Investigating the Evolution of the COVID-19 Pandemic in Germany Using Physics-Informed Neural Networks

Bachelor Thesis in Computer Science

submitted by
Phillip Rothenbeck
born February 22, 2002 in Eckernförde

written at
Computer Vision Group
Department of Mathematics and Computer Science
Friedrich-Schiller-Universität Jena

Supervisor:	Prof. Dr.-Ing. Joachim Denzler
Advisor:	Niklas Penzel, Sai Karthikeya Vemuri
Started:	May 1, 2024
Finished:	September 14, 2024

Eigenständigkeitserklärung

1. Hiermit versichere ich, dass ich die vorliegende Arbeit - bei einer Gruppenarbeit die von mir zu verantwortenden und entsprechend gekennzeichneten Teile - selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Die aus fremden oder auch eigenen, älteren Quellen wörtlich oder sinngemäß übernommenen Textstellen, Gedankengänge, Konzepte, Grafiken etc. in meinen Ausführungen habe ich als solche eindeutig gekennzeichnet und mit vollständigen Verweisen auf die jeweilige Quelle versehen. Alle weiteren Inhalte dieser Arbeit ohne entsprechende Verweise stammen im urheberrechtlichen Sinn von mir.

2. Ich weiß, dass meine Eigenständigkeitserklärung sich auch auf nicht zitierfähige, generierende KI-Anwendungen (nachfolgend „generierende KI“) bezieht. Mir ist bewusst, dass die Verwendung von generierender KI unzulässig ist, sofern nicht deren Nutzung von der prüfenden Person ausdrücklich freigegeben wurde (Freigabeerklärung). Sofern eine Zulassung als Hilfsmittel erfolgt ist, versichere ich, dass ich mich generierender KI lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss deutlich überwiegt. Ich verantworte die Übernahme der von mir verwendeten maschinell generierten Passagen in meiner Arbeit vollumfänglich selbst. Für den Fall der Freigabe der Verwendung von generierender KI für die Erstellung der vorliegenden Arbeit wird eine Verwendung in einem gesonderten Anhang meiner Arbeit kenntlich gemacht. Dieser Anhang enthält eine Angabe oder eine detaillierte Dokumentation über die Verwendung generierender KI gemäß den Vorgaben in der Freigabeerklärung der prüfenden Person. Die Details zum Gebrauch generierender KI bei der Erstellung der vorliegenden Arbeit inklusive Art, Ziel und Umfang der Verwendung sowie die Art der Nachweispflicht habe ich der Freigabeerklärung der prüfenden Person entnommen.

3. Ich versichere des Weiteren, dass die vorliegende Arbeit bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt wurde oder in deutscher oder einer anderen Sprache als Veröffentlichung erschienen ist.

4. Mir ist bekannt, dass ein Verstoß gegen die vorbenannten Punkte prüfungsrechtliche Konsequenzen haben und insbesondere dazu führen kann, dass meine Prüfungsleistung als Täuschung und damit als mit „nicht bestanden“ bewertet werden kann. Bei mehrfachem oder schwerwiegendem Täuschungsversuch kann ich befristet oder sogar dauerhaft von der Erbringung weiterer Prüfungsleistungen in meinem Studiengang ausgeschlossen werden.

5. Die Richtlinien des Lehrstuhls für Examensarbeiten habe ich gelesen und anerkannt. Seitens des Verfassers/der Verfasserin bestehen keine Einwände, die vorliegende Examensarbeit für die öffentliche Benutzung zur Verfügung zu stellen.

Jena, den 14. September 2024 Phillip Rothenbeck

Überblick

German version of the abstract.
Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Abstract

English version of the abstract.
Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

Chapter 1

Introduction

In the early months of 2020, Germany, like many other countries, was struck by the novel *Coronavirus Disease* (COVID-19) [36]. The pandemic, which originates in Wuhan, China, had a profound impact on the global community, paralyzing it for over two years. In response to the pandemic, the German government employed a multifaceted approach [24], encompassing the introduction of vaccines and non-pharmaceutical mitigation policies such as lockdowns. Between mitigation policies and varying strains of COVID-19, which have exhibited varying degrees of infectiousness and lethality [27], Germany had recorded over 38,400,000 infection cases and 174,000 deaths, as of the end of June in 2023 [34]. In light of these figures the need for an analysis arises.

The dynamics of the spread of disease transmission in the real-world are complex. A multitude of factors influence the course of a disease, and it is challenging to gain a comprehensive understanding of these factors and develop tools that allows for the comparison of disease courses across different diseases and time points. The common approach in epidemiology to address this is the utilization of epidemiological models that approximate the dynamics by focusing on specific factors and modeling these using mathematical tools. These models provide transition rates and parameters that determine the behavior of a disease within the boundaries of the model. A fundamental epidemiological model, is the *SIR model*, which was first proposed by Kermack and McKendrick [12] in 1927. The SIR model is a compartmentalized model that divides the entire population into three distinct groups: the *susceptible* compartment, S; the *infectious* compartment, I; and the *removed* compartment, R. In the context of the SIR model, the constant parameters of the transmission rate β and the recovery rate α serve to quantify and determine the course of a pandemic. However, pandemic is not a static entity, therefor, Liu and Stechlinski [16], and Setianto and Hidayat [31], propose an SIR model with time-dependent transition rates

and reproduction number $\mathcal{R}_t$. The SIR model is defined by a system of differential equations, that incorporate the transition rates, thereby depicting the fluctuation between the three compartments. For a given set of data, the transition rate can be identified by solving the set of differential systems. Recently, the data-driven approach of *physics-informed neural networks* (PINN) has gained attention due to its capability of finding solutions to differential equations by fitting its predictions to both given data and the governing system of differential equations. By employing this methodology, Shaier *et al.* [32] were able to find the transition rate on data for different diseases. Additionally, Millevoi *et al.* [18] were able to identify the reproduction number $\mathcal{R}_t$ for both synthetic and Italian COVID-19 data using an approach based on a reduced version of the SIR model.

The objective of this thesis is to identify the transition rates β and $alpha$, as well as the reproduction number $\mathcal{R}_t$ of COVID-19 over the first 1200 days of recorded data in Germany and its federal states. The Robert Koch Institute (RKI) has compiled data on both reported cases and associated moralities from the beginning of the outbreak in Germany to the present. We utilize and preprocess this data according to the required format of our approaches. As the raw data lacks information on recovery incidence, we introduce the recovery queue that simulates a recovery period. To estimate the transition rates we adopt the approach of Shaier *et al.* [32], which utilizes a physics-informed neural network learning the data, which consists of time point with their respective sizes of the S, I and R compartments, to predict the transition rates based on the data and the governing system of differential equations. Moreover, we utilize the methodology proposed by Millevoi *et al.* [18] that estimates the reproduction number for each day across the 1200-day span for each German state and Germany as a whole, in reduced SIR model. Thus needing only the size of the I group for each time step. To validate the effectiveness of these methods, we first conduct experiments on a small synthetic dataset before applying the techniques to real-world data. We then analyze the plausibility of our results by comparing them to real-world events and data such as vaccination ratios of each region or the peaks of impactful variants to demonstrate the relevance of these numbers. This analysis demonstrates the relevance of our findings and reveals a correlation between our results and real-world developments, thus supporting the effectiveness of our approach.

