in each compartment and the mean transition probabilities for a slightly more complex *SECIR* model. As the authors consider additional compartments, we cannot transfer them entirely but have to consider how the parameters of both models are related. For the *SEIR* model, the mean times spent in the Exposed compartment T_{E_i} and in the Infected compartment T_{I_i} are required. For the length of the latent stage E, we adopt the values for T_{E_i}. Compared to the *SEIR* model, the *SECIR* model has one additional infectious compartment, the pre- or asymptomatic stage "Carrier". For the mean times spent in the infectious compartment of the *SEIR* model, we have to consider both compartments C and I of the *SECIR* model. Only a fraction μ_C^I of the individuals in C develop symptoms and remain infectious, whereas the rest recovers directly. On average, a person remains infectious for $T_{I_i} = \widehat{T}_{C_i} + \mu_C^I \widehat{T}_{I_i}$ days, where $\widehat{T}_{C_i}$, $\widehat{T}_{I_i}$ denote the parameters of the *SECIR* model. As the transmission probability on contact does not change with the model, we can adopt these values as well. In some simulations, the age structure of the populations may be excluded, in which case, age-independent parameters become necessary. For this, we weight the values in each age group with the relative share of the age group in the total population according to the demographic data obtained from [35]. The inferred parameters are given in Table 5.1.

For the mean number of contacts in the age-resolved population, we use a contact

Parameter	0-4	5-14	15-34	35-59	60-79	80+	Weighted Average
ρ_i	0.03	0.06	0.06	0.06	0.09	0.175	0.0733
T_{E_i}	3.335	3.335	3.335	3.335	3.335	3.335	3.335
T_{I_i}	8.0097	8.0097	8.2182	8.1158	8.0330	7.985	8.0976

Table 5.1: Age-resolved epidemiological parameters for the simulations of SARS-CoV-2 with a (metapopulation) *SEIR* model, adopted from [30]. Age-independent parameters are inferred through weighting the age-dependent parameters by the relative share of the age group in the total population according to [35].

matrix $(\phi_{ij})_{i,j\in\{1,\dots,j\}}$ representing a realistic contact pattern for Germany. The contact patterns derived from [28, 31, 10] are categorized into the following: "Home", "School", "Work" and "Other". These categories are then combined to generate a contact matrix compatible with the specified age groups. A detailed description of the matrix and how it was derived can be found in [20, p. 5]. For the age-independent simulations, we fix the mean number of contacts of $\phi = 7.95$, in accordance with [28, p. 384].

The population that is simulated depends on the purpose of our simulations and is described in each section. The same holds for the parameters concerning the spatial resolution of the model setting.

5.1 Runtime analysis with a fixed number of steps

We compare both modeling approaches in terms of computation time as a function of the number of regions. All timing runs were performed on a small-size internal cluster with 4×56 cores. In particular, we conducted all simulations on a single core of an Intel Xeon "Skylake" Gold 6132 (2.60 GHz). In order to initialize models with any number of regions, we use a synthetic population, consisting of 60 000 individuals, that is assigned to every region. To compare the results with the complexity analysis in Section 4.3, we initially study a population which is not age-stratified. All individuals are assigned to the Susceptible compartment. The infectious disease outbreak then starts with 100 individuals of one region in the Exposed compartment. For simplicity, we assume that half of each population commutes, with the commuting individuals equally distributed across all other regions, i.e., for M regions a share of $(2M)^{-1}$ commutes from a region $\mathcal{P}_n$ to another region $\mathcal{P}_m$ for all $n,\ m \in \{1,\dots,M\}$ with $n \neq m$. We simulate the spread of the disease for 20 days with the explicit Euler method presented in Listing 4.2, using a step size of $h = 0.1$. When analyzing the runtime, the initialization and setup process is excluded in order to ensure the comparability of the results. Consequently, the analyzed parts correspond to the parts we have previously analyzed in Section 4.3 with regard to

the required number of FLOP. For an appropriate runtime analysis, we perform 10 warm up runs and then perform 100 runs over which the runtime is averaged.

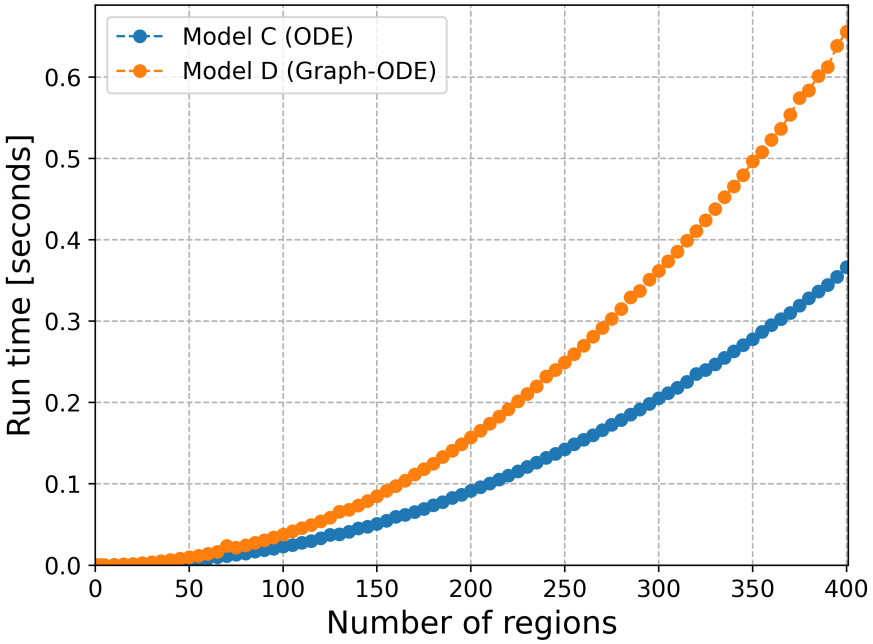


Figure 5.1: **Runtimes for varying numbers of regions with a non-age-stratified population.** The time to solution for both models is shown for a varying number of regions, with the simulated population being not divided into age groups. Both models are solved using an explicit Euler method with a step size of $h = 0.1$ and a simulation time of 20 days.

In Fig. 5.1, we compare the runtimes of both models with respect to varying numbers of regions. We conduct simulations for $M \in \{1, 5, 10, \ldots, 400\}$ many regions, and the plot shows the runtime of the equation-based Model C (blue) and the graph-based Model D (orange). In both models, the runtime increases quadratically with regard to the number of regions. In line with the lower computed number of FLOP, the equation-based model consistently performs better, with its runtime being lower across all numbers of regions. As the number of regions increases, the discrepancy between the two models also increases.

The obtained runtimes are now compared against the calculated number of FLOP, as derived in Section 4.3. Fig. 5.2 shows a dual-axis plot, where the left y-axis represents the mean runtimes and the right y-axis shows the calculated number of FLOP for a simulation time of 20 days. The x-axis corresponds to the number of regions in both models. Both the calculated numbers of FLOP and the runtimes follow a quadratic function between 1 and 400 regions. While the number of FLOP for the graph-based Model D increases approximately 1.075 as fast as the number of FLOP for the equation-based Model C, we

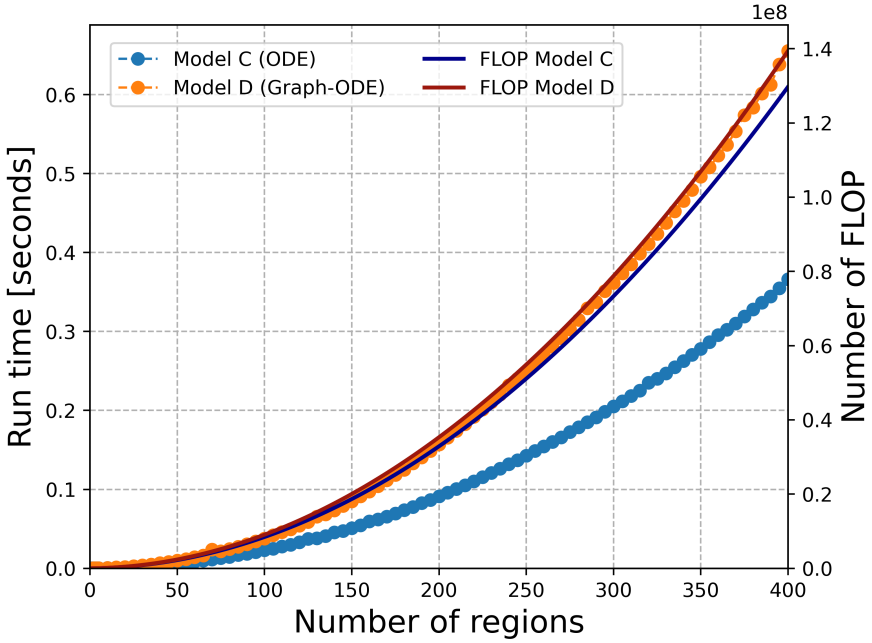


Figure 5.2: **Comparison of theoretical number of FLOP and runtimes for non-age-stratified population.** On the left y-axis, we show the mean runtime for the numerical solution of both models, averaged over 100 runs. On the right y-axis, we show the calculated number of FLOP required for the numerical solution of both models derived in Section 4.3. The x-axis corresponds to the number of regions in both models. While the calculated numbers of FLOP diverge only slightly, the discrepancy in the runtimes is significantly higher.

observe that the ratio of the runtimes is significantly higher. The ratio of the runtime for Model D to the runtime of Model C stabilizes at approximately 1.75. At 400 regions, the calculated numbers of FLOP exhibit a slight divergence, with a relative difference of approximately 7.2%. On the contrary, the runtimes diverge significantly, with a relative difference of approximately 56.6% for 400 regions. This observation indicates that the theoretical complexity analysis in Section 4.3 gives an idea about the complexity of the algorithms but does not allow drawing conclusions regarding the performance of both models. The origin of this discrepancy should be discussed in future research. An examination of the memory consumption in each model and a performance analysis using tools such as *Likwid* [11] could provide useful insights. Preliminary investigations indicate that the graph-based Model D requires a higher bandwidth than Model C. To draw definitive conclusions, a thorough investigation is necessary.

In Section 4.3, we already remarked that the introduction of age groups to the models does not lead to a significant change in the complexity. In both models, we concluded a

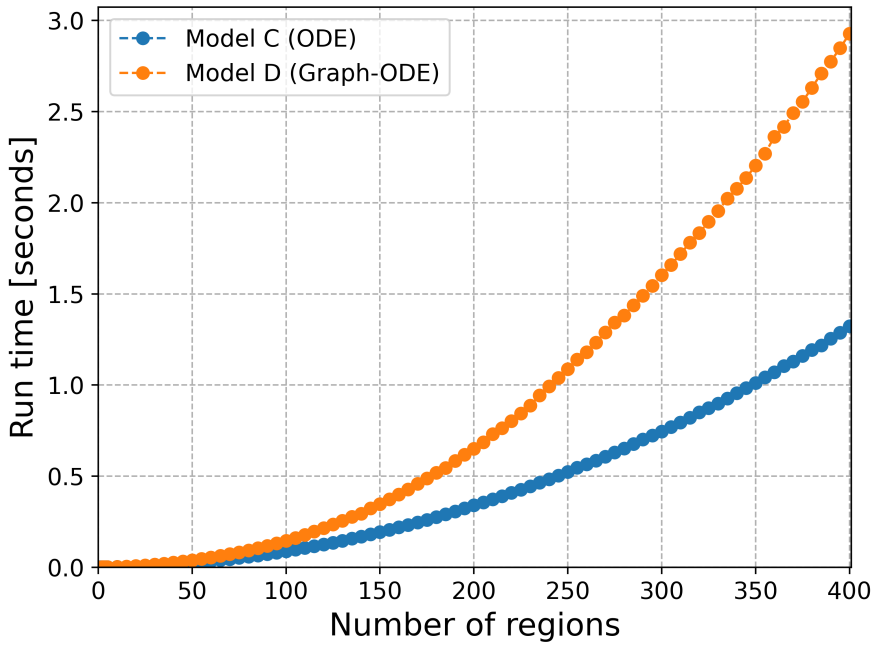


Figure 5.3: **Runtimes for varying numbers of regions with an age-stratified population.** The time to solution for both models is shown for a varying number of regions, with the simulated population being divided into six age groups. Both models are solved using an explicit Euler method with a step size of $h = 0.1$ and a simulation time of 20 days.

complexity of

$$\mathcal{F}\left(\mathcal{M}^{\mathrm{C}}_{J,M}\right) = \mathcal{F}\left(\mathcal{M}^{\mathrm{D}}_{J,M,M^2}\right) = \mathcal{O}\left(JM^2 + J^2M\right). \tag{5.1.1}$$

Therefore, we expect the runtimes of both models to show a similar pattern for an age-resolved population as for the non-age-stratified population. For this reason, we divide the population from before into six age groups, with 10 000 individuals each. Fig. 5.3 shows the mean runtime in seconds for $M \in \{1, 5, 10, \ldots, 400\}$ regions, again averaged over 100 runs. With an increasing number of regions, Model C takes up to 1.3 seconds per run, whereas Model D has a maximum of approximately 3 seconds per run. Accordingly, for a simulation of an age-resolved population, the equation-based model performs better with regard to the runtime, across all numbers of regions until 400.