1.1 Related work

In this section, we categorize our work into the context of existing literature on the topic of solving the epidemiological models for real-world data. The first work, by Smirnova *et al.* [33], endeavors to identify a stochastic methodology for estimating the time-dependent transmission rate $\beta(t)$. They achieve this by projecting the time-dependent transmission rate onto a finite subspace, that is defined by Legendre polynomials. Subsequently, they compare the three regularization techniques of variational (Tikhonov's) regularization, truncated singular value decomposition (TSVD), and modified TSVD to ascertain the most reliable method for forecasting with limited data. Their findings indicate that modified TSVD provides the most stable forecasts on limited data, as demonstrated on both simulated data and real-world data from the 1918 influenza pandemic and the Ebola epidemic. In contrast, we utilize physics-informed neural networks (PINN) to find the constant transition rates and the reproduction number for Germany and its states

Some related works similarly to us apply PINN approaches to COVID-19 and other real-world disease data such as [32, 2, 21, 18]. Specifically in [32], Shaier *et al.* put forth a data-driven approach which they refer to as disease informed neural networks (DINN). In their work, they demonstrate the capacity of DINNs to forecast the trajectory of epidemics and pandemics. They underpin the efficacy of their approach by applying it to 11 diseases, that have previously been modeled. In their experiments they employ the SIDR (susceptible, infectious, dead, recovered) model. Finally, they present that this method is a robust and effective means of identifying the parameters of a SIR model.

Similarly in [2], Berkhahn and Ehrhard employ the susceptible, vaccinated, infectious, hospitalized and removed (SVIHR) model. The proposed PINN methodology initially estimates the SVIHR model parameters for German COVID-19 data, covering the time span from the inceptions of the outbreak to the end of 2021. For comparative purposes, Berkhahn and Ehrhard employ the method of non-standard finite differences (NSFD) as well. The authors employ both methods the two forecasting methods project the trajectory of COVID-19 from mid-April 2023 onwards. Berkhahn and Ehrhard find that the PINN is able to adapt to varying vaccination rates and emerging variants.

Furthermore, Olumoyin *et al.* [21] employ an alternative methodology for identifying the time-dependent transmission rate of an asymptomatic-SIR model accounting for unreported infectious cases. The PINN approach they introduce, utilizes the cumulative and daily reported infection cases and symptomatic recovered cases, to demonstrate the effect of different mitigation measures and to ascertain the proportion of non-symptomatic individuals and asymptomatic recovered individuals. With this they can illustrate the influence of vaccination and a set non-pharmaceutical mitigation methods on the transmission of COVID-19 on data from Italy, South Korea, the United Kingdom, and the United States.

Finally, Millevoi *et al.* [18] address the issue of the changes in the transmission rate due to the dynamics of a pandemic. The authors employ the reproduction number to reduce the system of differential equations to a single equation and introduce a reduced-split version of the PINN, which initially trains on the data and then trains to minimize the residual of the ODE. They test their approach on five synthetic and two real-world scenarios from the early stages of the COVID-19 pandemic in Italy. This method yields an increase in both accuracy and training speed. In contrast, to these works, we estimate the rates and the reproduction number for Germany for the entirety of the span from early March in 2020 to late June in 2023.

1.2 Overview

This thesis is comprised of four chapters. Chapter 2 presents with the theoretical overview of mathematical modeling in epidemiology, with a particular focus on the SIR model. Subsequently, it shifts its focus to neural networks, specifically on the background of physics-informed neural networks (PINN) and their use in solving ordinary differential equations. Chapter 3 outlines the methodology employed in this thesis. First we present the data, that was collected by the Robert Koch Institute (RKI). Then we present the PINN approaches, which are inspired by the work of Shaier *et al.* [32] and Millevoi *et al.* [18]. Chapter 4 presents the setups and results of the experiments that we conduct. This chapter is divided into two sections. The first section presents and discusses the results concerning the transition rates of β and α. The subsequent section presents the results concerning the reproduction value $\mathcal{R}_t$. Finally, in Chapter 5, we connect our results with the events of the real-world and give an overview of potential further work.

Chapter 2

Theoretical Background

This chapter introduces the theoretical foundations for the work presented in this thesis. In Section 2.1 and Section 2.2, we describe differential equations and the underlying theory. In these Sections both the explanations and the approach are based on a book on analysis by Rudin [29] and a book about ordinary differential equations by Tenenbaum and Pollard [35]. Subsequently, we employ this knowledge to examine various pandemic models in Section 2.3. Finally, we address the topic of neural networks with a focus on the multilayer perceptron in Section 2.4 and physics informed neural networks in Section 2.5.

2.1 Mathematical Modelling using Functions

To model a physical problem mathematically, it is necessary to define a fundamental set of numbers or quantities upon which the subsequent calculations will be based. These sets may represent, for instance, a specific time interval or a distance. The term *domain* describes these fundamental sets of numbers or quantities [29]. A *variable* is a changing entity living in a certain domain. In this thesis, we will focus on domains of real numbers in $\mathbb{R}$.

The mapping between variables enables the modeling of a physical process and may depict semantics. We use functions in order to facilitate this mapping. Let $A, B \subset \mathbb{R}$ be to subsets of the real numbers, then we define a function as the mapping

$$f : A \to B. \tag{2.1}$$

In other words, the function f maps elements $x \in A$ to values $f(x) \in B$. A is the *domain* of f, while B is the *codomain* of f. Functions are capable of representing the state of a system as a value based on an input value from their domain. One

illustrative example is a function that maps a time step to the distance covered since a starting point. In this case, time serves as the domain, while the distance is the codomain.

2.2 Mathematical Modelling using Differential Equations

Often, the behavior of a variable or a quantity across a domain is more interesting than its current state. Functions are able to give us the latter, but do not contain information about the change of a system. The objective is to determine an effective method for calculating the change of a function across its domain. Let f be a function and $[a, b] \subset \mathbb{R}$ an interval of real numbers. The expression

$$m = \frac{f(b) - f(a)}{a - b} \tag{2.2}$$

gives the average rate of change. While the average rate of change is useful in many cases, the momentary rate of change is more accurate. To calculate this, we need to narrow down, the interval to an infinitesimal. For each $x \in [a, b]$ we calculate

look up in Rudin - cite (wordly)

$$\frac{df}{dx} = \lim_{t \to x} \frac{f(t) - f(x)}{t - x}, \tag{2.3}$$

if it exists. As the Tenenbaum and Pollard [35] define, df/dx is the *derivative*, which is "the rate of change of a variable with respect to another". The relation between a variable and its derivative is modeled in a *differential equation*. The derivative of df/dx yields d^2f/dx^2, which is the function that calculates the rate of change of the rate of change and is called the *second order derivative*. Iterating this n times results in d^nf/dx^n, the derivative of the n'th order. A method for obtaining a differential equation is to derive it from the semantics of a problem. For example, in physics a differential equation can be derived from the law of the conservation of energy [4]. Differential equations find application in several areas such as engineering *e.g.*, the Kirchhoff's circuit laws [14] to describe the relation between the voltage and current in systems with resistors, inductors, and capacitors; physics with, *e.g.*, the Schrödinger equation, which predicts the probability of finding particles like electrons in specific places or states in a quantum system; economics, *e.g.*, Black-Scholes equation [20] predicting the price of financial derivatives, such as options, over time; epidemiology with the SIR Model [12]; and beyond.

In the context of functions, it is possible to have multiple domains, meaning that function has more than one parameter. To illustrate, consider a function operating in two-dimensional space, wherein each parameter represents one axis. Another example would be a function, that maps its inputs of a location variable and a time variable on a height. The term *partial differential equations* (*PDE*'s) describes differential equations of such functions, which contain partial derivatives with respect to each individual domain. In contrast, *ordinary differential equations* (*ODE*'s) are the single derivatives for a function having only one domain [35]. In this thesis, we restrict ourselves to ODE's.

A *system of differential equations* is the name for a set of differential equations. The derivatives in a system of differential equations each have their own codomain, which is part of the problem, while they all share the same domain.

Tenenbaum and Pollard [35] provide many examples for ODE's, including the *Motion of a Particle Along a Straight Line*. Further, Newton's second law states that "the rate of change of the momentum of a body ($momentum = mass \cdot velocity$) is proportional to the resultant external force F acted upon it" [35]. Let m be the mass of the body in kilograms, v its velocity in meters per second and t the time in seconds. Then, Newton's second law translates mathematically to

$$F = m\frac{dv}{dt}. \tag{2.4}$$

It is evident that the acceleration, $a = \frac{dv}{dt}$, as the rate of change of the velocity is part of the equation. Additionally, the velocity of a body is the derivative of the distance traveled by that body. Based on these findings, we can rewrite the Equation (2.4) to

$$F = ma = m\frac{d^2s}{dt^2}. \tag{2.5}$$

To conclude, note that this explanation of differential equations focuses on the aspects deemed crucial for this thesis and is not intended to be a complete explanation of the subject. To gain a better understanding of it, we recommend the books mentioned above [29, 35]. In the following section we describe the application of these principles in epidemiological models.

2.3 Epidemiological Models

Pandemics, like *COVID-19*, which have resulted in a significant number of fatalities. Hence, the question arises: How should we analyze a pandemic effectively? It is essential to study whether the employed countermeasures are efficacious in combating the pandemic. Given the unfavorable public response to measures such as lockdowns, it is imperative to investigate that their efficacy remains commensurate with the costs incurred to those affected. In the event that alternative and novel technologies were in use, such as the mRNA vaccines in the context of COVID-19, it is needful to test the effect and find the optimal variant. In order to shed light on the aforementioned events, we need a method to quantify the pandemic along with its course of progression.

The real world is a highly complex system, which presents a significant challenge attempting to describe it fully in a mathematical model. Therefore, the model must reduce the complexity while retaining the essential information. Furthermore, it must address the issue of limited data availability. For instance, during COVID-19 institutions such as the Robert Koch Institute (RKI)[1] were only able to collect data on infections and mortality cases. Consequently, we require a model that employs an abstraction of the real world to illustrate the events and relations that are pivotal to understanding the problem.

2.3.1 SIR Model

In 1927, Kermack and McKendrick [12] introduced the *SIR Model*, which subsequently became one of the most influential epidemiological models. This model enables the modeling of infections during epidemiological events such as pandemics. The book *Mathematical Models in Biology* [6] reiterates the model and serves as the foundation for the following explanation of SIR models.

The SIR model is capable of illustrating diseases, which are transferred through contact or proximity of an individual carrying the illness and a healthy individual. This is possible due to the distinction between infected individuals who are carriers of the disease and the part of the population, which is susceptible to infection. In the model, the mentioned groups are capable to change, *e.g.*, healthy individuals becoming infected. The model assumes the size N of the population remains constant

[1] `https://www.rki.de/EN/Home/homepage_node.html`

throughout the duration of the pandemic. The population N comprises three distinct compartments: the *susceptible* group S, the *infectious* group I and the *removed* group R (hence SIR model). Let $\mathcal{T} = [t_0, t_f] \subseteq \mathbb{R}_{\geq 0}$ be the time span of the pandemic, then,

$$S : \mathcal{T} \to \mathbb{N}, \quad I : \mathcal{T} \to \mathbb{N}, \quad R : \mathcal{T} \to \mathbb{N}, \tag{2.6}$$

give the values of S, I and R at a certain point of time $t \in \mathcal{T}$. For S, I, R and N applies:

$$N = S + I + R. \tag{2.7}$$

The model makes another assumption by stating that recovered people are immune to the illness and infectious individuals can not infect them. The individuals in the R group are either recovered or deceased, and thus unable to transmit or carry the disease. As visualized in the Figure 2.1 the individuals may transition between groups

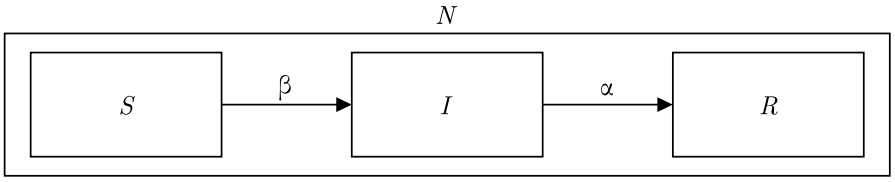


Figure 2.1: A visualization of the SIR model, illustrating N being split in the three groups S, I and R.

based on transition rates. The transmission rate β is responsible for individuals becoming infected, while the rate of removal or recovery rate α (also referred to as δ or ν, *e.g.*, [6, 18]) moves individuals from I to R.

We can describe this problem mathematically using a system of differential equations (see Section 2.2). Thus, Kermack and McKendrick [12] propose the following set of differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI, \\ \frac{dI}{dt} &= \beta SI - \alpha I, \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \tag{2.8}$$

This set of differential equations, is based on the following assumption: "The rate of transmission of a microparasitic disease is proportional to the rate of encounter of

susceptible and infective individuals modelled by the product (βSI)", according to Edelstein-Keshet [6]. The system shows the change in size of the groups per time unit due to infections, recoveries, and deaths.

The term βSI describes the rate of encounters of susceptible and infected individuals. This term is dependent on the size of S and I, thus Anderson and May [1] propose a modified model:

$$\begin{aligned} \frac{dS}{dt} &= -\beta\frac{SI}{N}, \\ \frac{dI}{dt} &= \beta\frac{SI}{N} - \alpha I, \\ \frac{dR}{dt} &= \alpha I. \end{aligned} \tag{2.9}$$