5.2 Impact and possibilities of the numerical solver

So far, we restricted ourselves to solving both models with an Euler method to ensure an equal number of steps. A method with a higher order and an adaptive step size control allows a more efficient and accurate solution to the model, so as a next comparison, the

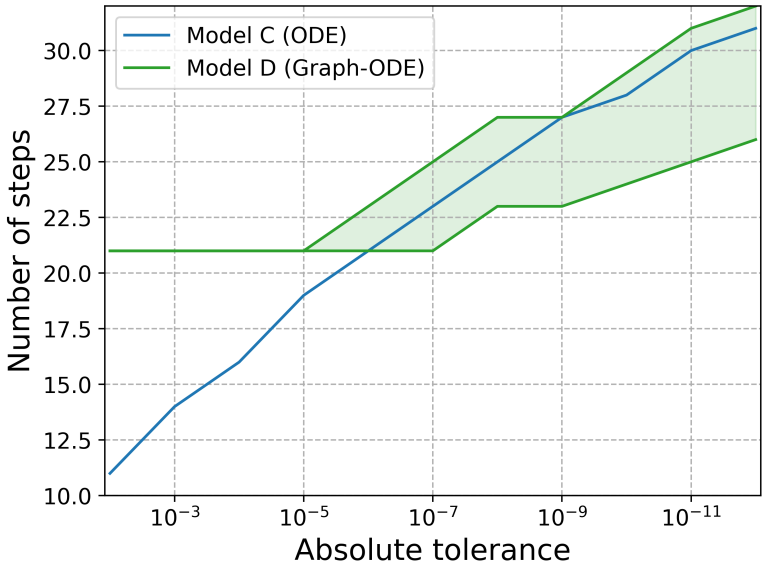


Figure 5.4: **Required number of steps of an adaptive solver.** Number of steps a Runge-Kutta Cash-Karp 5(4) solver performs for a simulation of 100 regions and a simulation time of 10 days with respect to the absolute tolerance for the step size control. For the graph-based model, the plot shows the minimum and maximum number of steps across all regions.

behavior of such a solver on both models will be examined. We use a pre-implemented Runge-Kutta Cash-Karp 5(4) method of the library boost which is one of the standard methods for solving ODEs [1]. A key difference between both modeling approaches is the semi-discreteness of the graph-based Model D, as information has to be exchanged between regions every half simulation day. While, in theory, an adaptive solver can perform arbitrary large steps for Model C, it is forced to make steps smaller or equal to half a day for Model D. Both models are set up with 100 regions and a population of 60 000 individuals, divided into six age groups with 10 000 individuals each. All except 100 individuals of one region are assigned to the Susceptible compartment, while these 100 individuals are assigned to the Exposed compartment. The commuting strengths are given by $h_{nm} = 0.005$ for all $n, m \in \{1, \dots, 100\}$ with $n \neq m$. Each model is advanced for 10 days, with an initial step size of $h = 0.1$. The minimal step size for the solver is set to the minimum double value larger than zero for both models. The maximal step size for the ODE Model C is set to the maximum double value, while it is set to 0.5 for the graph-based Model D. We now analyze how many steps the solver performs with respect to a varying tolerance for the step size control, shown in Fig. 5.4. We recall that in the graph-based model, each region is represented by a different *SEIR* model. Consequently, the solver is advanced independently for each region and might therefore perform a different number of steps, depending on the disease dynamics at the

corresponding region. For the graph-based Model D, we therefore plot the minimum and the maximum numbers of steps of all regions (green). We observe that for a large tolerance between 10^{-2} to 10^{-5}, the solver of Model D performs exactly 21 steps in the hotspot region, as well as in all other regions. As the maximum step size is 0.5 days and the simulation time is 10 days, we conclude that the solver always chooses the maximum step size. This outcome is not unexpected given the simplicity of the example and the mean stay times of several days in each compartment. In the higher tolerance range the solver requires fewer steps for Model C than for Model D, since its maximum step size is not restricted to 0.5. For smaller tolerance values, the solver performs a similar number of steps and Model C does not appear to have an advantage. Since, for more reasonable and smaller tolerance values, both numerical models make a similar number of steps, we expect the results regarding the runtime from the previous section to be transferable to the runtime for simulations with an adaptive solver.

5.3 Basic reproduction numbers

In previous chapters, we derived the next generation matrices for Model A (2.3.1), as well as the equation-based metapopulation models, Model B (3.1.26) and Model C (3.1.25). As the solution to the eigenvalue problem yields lengthy and complex formulas, especially for the metapopulation models, we did not compute the eigenvalues analytically. Instead, we implement the next generation matrices for all three models and compute the maximum eigenvalue which represents the basic reproduction number, numerically. We now examine, how the basic reproduction number behaves for different numbers of regions. When deriving the metapopulation models, we divided a population into several subpopulations, representing the regions. The metapopulation models, Model B (3.1.26) and Model C (3.1.25), are therefore refinements of the basic Model A (2.3.1). Aggregating identical disease dynamics over all regions, we expect those models to have the same basic reproduction number. We use the age-resolved synthetic population of Section 5.1, which allows us to easily add new regions to the model. This population can be used for both metapopulation models, and, since every region has the same population sizes, the population of Model A is determined by multiplying the population of each age group with the number of regions. Note that the number of regions only affects Model A in terms of its population size. We again assume that half of each metapopulation commutes, and the target locations are equally distributed across all other locations. For each of the models, we set up the matrices F and V according to the computations we did earlier. Next, we compute the inverse of V using the *Eigen* library [12] and compute the next generation matrix FV^{-1}. We use the `Eigenvalues` module of the same library to compute the eigendecomposition of the next generation matrix and infer the eigenvalues. The greatest eigenvalue in absolute value is the desired basic reproduction number.

Number of Regions	Model A	Model B	Model C
1	4.48161	4.48161	4.48161
2	4.48161	5.60213	4.48161
5	4.48161	6.27443	4.48161
10	4.48161	6.49854	4.48161
20	4.48161	6.61059	4.48161
50	4.48161	6.67782	4.48161

Table 5.2: **Basic reproduction numbers for different numbers of regions.** Comparison of the basic reproduction numbers of the simple *SEIR* Model A (2.3.1) and the metapopulation models, Model B (3.1.26) and Model C (3.1.25), for a varying number of regions. Note that the number of regions only affects Model A in terms of population size. While Model A and C produce an equal, constant basic reproduction number for all numbers of regions, the basic reproduction number of Model B is significantly higher.

In Table 5.2, we summarize the calculated values for the basic reproduction numbers for a varying number of regions. For a single region, both metapopulation models simplify to the basic *SEIR* model, so it is evident that the basic reproduction numbers agree. As the number of regions increases, we observe that the basic reproduction number of Model A remains constant, which we also noticed in Section 2.3. While the basic reproduction number of Model C agrees with Model A for all numbers of regions, the one of Model B diverges from the basic reproduction number of Model A. This suggests that the modeled disease spreads faster in Model B than in the other models. We will investigate this further in the next section, which will address how a disease spreads from a hotspot.

5.4 Spreading from the hotspot

In this section, we simulate a scenario where an infectious disease spreads from a hotspot and compare the resulting disease dynamics of the considered metapopulation models. As a model setting, we chose the federal state of North Rhine-Westphalia (NRW) in Germany, where the patches of the metapopulation model are represented by the 52 counties in NRW. The population is set according to [35], from where we infer the number of individuals in each county by age. We divide this data into the age groups from the RKI data, in accordance with the parameters above. In the beginning, all persons are assigned to the Susceptible compartment. We will later implement a hotspot by manually exposing individuals to the virus. The commuting strength matrix $H = (h_{nm})_{n,m\in\{1,\dots,52\}}$ is inferred from the German Federal Employment Agency [8]. From there, we get the mean number of commuters between all 400 counties in Germany, rounded to the next

multiple of ten. The data is downloaded and preprocessed using the `memilio-epidata` package of the software MEmilio [21]. Since we focus on a smaller part of Germany, i.e., the federal state of North Rhine-Westphalia, we further adjust the data by discarding all rows and columns which correspond to counties outside NRW. The epidemiological parameters can be found in Table 5.1 and have been described already, as is the case for the contact pattern used.

To start the dynamics, we expose 100 individuals of the county of Gelsenkirchen to the virus and simulate the outbreak for 100 days, using the adaptive Runge-Kutta Cash-Karp 5(4) method. We compare the disease dynamics resulting from the graph-based model, Model D to the ODE metapopulation models, Model B(3.1.26) and Model C (3.1.25). In Fig. 5.5 we can see the simulated disease progression in all three models, given as the share of infected individuals in each county of NRW. In the rows of the figure we present the outcomes of Model B, Model C and Model D for a particular day, while the columns correspond to one model over simulation days 10, 30, 50, 70 and 90. For every simulation day in steps of 20 we provide the infectious shares in logarithmic scale for Model B, Model C and Model D. At the bottom of the figure, we provide a shared color bar for all models, where the color gradient shows the presence of infected individuals in each county. On day 10, we see that the disease has started to spread from the hotspot in each of the three models. As time progresses, the disease spreads across NRW, and more and more individuals become infected. Throughout the figure, we do not notice any visual difference in the results of Model C and Model D. The dynamics resulting from Model B, however, show a faster increase in infections compared to the other models.

This can be even better observed when looking at the compartment sizes aggregated over all counties, compare Fig. 5.6. The figure shows the number of individuals in each compartment across all regions, where the x-axis is given by the time in days. Compared to the other models, the resulting dynamics of Model B is characterized by a higher number of infections over a shorter period of time. After day 70, almost all individuals of the population in Model B have left the Susceptible compartment and have been exposed to the virus. The more vivid dynamics of Model B can be explained by the fact that this model does not reflect the commuting in sufficient detail, see the explanation in Section 3.1.1. When deriving the ODE models in Section 3.1, we already remarked this shortcoming of that model, and it is visible in this simulation as well. Furthermore, it is supported by the basic reproduction number, which is significantly higher than the one of Model C.

Although Model C and Model D lead to similar results concerning the prognosis of the disease dynamics, they do not overlap completely. Therefore, we examine the differences between the predicted shares of infectious individuals over time, i.e., for a county we determine the difference between the predicted numbers of infectious individuals and