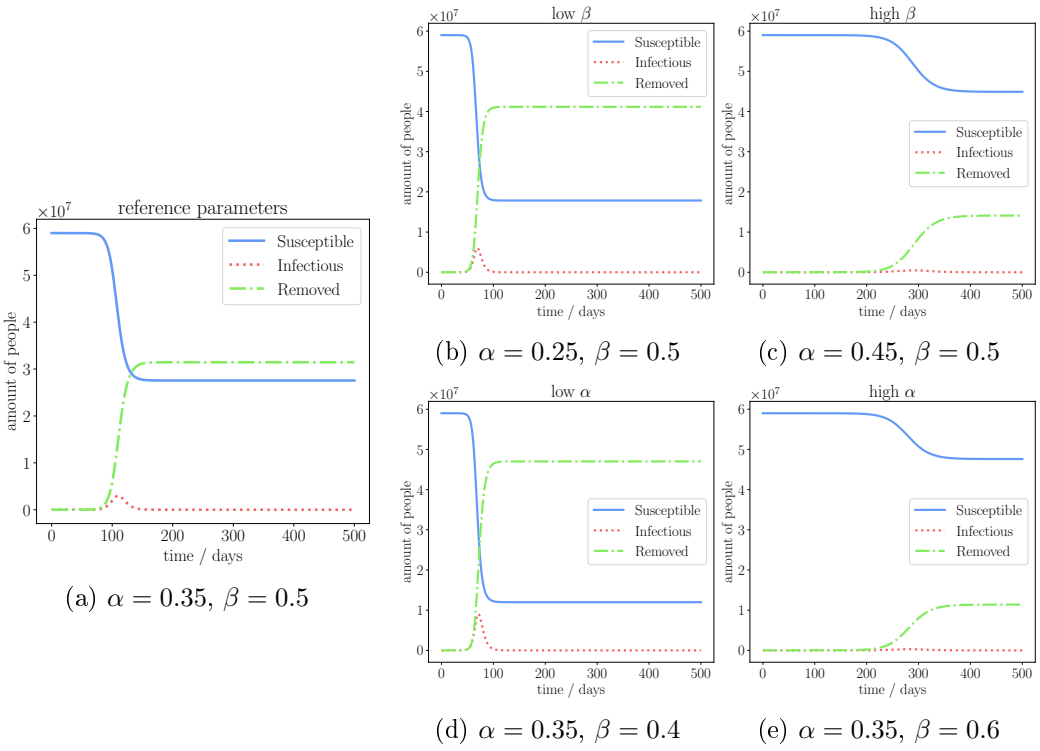
In Equation (2.9) βSI gets normalized by N, which is more correct in a real world aspect [1].

The initial phase of a pandemic is characterized by the infection of a small number of individuals, while the majority of the population remains susceptible. The infectious group has not yet infected any individuals thus neither recovery nor mortality is possible. Let $I_0 \in \mathbb{N}$ be the number of infected individuals at the beginning of the disease. Then,

$$\begin{aligned} S(0) &= N - I_0, \\ I(0) &= I_0, \\ R(0) &= 0, \end{aligned} \tag{2.10}$$

describes the initial configuration of a system in which a disease has just emerged.

In the SIR model the temporal occurrence and the height of the peak (or peaks) of the infectious group are of paramount importance for understanding the dynamics of a pandemic. A low peak occurring at a late point in time indicates that the disease is unable to keep pace with the rate of recovery, resulting in its demise before it can exert a significant influence on the population. In contrast, an early and high peak means that the disease is rapidly transmitted through the population, with a significant proportion of individuals having been infected. Figure 2.1 illustrates this effect by varying the values of β or α while simulating a pandemic using a model such



(a) $\alpha = 0.35$, $\beta = 0.5$

(b) $\alpha = 0.25$, $\beta = 0.5$

(c) $\alpha = 0.45$, $\beta = 0.5$

(d) $\alpha = 0.35$, $\beta = 0.4$

(e) $\alpha = 0.35$, $\beta = 0.6$

Figure 2.2: Synthetic data, using Equation (2.9) and $N = 7.9 \cdot 10^6$, $I_0 = 10$ with different sets of parameters. We visualize the case with the reference parameters in (a). In (b) and (c) we keep α constant, while varying the value of β. In contrast, (d) and (e) have varying values of α.

as Equation (2.9). It is evident that both the transmission rate β and the recovery rate α influence the height and time of the peak of I. When the number of infections exceeds the number of recoveries, the peak of I will occur early and will be high. On the other hand, if recoveries occur at a faster rate than new infections the peak will occur later and will be low. Thus, it is crucial to know both β and α, as these parameters characterize how the pandemic evolves.

The SIR model makes a number of assumptions that are intended to reduce the model's overall complexity while simultaneously increasing its divergence from actual reality. One such assumption is that the size of the population, N, remains constant, as the daily change is negligible to the total population. This depiction is not an accurate representation of the actual relations observed in the real world, as the size of a population is subject to a number of factors that can contribute to change. The population is increased by the occurrence of births and decreased by the occurrence

other assumptions in a bad light?

of deaths. Other examples are the impossibility for individuals to be susceptible again, after having recovered, or the possibility for the transition rates to change due to new variants or the implementation of new countermeasures. We address this latter option in the next Section 2.3.2.

2.3.2 Reduced SIR Model and the Reproduction Number

The Section 2.3.1 presents the classical SIR model. This model contains two scalar parameters β and α, which describe the course of a pandemic over its duration. This is beneficial when examining the overall pandemic; however, in the real world, disease behavior is dynamic, and the values of the parameters β and α change throughout the course of the disease. The reason for this is due to events such as the implementation of countermeasures that reduce the contact between the infectious and susceptible individuals, the emergence of a new variant of the disease that increases its infectivity or deadliness, or the administration of a vaccination that provides previously susceptible individuals with immunity without ever being infected. To address this, based on the time-dependent transition rates introduced by Liu and Stechlinski [16], and Setianto and Hidayat [31], Millevoi *et al.* [18] present a model that simultaneously reduces the size of the system of differential equations and solves the problem of time scaling at hand.

First, they alter the definition of β and α to be dependent on the time interval $\mathcal{T} = [t_0, t_f] \subseteq \mathbb{R}_{\geq 0}$,

$$\beta : \mathcal{T} \to \mathbb{R}_{\geq 0}, \quad \alpha : \mathcal{T} \to \mathbb{R}_{\geq 0}. \tag{2.11}$$

Another crucial element is $D(t) = \frac{1}{\alpha(t)}$, which represents the initial time span an infected individual requires to recuperate. Subsequently, at the initial time point t_0, the *reproduction number*,

$$\mathcal{R}_0 = \beta(t_0)D(t_0) = \frac{\beta(t_0)}{\alpha(t_0)}, \tag{2.12}$$

represents the number of susceptible individuals, that one infectious individual infects at the onset of the pandemic. In light of the effects of β and α (see Section 2.3.1), $\mathcal{R}_0 < 1$ indicates that the pandemic is emerging. In this scenario α is relatively low due to the limited number of infections resulting from $I(t_0) << S(t_0)$.
Further, $\mathcal{R}_0 > 1$ leads to the disease spreading rapidly across the population, with an increase in I occurring at a high rate. Nevertheless, $\mathcal{R}_0$ does not cover the entire

time span. For this reason, Millevoi *et al.* [18] introduce $\mathcal{R}_t$ which has the same interpretation as $\mathcal{R}_0$, with the exception that $\mathcal{R}_t$ is dependent on time. The time-dependent reproduction number is defined as,

$$\mathcal{R}_t = \frac{\beta(t)}{\alpha(t)} \cdot \frac{S(t)}{N}, \tag{2.13}$$

on the time interval $\mathcal{T}$. This definition includes the transition rates for information about the spread of the disease and information of the decrease of the ratio of susceptible individuals in the population. In contrast to β and α, $\mathcal{R}_t$ is not a parameter but a state variable in the model and enabling the following reduction of the SIR model.

Sai comment - earlier?