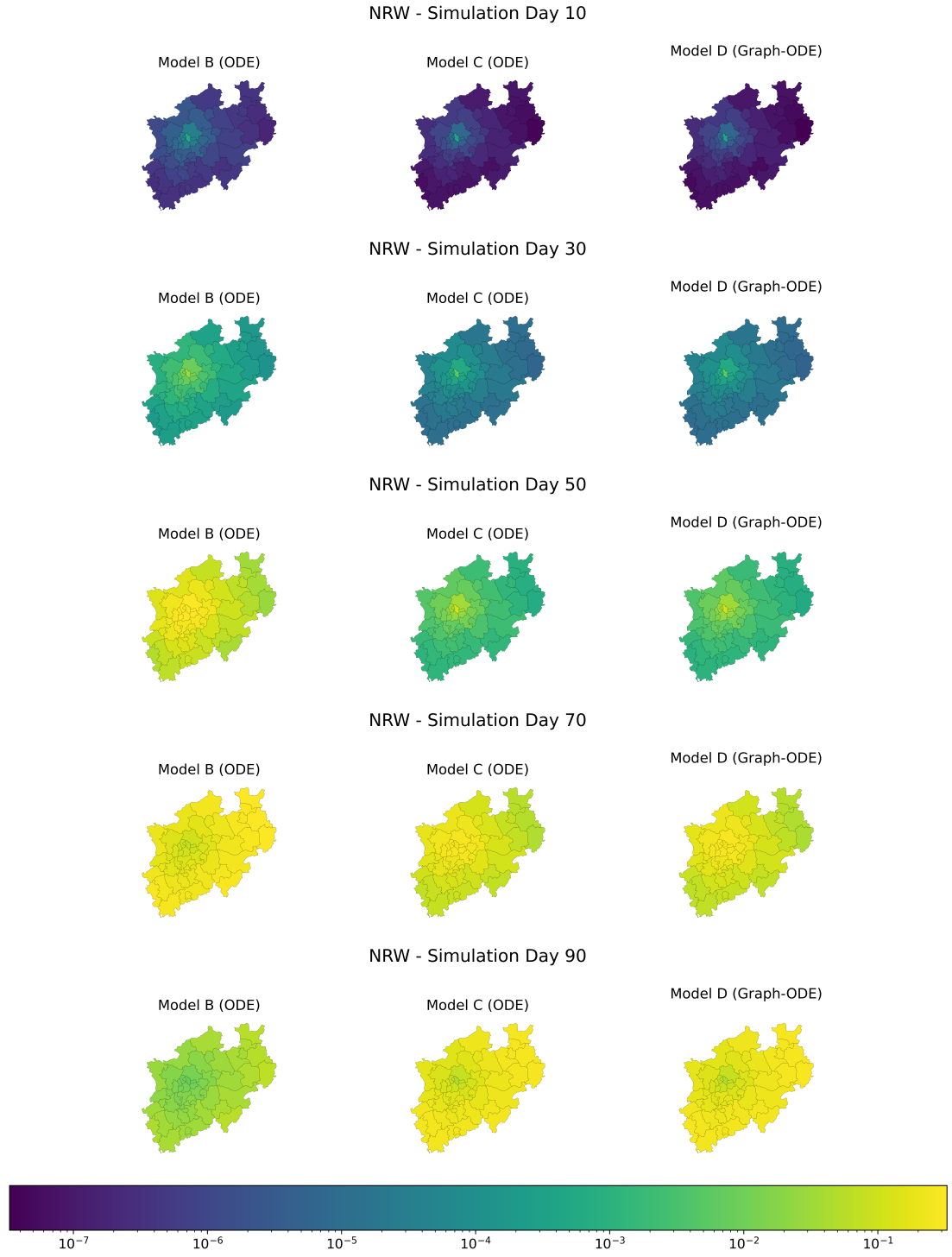


Figure 5.5: **Spatial spread of SARS-CoV-2 over 90 days in all three models.** Comparison of the spatial spread of the disease in the equation-based models, Model B (3.1.26) and Model C (3.1.25), and the graph-based model, Model D. The color gradient indicates the share of infectious individuals in each county.

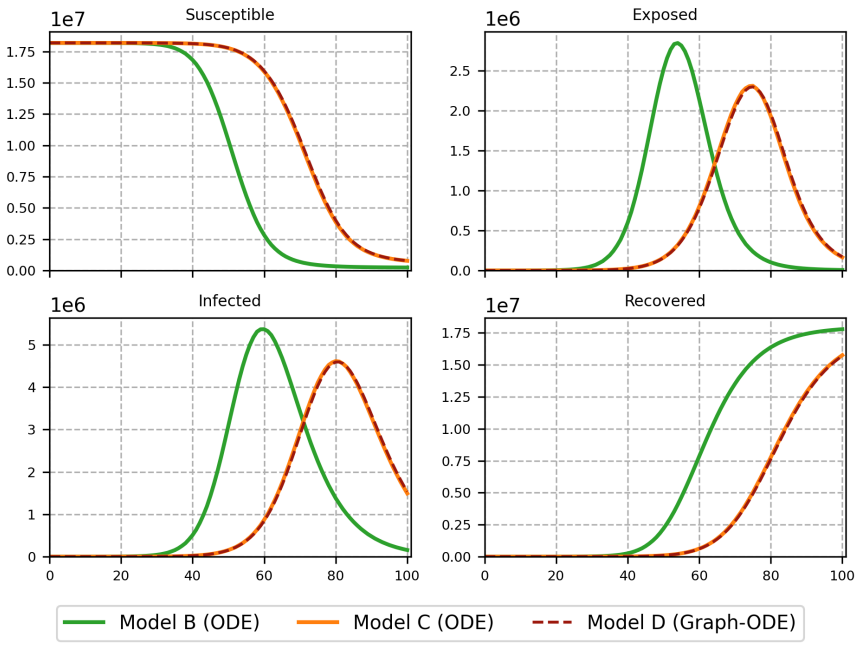


Figure 5.6: **Compartment sizes over time for Model B, Model C, and Model D.** Representation of the simulation results of all three models in terms of compartment sizes over time. For each model, the absolute number of individuals in a compartment is aggregated over all regions.

normalize by the population size of the county.

This is shown in Fig. 5.7 where the colorbar implies whether Model C or Model D predicts a higher number of infectious individuals. A value of 0, in which case both models predict the same number of infectious individuals, is represented by a white color. A red color indicates that Model C predicts a higher number of infectious individuals than Model D and a blue color vice versa. For the first 30 days, both models predict a comparable outcome, which is indicated by the predominantly colorless maps. As of days 40 and 50, the equation-based model, Model C predicts a lower number of infections in the hotspot and a slightly higher number of infections in the surrounding area. This effect is intensified on day 60, and a predominantly red map is observed on day 70. From day 80 to 100, a spread of blue color is apparent across the map, indicating that the graph-based model predicts a higher number of infectious individuals. This progression suggests that the ODE model anticipates the spread of the disease slightly earlier, which is further supported by Fig. 5.6 that shows the slightly shifted curves in the Infected compartment. Overall, the maximum observed difference in a county is 0.01% which indicates the highly comparable results.

Fig. 5.8 shows the differences aggregated over all regions. The difference in the number of predicted infectious individuals between Model C and Model D are accumulated over

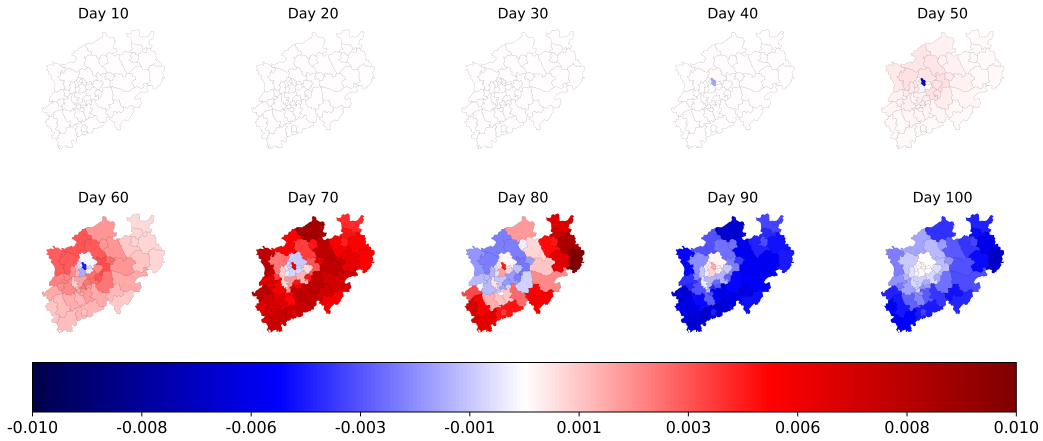


Figure 5.7: **Spatial visualization of the difference in the results of Model C and Model D.** Difference of infectious shares, i.e., the difference in numbers of infected individuals normalized by the population size of the respective county, between Model C and Model D over 100 days. For counties with a red color, Model C predicts a higher number of infectious individuals, while it is lower for counties with a blue color.