Equation (2.7) allows for the calculation of the value of the group R using S and I, with the term $R(t) = N - S(t) - I(t)$. Thus,

$$\begin{aligned} \frac{dS}{dt} &= \alpha(\mathcal{R}_t - 1)I(t), \\ \frac{dI}{dt} &= -\alpha\mathcal{R}_t I(t), \end{aligned} \tag{2.14}$$

is the reduction of Equation (2.8) on the time interval $\mathcal{T}$ using this characteristic and the reproduction number $\mathcal{R}_t$ (see Equation (2.13)). Another issue that Millevoi *et al.* [18] seek to address is the extensive range of values that the SIR groups can assume. Accordingly, they initially scale the time interval $\mathcal{T}$ using its borders to calculate the scaled time $t_s = \frac{t-t_0}{t_f-t_0} \in [0, 1]$. Subsequently, they calculate the scaled groups,

$$S_s(t_s) = \frac{S(t)}{C}, \quad I_s(t_s) = \frac{I(t)}{C}, \quad R_s(t_s) = \frac{R(t)}{C}, \tag{2.15}$$

using a large constant scaling factor $C \in \mathbb{N}$. Applying this to the variable I, results in,

$$\frac{dI_s}{dt_s} = \alpha(t_f - t_0)(\mathcal{R}_t - 1)I_s(t_s), \tag{2.16}$$

which is a further reduced version of Equation (2.8). This less complex differential equation results in a less complex solution, as it entails the elimination of a parameter (β) and the two state variables (S and R). The reduced SIR model, is more precise in applications with a worse data situation, due to its fewer input variables.

2.4 Multilayer Perceptron

In Section 2.2, we demonstrate the significance of differential equations in systems, illustrating how they can be utilized to elucidate the impact of a specific parameter on the system's behavior. In Section 2.3, we show specific applications of differential equations in an epidemiological context. The final objective is to solve these equations by finding a function that fits. Fitting measured data points to approximate such a function, is one of the multiple methods to achieve this goal. The *Multilayer Perceptron* (MLP) [30] is a data-driven function approximator. In the following section, we provide a brief overview of the structure and training of these *neural networks*. For reference, we use the book *Deep Learning* by Goodfellow *et al.* [10] as a foundation for our explanations.

The objective is to develop an approximation method for any function f^*, which could be a mathematical function or a mapping of an input vector to the desired output. Let $\boldsymbol{x}$ be the input vector and $\boldsymbol{y}$ the label, class, or result. Then, $\boldsymbol{y} = f^*(\boldsymbol{x})$, is the function to approximate. In the year 1958, Rosenblatt [28] proposed the perceptron modeling the concept of a neuron in a neuroscientific sense. The perceptron takes in the input vector $\boldsymbol{x}$ performs an operation and produces a scalar result. This model optimizes its parameters θ to be able to calculate $\boldsymbol{y} = f(\boldsymbol{x}; \theta)$ as accurately as possible. As Minsky and Papert [19] demonstrate, the perceptron is only capable of approximating a specific class of functions. Consequently, there is a necessity for an expansion of the perceptron.

As Goodfellow *et al.* [10] proceed, the solution to this issue is to decompose f into a chain structure of the form,

$$f(\boldsymbol{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\boldsymbol{x}))). \tag{2.17}$$

This nested version of a perceptron is a multilayer perceptron. Each sub-function, designated as $f^{(i)}$, is represented in the structure of an MLP as a *layer*, which contains a linear mapping and a nonlinear mapping in form of an *activation function*. A multitude of *Units* (also *neurons*) compose each layer. A neuron performs the same vector-to-scalar calculation as the perceptron does. Subsequently, a nonlinear activation function transforms the scalar output into the activation of the unit. The layers are staggered in the neural network, with each layer being connected to its neighbors, as illustrated in Figure 2.3. The input vector $\boldsymbol{x}$ is provided to each

unit of the first layer $f^{(1)}$, which then gives the results to the units of the second layer $f^{(2)}$, and so forth. The final layer is the *output layer*. The intervening layers, situated between the first and the output layers are the *hidden layers*. The term *forward propagation* describes the process of information flowing through the network from the input layer to the output layer, resulting in a scalar loss. The alternating structure of linear and nonlinear calculation enables MLP's to approximate any function. As Hornik *et al.* [11] proves, MLP's are universal approximators.

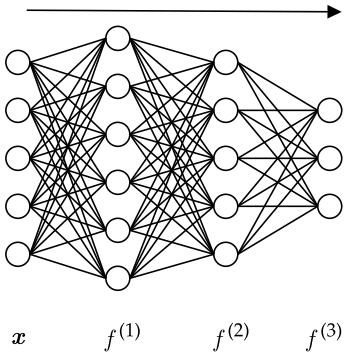


Figure 2.3: A illustration of an MLP with two hidden layers. Each neuron of a layer is connected to every neuron of the neighboring layers. The arrow indicates the direction of the forward propagation.

The term *training* describes the process of optimizing the parameters θ. In order to undertake training, it is necessary to have a set of *training data*, which is a set of pairs (also called training points) of the input data $\boldsymbol{x}$ and its corresponding true solution $\boldsymbol{y}$ of the function f^*. For the training process we must define a *loss function* $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$, using the model prediction $\hat{\boldsymbol{y}}$ and the true value $\boldsymbol{y}$, which will act as a metric for evaluating the extent to which the model deviates from the correct answer. One common loss function is the *mean square error* (MSE) loss function. Let N be the number of points in the set of training data. Then,

$$\mathcal{L}_{MSE}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{N} \sum_{i=1}^{N} ||\hat{\boldsymbol{y}}^{(i)} - \boldsymbol{y}^{(i)}||^2, \tag{2.18}$$

calculates the squared difference between each model prediction and true value of a training and takes the mean across the whole training data.

Ultimately, the objective is to utilize this information to optimize the parameters, in order to minimize the loss. One of the most fundamental optimization strategy is *gradient descent*. In this process, the derivatives are employed to identify the location of local or global minima within a function, which lie where the gradient is zero. Given that a positive gradient signifies ascent and a negative gradient indicates descent, we must move the variable by a *learning rate* (step size) in the opposite direction to that of the gradient. The calculation of the derivatives in respect to the parameters is a complex task, since our functions is a composition of many functions (one for each layer). We can address this issue taking advantage of Equation (2.17) and employing the chain rule of calculus. Let $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \theta)$ be the model prediction with the decomposed version $f(\boldsymbol{x}; \theta) = f^{(3)}(w; \theta_3)$ with $w = f^{(2)}(z; \theta_2)$ and $z = f^{(1)}(\boldsymbol{x}; \theta_1)$. $\boldsymbol{x}$ is the input vector and $\theta_3, \theta_2, \theta_1 \subset \theta$. Then,

$$\nabla_{\theta_3} \mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{d\mathcal{L}}{d\hat{\boldsymbol{y}}} \frac{d\hat{\boldsymbol{y}}}{df^{(3)}} \nabla_{\theta_3} f^{(3)}, \tag{2.19}$$

is the gradient of $\mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ in respect of the parameters θ_3. To obtain $\nabla_{\theta_2} \mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$, we have to derive $\nabla_{\theta_3} \mathcal{L}(\hat{\boldsymbol{y}}, \boldsymbol{y})$ in respect to θ_2. The name of this method in the context of neural networks is *back propagation* [30], as it propagates the error backwards through the neural network.

In practical applications, an optimizer often accomplishes the optimization task by executing back propagation in the background. Furthermore, modifying the learning rate during training can be advantageous. For instance, making larger steps at the beginning and minor adjustments at the end. Therefore, schedulers are implementations algorithms that employ diverse learning rate alteration strategies.

leave whole paragraph out? - Niklas

For a more in-depth discussion of practical considerations and additional details like regularization, we direct the reader to the book *Deep Learning* by Goodfellow *et al.* [10]. The next section will demonstrate the application of neural networks in approximating solutions to differential systems.

2.5 Physics Informed Neural Networks

In Section 2.4, we describe the structure and training of MLP's, which are wildely recognized tools for approximating any kind of function. In this section, we apply this capability to create a solver for ODE's and PDE's as Legaris *et al.* [15] describe

in their paper. In this approach, the model learns to approximate a function using provided data points while leveraging the available knowledge about the problem in the form of a system of differential equations. The *physics-informed neural network* (PINN) learns the system of differential equations during training, as it optimizes its output to align with the equations.

In contrast to standard MLP's, PINNs are not only data-driven. The loss term of a PINN comprises two components. The first term incorporates the equations of the aforementioned prior knowledge to pertinent the problem. As Raissi *et al.* [23] propose, the residual of each differential equation in the system must be minimized in order for the model to optimize its output in accordance with the theory. We obtain the residual r_i, with $i \in \{1, ..., N_d\}$, by rearranging the differential equation and calculating the difference between the left-hand side and the right-hand side of the equation. N_d is the number of differential equations in a system. As Raissi *et al.* [23] propose the *physics loss* of a PINN,

$$\mathcal{L}_{physics}(\boldsymbol{x}, \hat{\boldsymbol{y}}) = \frac{1}{N_d} \sum_{i=1}^{N_d} ||r_i(\boldsymbol{x}, \hat{\boldsymbol{y}})||^2, \tag{2.20}$$

takes the input data and the model prediction to calculate the mean square error of the residuals. The second term, the *observation loss* $\mathcal{L}_{obs}(\hat{\boldsymbol{y}}, \boldsymbol{y})$, employs the mean square error of the distances between the predicted and the true values for each training point. Additionally, the observation loss may incorporate extra terms of inital and boundary conditions. Let N_t denote the number of training points. Then,

$$\mathcal{L}_{PINN}(\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{1}{N_d} \sum_{i=1}^{N_d} ||r_i(\boldsymbol{x}, \hat{\boldsymbol{y}})||^2 + \frac{1}{N_t} \sum_{i=1}^{N_t} ||\hat{\boldsymbol{y}}^{(i)} - \boldsymbol{y}^{(i)}||^2, \tag{2.21}$$

represents the comprehensive loss function of a physics-informed neural network.

Given the nature of residuals, calculating the loss term of $\mathcal{L}_{physics}(\boldsymbol{x}, \hat{\boldsymbol{y}})$ requires the calculation of the derivative of the output with respect to the input of the neural network. As we outline in Section 2.4, during the process of back-propagation we calculate the gradients of the loss term in respect to a layer-specific set of parameters denoted by θ_l, where l represents the index of the respective layer. By employing the chain rule of calculus, the algorithm progresses from the output layer through each

hidden layer, ultimately reaching the first layer in order to compute the respective gradients. The term,

$$\nabla_{\boldsymbol{x}} \hat{\boldsymbol{y}} = \frac{d\hat{\boldsymbol{y}}}{df^{(2)}} \frac{df^{(2)}}{df^{(1)}} \nabla_{\boldsymbol{x}} f^{(1)}, \tag{2.22}$$

illustrates that, in contrast to the procedure described in eq. (2.19), this procedure the *automatic differenciation* goes one step further and calculates the gradient of the output with respect to the input $\boldsymbol{x}$. In order to calculate the second derivative $\frac{d\hat{\boldsymbol{y}}}{d\boldsymbol{x}} = \nabla_{\boldsymbol{x}}(\nabla_{\boldsymbol{x}} \hat{\boldsymbol{y}})$, this procedure must be repeated.

Above we present a method for approximating functions through the use of systems of differential equations. As previously stated, we want to find a solver for systems of differential equations. In problems, where we must solve an ODE or PDE, we have to find a set of parameters, that satisfies the system for any input $\boldsymbol{x}$. In terms of the context of PINN's this is the inverse problem, where we have a set of training data from measurements, for example, is available along with the respective differential equations but information about the parameters of the equations is lacking. To address this challenge, we set these parameters as distinct learnable parameters within the neural network. This enables the network to utilize a specific value, that actively influences the physics loss $\mathcal{L}_{physics}(\boldsymbol{x}, \hat{\boldsymbol{y}})$. During the training phase the optimizer aims to minimize the physics loss, which should ultimately yield an approximation of the true value.

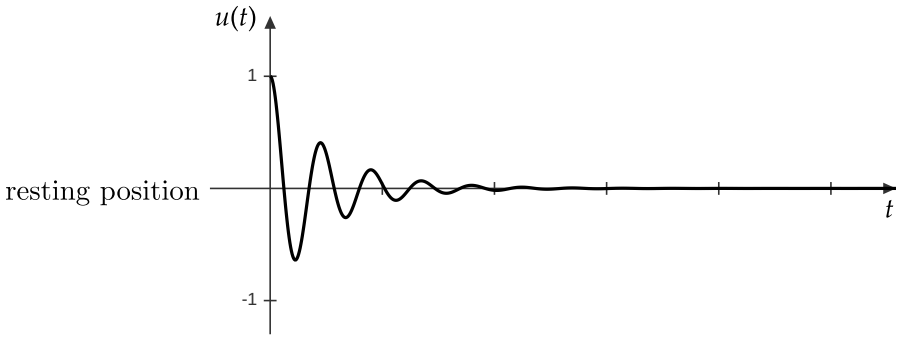


Figure 2.4: Illustration of of the movement of an oscillating body in the underdamped case. With $m = 1kg$, $\mu = 4\frac{Ns}{m}$ and $k = 200\frac{N}{m}$.

One illustrative example of a potential application for PINN's is the *damped harmonic oscillator* [4]. In this problem, we displace a body, which is attached to a spring, from its resting position. The body is subject to three forces: firstly, the in-

ertia exerted by the displacement u, which points in the direction the displacement u; secondly the restoring force of the spring, which attempts to return the body to its original position and thirdly, the friction force, which points in the opposite direction of the movement. In accordance with Newton's second law and the combined influence of these forces, the body exhibits oscillatory motion around its position of rest. The system is influenced by m the mass of the body, μ the coefficient of friction and k the spring constant, indicating the stiffness of the spring. The residual of the differential equation,

$$m\frac{d^2u}{dx^2} + \mu\frac{du}{dx} + ku = 0, \tag{2.23}$$

shows relation of these parameters in reference to the problem at hand. As Tenenbaum and Morris provide, there are three potential solutions to this issue. However only the *underdamped case* results in an oscillating movement of the body, as illustrated in Figure 2.4. In order to apply a PINN to this problem, we require a set of training data x. This consists of pairs of time points and corresponding displacement measurements $(t^{(i)}, u^{(i)})$, where $i \in \{1, ..., N_t\}$. In this hypothetical case, we know the mass $m = 1kg$, and the spring constant $k = 200\frac{N}{m}$ and the initial displacement $u^{(1)} = 1$ and $\frac{du(0)}{dt} = 0$. However, we do not know the value of the friction μ. In this case the loss function,

$$\mathcal{L}_{osc}(\boldsymbol{x}, \boldsymbol{u}, \hat{\boldsymbol{u}}) = (u^{(1)} - 1) + \frac{du(0)}{dt} + ||m\frac{d^2u}{dx^2} + \mu\frac{du}{dx} + ku||^2 + \frac{1}{N_t}\sum_{i=1}^{N_t} ||\hat{\boldsymbol{u}}^{(i)} - \boldsymbol{u}^{(i)}||^2, \tag{2.24}$$

includes the border conditions, the residual, in which μ is a learnable parameter and the observation loss.