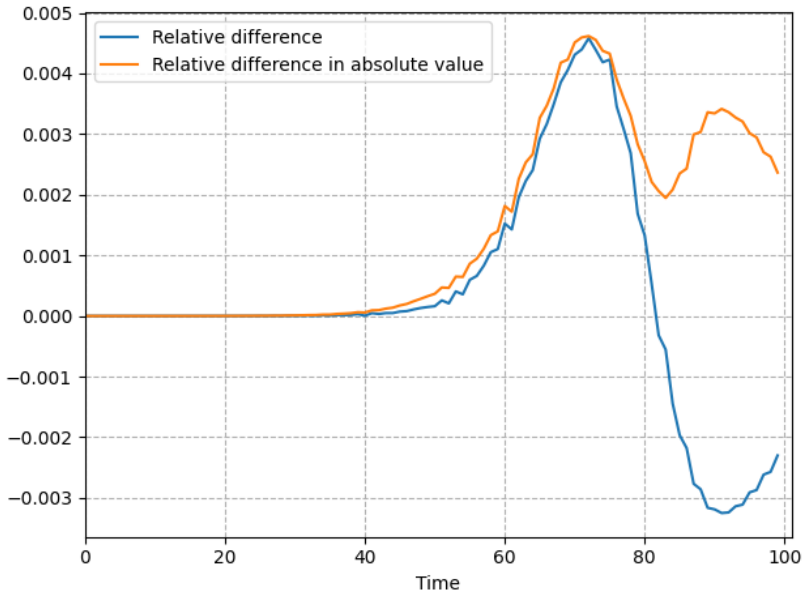


Figure 5.8: **Relative differences between Model C and Model D aggregated over all regions.** Visualization of the differences and the differences in absolute value of the predicted numbers of infectious individuals, aggregated over all regions and normalized by the total population size.

all regions and normalized by the total population size. In accordance with our previous observations, there is no visible difference until day 40. After that, we observe a fast increase until day 70, suggesting that the overall number of infectious individuals is predicted higher by the ODE model. This is followed by a rapid decrease into the negative range, indicating the lower number of predicted infectious individuals of Model C. It is evident that this approach results in the cancellation of differences with opposite signs, as we observe, for instance, in Fig. 5.7 around day 80. Although we see a lot of counties with either reddish or blueish shades, the difference aggregated over all counties shown in Fig. 5.8 is only 0.1%. To avoid this cancellation, we further plot the aggregated difference in absolute value, normalized by the total population (orange line) in Fig. 5.8. The resulting curve displays a similar trend, but also highlights the difference between the two models by eliminating the cancellation effect. For the maximum difference aggregated over all regions, we determine a value of 0.45%. Considering that these are two different models, a difference of 0.45% is remarkably small.

6 Conclusion

In this thesis, we have presented and evaluated different modeling strategies to represent mobility in epidemiological models for infectious disease dynamics. We introduced simple compartmental models based on ODEs and considered possible extensions concerning age and spatial resolution. This led us to *metapopulation models*, which divide a population into spatially separated subpopulations that interact with one another through connections between the respective patches in which they are located. Following [40, 23], we derived a metapopulation model based on ODEs which incorporates the spatial structure in its differential equations (Model B). Proceeding from there, we derived a similar model that overcomes certain shortcomings of the previous model (Model C). A different approach of introducing mobility to epidemiological models was given with a semi-discrete hybrid graph-based approach which is characterized by independent models for every patch and an exchange of information at discrete time points (Model D). Given the reduced necessity for synchronization between discrete time steps, the design of the graph-based model offers better parallelization properties. We have analyzed the implementations of both modeling approaches and performed a detailed complexity analysis based on the number of FLOP required, revealing the strong dependence on the number of age groups in the considered population. Based on numerical simulations, we have examined the performance of the models with regard to their required time to solution and the required steps performed by an adaptive solver. Using the basic reproduction number as an indicator, we have validated the equation-based models. A fictional simulation scenario using the example of the federal state of North Rhine-Westphalia has demonstrated the differences in the outcomes of the models.

While Model C and Model D presented a high degree of similarity in its predictions, the results of Model B deviate substantially. Compared to the other models, the outcome of Model B shows a faster spread of disease and more severe outcome. This observation was supported by the greater reproduction number, and can be attributed to the limitations we have overcome with Model C.

In terms of implementability, the equation-based models, Model B and Model C, have demonstrated a clear advantage over the graph-based model. In addition to the framework for the compartmental models, which is necessary for both approaches, the graph-based model also requires a compatible graph framework. Moreover, while the equation-based models can be solved with pre-implemented solvers, the solving of the graph-based

model is more complex and demands further considerations. Consequently, the graph-based model also shows a deficit in terms of comprehensibility. Nevertheless, it should be noted that the graph-based model offers a higher degree of flexibility, as it is not limited to equation-based models.

A comparison of the complexity of both models demonstrated that, depending on the fixed step size of the solver, the equation-based model requires a slightly lower number of FLOP than the graph-based model. The runtime analysis yielded results which were consistent with these observations, for both a population with a single age group, and for an age-resolved population. We could transfer these findings from a solver with a fixed step size to adaptive solvers by showing that both models, when run with an adaptive solver using an acceptable tolerance for the step size control, require a similar number of steps.

The potential for further research and enhancements is present in both models. The authors of [47] extended the graph-based Model D to account for transmissions during commuting. A similar objective has been followed by [23] for the equation-based metapopulation model, extending Model B. Following their approach, Model C can be extended to reflect the impact of mobility in greater detail. Moreover, we already mentioned possibilities for generalization when deriving Model C, such as the introduction of age- and health-status-dependent commuting strengths, which is already reflected in Model D.

Computational complexity of the graph-based Model D can be reduced by eliminating multi-edges with minimal weight. While this is not supported by the current implementation of Model C, exchanging the commuting strengths matrix of Model C by a sparse matrix may reduce complexity for this model as well. Additionally, Section 5.1 emphasizes that further investigation is necessary to assess the memory consumption and performance of the models.

A promising starting point for further research can be also found in the parallelization of both models. Given the structure of the graph-based model, which provides great parallelization properties, it is worth investigating to determine whether improvements can be made to the runtime and performance.

Bibliography

[1] K. Ahnert and M. Mulansky, *Class template runge_kutta_cash_karp54*, boost C++ libraries. [Online]. Available: https://www.boost.org/doc/libs/1_87_0/libs/numeric/odeint/doc/html/doxygen/odeint_reference/classboost_1_1numeric_1_1odeint_1_1runge__kutta__cash__karp54.html (visited on 12/27/2024).