2.5.1 Disease Informed Neural Networks

In this section, we describe the capability of MLP's to solve systems of differential equations. In Section 2.3.1, we describe the SIR model, which models the relations of susceptible, infectious and removed individuals and simulates the progress of a disease in a population with a constant size. A system of differential equations models these relations. Shaier *et al.* [32] propose a method to solve the equations of the SIR model using a PINN, which they call a *disease-informed neural network* (DINN).

To solve Equation (2.8) we need to find the transmission rate β and the recovery rate α. As Shaier *et al.* [32] point out, there are different approaches to solve this set

of equations. For instance, building on the assumption, that at the beginning one infected individual infects $-n$ other people, concluding in $\frac{dS(0)}{dt} = -n$. Then,

$$\beta = -\frac{\frac{dS}{dt}}{S_0 I_0} \tag{2.25}$$

would calculate the initial transmission rate using the initial size of the susceptible group S_0 and the infectious group I_0. The recovery rate, then could be defined using the amount of days a person between the point of infection and the start of isolation d, $\alpha = \frac{1}{d}$. The analytical solutions to the SIR models often use heuristic methods and require knowledge like the sizes S_0 and I_0. A data-driven approach such as the one that Shaier *et al.* [32] propose does not have these problems. Since the model learns the parameters β and α while learning the training data consisting of the time points $\boldsymbol{t}$, and the corresponding measured sizes of the groups $\boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R}$. Let $\hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}}$ be the model predictions of the groups and $r_S = \frac{d\hat{\boldsymbol{S}}}{dt} + \beta\hat{\boldsymbol{S}}\hat{\boldsymbol{I}}, r_I = \frac{d\hat{\boldsymbol{I}}}{dt} - \beta\hat{\boldsymbol{S}}\hat{\boldsymbol{I}} + \alpha\hat{\boldsymbol{I}}$ and $r_R = \frac{d\hat{\boldsymbol{R}}}{dt} - \alpha\hat{\boldsymbol{I}}$ the residuals of each differential equation using the model predictions. Then,

$$\begin{aligned}\mathcal{L}_{SIR}() = ||r_S||^2 + ||r_I||^2 + ||r_R||^2 + \frac{1}{N_t}\sum_{i=1}^{N_t} ||\hat{\boldsymbol{S}}^{(i)} - \boldsymbol{S}^{(i)}||^2 + \\ ||\hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)}||^2 + \\ ||\hat{\boldsymbol{R}}^{(i)} - \boldsymbol{R}^{(i)}||^2,\end{aligned} \tag{2.26}$$

is the loss function of a DINN, with α and *beta* being learnable parameters.

Chapter 3

Methods

This chapter provides the methods, that we employ to address the problem that we present in Chapter 1. Section 3.1 outlines our approaches for preprocessing of the available data and has two sections. The first section describes the publicly available data provided by the *Robert Koch Institute* (RKI)[1]. The second section outlines the techniques we use to process this data to fit our project's requirements. Subsequently, we give a theoretical overview of the PINN's that we employ. These latter sections, establish the foundation for the implementations described in Section 4.1.1 and Section 4.2.1.

3.1 Epidemiological Data

In order for the PINNs to be effective with the data available to us, it is necessary for the data to be in the format required by the epidemiological models, which the PINNs will solve. Let N_t be the number of training points, then let $i \in \{1, ..., N_t\}$ be the index of the training points. The data required by the PINN for solving the SIR model (see Section 2.5.1), consists of pairs $(\boldsymbol{t}^{(i)}, (\boldsymbol{S}^{(i)}, \boldsymbol{I}^{(i)}, \boldsymbol{R}^{(i)}))$. Given that the system of differential equations representing the reduced SIR model (see Section 2.3.2) consists of a single differential equation for I, it is necessary to obtain pairs of the form $(\boldsymbol{t}^{(i)}, \boldsymbol{I}^{(i)})$. This section, focuses on the structure of the available data and the methods we employ to transform it into the correct structure.

3.1.1 RKI Data

The Robert Koch Institute is responsible for the on monitoring and prevention of diseases. As the central institution of the German government in the field of biomedicine, one of its tasks during the COVID-19 pandemic was it to track the

[1] `https://www.rki.de/EN/Home/homepage_node.html`

number of infections and death cases in Germany. The data was collected by university hospitals, research facilities and laboratories through the conduction of tests. Each new case must be reported within a period of 24 hours at the latest to the respective state authority. Each state authority collects the cases for a day and must report them to the RKI by the following working day. The RKI then refines the data and releases statistics and updates its repositories holding the information for the public to access. For the purposes of this thesis we concentrate on two of these repositories.

The first repository is called *COVID-19-Todesfälle in Deutschland*[2]. The dataset comprises discrete data points, each with a date indicating the point in time at which the respective data was collected. The dates span from March 9, 2020, to the present day. For each date, the dataset provides the total number of infection and death cases, the number of new deaths, and the case-fatality ratio. The total number of infection and death cases represents the sum of all cases reported up to that date, including the newly reported data. The dataset includes two additional datasets, that contain the death case information organized by age group or by the individual states within Germany on a weekly basis.

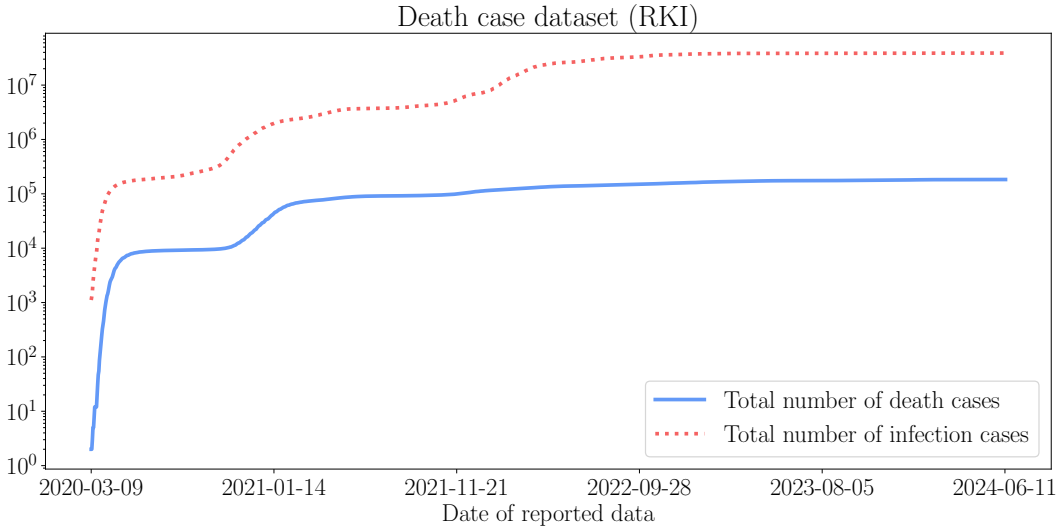


Figure 3.1: A visualization of the total death case and infection case data for each day from the data set *COVID-19-Todesfälle in Deutschland*. Status of the 20'th of August 2024.