[2] L. J. Allen, "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 128–142, May 2017, ISSN: 24680427. DOI: 10.1016/j.idm.2017.03.001.

[3] J. Arino, R. Jordan, and P. Van Den Driessche, "Quarantine in a multi-species epidemic model with spatial dynamics," *Mathematical Biosciences*, vol. 206, no. 1, pp. 46–60, Mar. 2007, ISSN: 00255564. DOI: 10.1016/j.mbs.2005.09.002.

[4] S. Arregui, M. J. Iglesias, S. Samper, D. Marinova, C. Martin, J. Sanz, and Y. Moreno, "Data-driven model for the assessment of *Mycobacterium tuberculosis* transmission in evolving demographic structures," *Proceedings of the National Academy of Sciences*, vol. 115, no. 14, E3238–E3245, Apr. 3, 2018. DOI: 10.1073/pnas.1720606115.

[5] S. Bansal, B. T. Grenfell, and L. A. Meyers, "When individual behaviour matters: Homogeneous and network models in epidemiology," *Journal of The Royal Society Interface*, vol. 4, no. 16, pp. 879–891, Oct. 22, 2007, ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2007.1100.

[6] J. Bicker, R. Schmieding, M. Meyer-Hermann, and M. J. Kühn, "Hybrid metapopulation agent-based epidemiological models for efficient insight on the individual scale: A contribution to green computing," *Infectious Disease Modelling*, vol. 10, no. 2, pp. 571–590, Jun. 2025, ISSN: 2468-0427. DOI: 10.1016/j.idm.2024.12.015.

[7] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical Models in Epidemiology* (Texts in Applied Mathematics). New York, NY: Springer New York, 2019, vol. 69, ISBN: 978-1-4939-9828-9. DOI: 10.1007/978-1-4939-9828-9.

[8] Bundesagentur für Arbeit, *Jahresdaten zu Ein- und Auspendlern für Kreise und Gemeinden in Deutschland*, Dec. 6, 2024. [Online]. Available: https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Pendler/Pendler-Nav.html?Thema%3DEinpendler%26DR_Land%3D08000000%26DR_Gebiete%3Dall%26toggleswitch%3D0.

[9] W. Duan, Z. Fan, P. Zhang, G. Guo, and X. Qiu, "Mathematical and computational approaches to epidemic modeling: A comprehensive review," *Frontiers of Computer Science*, vol. 9, no. 5, pp. 806–826, Oct. 2015, ISSN: 2095-2228, 2095-2236. DOI: 10.1007/s11704-014-3369-2.

[10] L. Fumanelli, M. Ajelli, P. Manfredi, A. Vespignani, and S. Merler, "Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread," *PLoS Computational Biology*, vol. 8, no. 9, M. Salathé, Ed., e1002673, Sep. 13, 2012, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002673.

[11] T. Gruber, M. Panzlaff, J. Eitzinger, G. Hager, and G. Wellein, *Likwid*, Dec. 2024. DOI: 10.5281/zenodo.4275676..

[12] G. Guennebaud, B. Jacob, *et al.*, *Eigen v3*, http://eigen.tuxfamily.org, 2010.

[13] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000. DOI: 10.1137/S0036144500371907.

[14] D. Kerkmann, S. Korf, K. Nguyen, D. Abele, A. Schengen, C. Gerstein, J. H. Göbbert, A. Basermann, M. J. Kühn, and M. Meyer-Hermann, *Agent-based modeling for realistic reproduction of human mobility and contact behavior to evaluate test and isolation strategies in epidemic infectious disease spread*, Version Number: 1, 2024. DOI: 10.48550/ARXIV.2410.08050.

[15] W. O. Kermack and A. G. McKendrick, "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, ISSN: 0950-1207, 2053-9150. DOI: 10.1098/rspa.1927.0118.

[16] S. Khailaie, T. Mitra, A. Bandyopadhyay, M. Schips, P. Mascheroni, P. Vanella, B. Lange, S. C. Binder, and M. Meyer-Hermann, "Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures," *BMC Medicine*, vol. 19, no. 1, p. 32, Dec. 2021, ISSN: 1741-7015. DOI: 10.1186/s12916-020-01884-4.

[17] W. Koslow, M. J. Kühn, S. Binder, M. Klitz, D. Abele, A. Basermann, and M. Meyer-Hermann, "Appropriate relaxation of non-pharmaceutical interventions minimizes the risk of a resurgence in SARS-CoV-2 infections in spite of the delta variant," *PLOS Computational Biology*, vol. 18, no. 5, C. J. Struchiner, Ed., e1010054, May 16, 2022, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010054.

[18] J. Kottarathil, S. Naduvath, and J. V. Kureethara, *Graph Theory and Decomposition*, 1st ed. Boca Raton: Chapman and Hall/CRC, Feb. 20, 2024, ISBN: 978-1-00-339167-8. DOI: 10.1201/9781003391678.

[19] M. J. Kühn, D. Abele, S. Binder, K. Rack, M. Klitz, J. Kleinert, J. Gilg, L. Spataro, W. Koslow, M. Siggel, M. Meyer-Hermann, and A. Basermann, "Regional opening strategies with commuter testing and containment of new SARS-CoV-2 variants in Germany," *BMC Infectious Diseases*, vol. 22, no. 1, p. 333, Dec. 2022, ISSN: 1471-2334. DOI: 10.1186/s12879-022-07302-9.

[20] M. J. Kühn *et al.*, "Assessment of effective mitigation and prediction of the spread of SARS-CoV-2 in germany using demographic information and spatial resolution," *Mathematical Biosciences*, vol. 339, p. 108 648, Sep. 2021, ISSN: 00255564. DOI: 10.1016/j.mbs.2021.108648.

[21] M. J. Kühn *et al.*, *MEmilio v1.3.0 - a high performance modular EpideMIcs simuLatIOn software*, Nov. 28, 2024. DOI: 10.5281/ZENODO.14237545.

[22] P. Leleux, C. Schwarz, M. J. Kühn, C. Kruse, and U. Rüde, "Complexity analysis and scalability of a matrix-free extrapolated geometric multigrid solver for curvilinear coordinates representations from fusion plasma applications," working paper or preprint, Dec. 2023. [Online]. Available: https://hal.science/hal-04356523.

[23] J. Liu, G. P. Ong, and V. J. Pang, "Modelling effectiveness of COVID-19 pandemic control policies using an area-based SEIR model with consideration of infection during interzonal travel," *Transportation Research Part A: Policy and Practice*, vol. 161, pp. 25–47, Jul. 2022, ISSN: 09658564. DOI: 10.1016/j.tra.2022.05.003.

[24] J. Liu and S. Xia, *Computational Epidemiology: From Disease Transmission Modeling to Vaccination Decision Making* (Health Information Science). Cham: Springer International Publishing, 2020, ISBN: 978-3-030-52109-7. DOI: 10.1007/978-3-030-52109-7.

[25] M. Martcheva, *An Introduction to Mathematical Epidemiology* (Texts in Applied Mathematics). Boston, MA: Springer US, 2015, vol. 61, ISBN: 978-1-4899-7612-3. DOI: 10.1007/978-1-4899-7612-3.

[26] S. A. McDonald, A. C. Teirlinck, M. Hooiveld, L. Van Asten, A. Meijer, M. De Lange, A. B. Van Gageldonk-Lafeber, and J. Wallinga, "Inference of age-dependent case-fatality ratios for seasonal influenza virus subtypes a(h3n2) and a(h1n1)pdm09 and b lineages using data from the netherlands," *Influenza and Other Respiratory Viruses*, vol. 17, no. 6, e13146, Jun. 2023. DOI: 10.1111/irv.13146.

[27] R. Milwid, A. Steriu, J. Arino, J. Heffernan, A. Hyder, D. Schanzer, E. Gardner, M. Haworth-Brockman, H. Isfeld-Kiely, J. M. Langley, and S. M. Moghadas, "Toward standardizing a lexicon of infectious disease modeling terms," *Frontiers in Public Health*, vol. 4, Sep. 28, 2016, ISSN: 2296-2565. DOI: 10.3389/fpubh.2016.00213.

[28] J. Mossong *et al.*, "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS Medicine*, vol. 5, no. 3, S. Riley, Ed., e74, Mar. 25, 2008, ISSN: 1549-1676. DOI: 10.1371/journal.pmed.0050074.

[29] D. V. Ouellette, "Schur complements and statistics," *Linear Algebra and its Applications*, vol. 36, pp. 187–295, Mar. 1981, ISSN: 00243795. DOI: 10.1016/0024-3795(81)90232-9.

[30] L. Plötzke, A. Wendler, R. Schmieding, and M. J. Kühn, *Revisiting the linear chain trick in epidemiological models: Implications of underlying assumptions for numerical solutions*, Version Number: 1, 2024. DOI: 10.48550/ARXIV.2412.09140.

[31] K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLOS Computational Biology*, vol. 13, no. 9, B. Halloran, Ed., e1005697, Sep. 12, 2017, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005697.

[32] Robert Koch-Institut, *SARS-CoV-2 Infektionen in Deutschland*, in collab. with M. Diercke, H. Claus, L. Grabenhenrich, and H. Wuensche, version 2025-01-21, Jan. 21, 2025. DOI: 10.5281/ZENODO.14709565.

[33] K. H. Rosen, *Discrete mathematics and its applications*, Eighth edition. New York, NY: McGraw-Hill, 2019, 942 pp., ISBN: 978-1-259-67651-2.

[34] L. Saker, K. Lee, B. Cannito, A. Gilmore, and D. H. Campbell-Lendrum, *Globalization and infectious diseases : A review of the linkages*, 2004.

[35] Statistisches Bundesamt (DESTATIS), *Population: Germany, reference date, age.* [Online]. Available: https://www-genesis.destatis.de/datenbank/online/url/d7546ce2 (visited on 11/26/2024).

[36] S. Straif-Bourgeois, J. L. Tonzel, M. Kretzschmar, and R. Ratard, "Infectious disease epidemiology," in *Handbook of Epidemiology*, W. Ahrens and I. Pigeot, Eds. New York, NY: Springer New York, 2019, pp. 1–79, ISBN: 978-1-4614-6625-3. DOI: 10.1007/978-1-4614-6625-3_34-1.

[37] The World Bank, *World bank country and lending groups.* [Online]. Available: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (visited on 01/15/2025).

[38] P. Van Den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, no. 1, pp. 29–48, Nov. 2002, ISSN: 00255564. DOI: 10.1016/S0025-5564(02)00108-6.

[39] R. Verity *et al.*, "Estimates of the severity of coronavirus disease 2019: A model-based analysis," *The Lancet Infectious Diseases*, vol. 20, no. 6, pp. 669–677, Jun. 2020, ISSN: 14733099. DOI: 10.1016/S1473-3099(20)30243-7.

[40] J. Wang, X. Wang, and J. Wu, "Inferring metapopulation propagation network for intra-city epidemic control and prevention," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London United Kingdom: ACM, Jul. 19, 2018, pp. 830–838, ISBN: 978-1-4503-5552-0. DOI: 10.1145/3219819.3219865.

[41] L. Wei, *C++ Template Metaprogramming in Practice*, 1st ed. First edition. | Boca Raton, FL : CRC Press, 2021.: Auerbach Publications, Dec. 1, 2020, ISBN: 978-1-00-310231-1. DOI: 10.1201/9781003102311.

[42] A. Wendler, L. Plötzke, H. Tritzschak, and M. J. Kühn, *A nonstandard numerical scheme for a novel SECIR integro-differential equation-based model allowing nonexponentially distributed stay times*, Version Number: 1, 2024. DOI: 10.48550/ARXIV.2408.12228.

[43] World Health Organization (WHO), *World Health Statistics 2023: Monitoring Health for the SDGs, Sustainable Development Goals*, 1st ed. Geneva: World Health Organization, 2023, 1 p., ISBN: 978-92-4-007432-3.

[44] World Health Organization (WHO), "COVID-19 epidemiological update," 170, Aug. 13, 2024. [Online]. Available: https://www.who.int/publications/m/item/covid-19-epidemiological-update-edition-170 (visited on 08/22/2024).

[45] World Health Organization (WHO), *Fact sheet: The top 10 causes of death*, en. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (visited on 01/15/2025).

[46] World Health Organization (WHO), *Statement on the fifteenth meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic*, en. [Online]. Available: https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic (visited on 01/15/2025).

[47] H. Zunker, R. Schmieding, D. Kerkmann, A. Schengen, S. Diexer, R. Mikolajczyk, M. Meyer-Hermann, and M. J. Kühn, "Novel travel time aware metapopulation models and multi-layer waning immunity for late-phase epidemic and endemic scenarios," *PLOS Computational Biology*, vol. 20, no. 12, T. Britton, Ed., e1012630, Dec. 16, 2024, ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012630.

Versicherung an Eides statt

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit “Evaluation of different mobility models in equation-based infectious disease dynamics” selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten fremden Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte Druckfassung der eingereichten elektronischen Fassung vollständig entspricht.

Ort, Datum Unterschrift