[2] `https://github.com/robert-koch-institut/COVID-19-Todesfaelle_in_Deutschland.git`

The second repository is entitled *SARS-CoV-2 Infektionen in Deutschland*[3]. This dataset contains comprehensive data regarding the infections of each county on a daily basis. The counties are encoded using the *Community Identification Number*[4], wherein the first two digits denote the state, the third digit represents the government district, and the last two digits indicate the county. Each data point displays the gender, the age group, number death, infection and recovery cases and the reference and report date. The reference date marks the onset of illness in the individual. In the absence of this information, the reference date is equivalent to the report date.

The RKI assumes that the duration of the illness under normal conditions is 14 days, while the duration of severe cases is assumed to be 28 days. The recovery cases in the dataset are calculated using these assumptions, by adding the duration on the reference date if it is given. As stated in the ReadMe, the recovery data should be used with caution. Since we require the recovery data for further calculations, the following section presents the solutions we employed to address this issue.

3.1.2 Data Preprocessing

At the outset of this section, we establish the format of the data, that is necessary for training the PINNs. In this subsection, we present the method, that we employ to preprocess and transform the RKI data (see Section 3.1.1) into the training data.

In order to obtain the SIR data we require the size of each SIR compartment for each time point. The infection case data for the German states is available on a daily basis. To obtain the daily cases for the entire country we need to differentiate the total number of cases. The size of the population is defined as the respective size at the beginning of 2020. Using the starting conditions of Equation (2.10), we iterate through each day, modifying the sizes of the groups in a consecutive manner. For each iteration we subtract the new infection cases from $\boldsymbol{S}^{(i-1)}$ to obtain $\boldsymbol{S}^{(i)}$, for $\boldsymbol{I}^{(i)}$, we add the new cases and subtract deaths and recoveries, and the size of $\boldsymbol{R}^{(i)}$ is obtained by adding the new deaths and recoveries as they occur.

As previously stated in Section 3.1.1 the data on recoveries may either be unreliable or is entirely absent. To address this, we propose a method for computing

[3] `https://github.com/robert-koch-institut/SARS-CoV-2-Infektionen_in_Deutschland.git`
[4] `https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html`

the number of recovered individuals per day. Under the assumption that recovery takes D days, we present the recovery queue, a data structure that holds the number of infections for a given day, retains them for D days, and releases them into the removed group D days later.

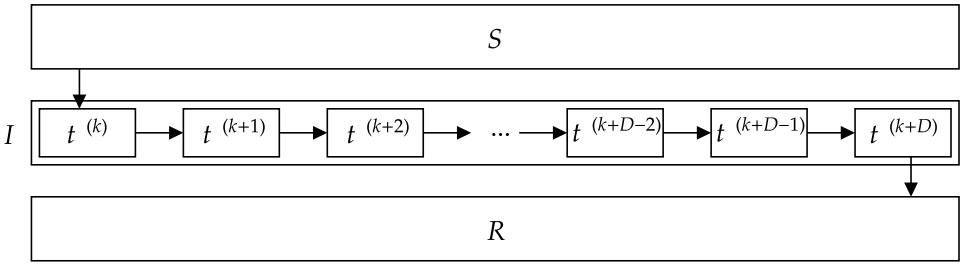


Figure 3.2: The recovery queue takes in the infected individuals for the k'th day and releases them D days later into the removed group.

In order to solve the reduced SIR model, we employ a similar algorithm to that used for the SIR model. However, in contrast to the recovery queue, we utilize the set recovery rate α to transfer a portion $\alpha \boldsymbol{I}^{(i)}$ of infections, which have recovered on the i and put them into the $\boldsymbol{R}^{(i)}$ compartment, which is irrelevant to our purposes.

The transformed data for both the SIR model and the reduced SIR model are then employed by the PINN models, which we describe in the subsequent section.

3.2 Estimating Epidemiological Parameters using PINNs

In the preceding section, we present the methods we employ to preprocess and format the data from the RKI in accordance with the specifications required for the work of this thesis. In this section, we will present the method we employ to identify the non-time-dependent SIR parameters β and α for the data. As a foundation for our work, we draw upon the work of Shaier et al. [32], to solve the SIR system of ODEs using PINNs.

In order to conduct an analysis of a pandemic, it is necessary to have a quantifiable measure that indicates whether the disease in question has the capacity to spread rapidly through a population or is it not successful in infecting a significant number of individuals. We employ the SIR model to construct an abstraction of

the complex relations inherent to real-world pandemics. The SIR model divides the population into three compartments. It is accompanied by a with system of ODEs that encapsulates the fluctuations and relationships between these compartments (see Equation (2.8)). The transmission rate β and the recovery rate α work as the aforementioned quantifiers. We obtain data from the preprocessing stage. It provides insight into the progression of the COVID-19 pandemic in Germany. The objective is to identify a function that solves the system of differential equations of the SIR model, by returning the size of each compartment at a specific point in time. This function is supposed to be able to reconstruct the training data and is defined by the values of the transition rates β and α. From a mathematical and semantic perspective, it is essential to determine these values of the parameter.

In order to ascertain the transmission rate β and the recovery rate α from the preprocessed RKI data of $(\boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R})$ for a given set of time points, it is necessary to employ a data-driven approach that outputs a model prediction of $(\hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}})$ for a set of time points, with the aim of minimizing the term,

$$\left\|\hat{\boldsymbol{S}}^{(i)} - \boldsymbol{S}^{(i)}\right\|^2 + \left\|\hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)}\right\|^2 + \left\|\hat{\boldsymbol{R}}^{(i)} - \boldsymbol{R}^{(i)}\right\|^2, \tag{3.1}$$

for each data point in the set of training dataset of a cardinality $N_t t$ and with $i \in \{1, ..., N_t\}$. Moreover, the aforementioned parameters must satisfy the system of differential equations that govern the SIR model. For this reason, Shaier *et al.* [32] utilize a PINN framework to satisfy both requirements. Their approach, which they refer to as the *disease-informed neural network* (see Section 2.5.1), takes epidemiological data as the input and returns the two transition rates α and β. This method achieves this by finding an approximate solution of to the inverse problem of physics-informed neural networks (see Section 2.5). In terms of the terms of the SIR model, a PINN addresses the inverse problemin two ways. First, it minimizes the mean of Equation (3.1) by bringing the model predictions $(\boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R})$ closer to the actual values $(\hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}})$ for each time point. Second, it reduces the residuals of the ODEs that constitute the SIR model. While the forward problem concludes at this point, the inverse problem presets that a parameter is unknown. Thus, we designate the parameters β and α as free, learnable parameters, $\widehat{\beta}$ and $\widehat{\alpha}$. These separate trainable parameters are values that are optimized during the training process and must fit the equations of the set of ODEs. Furthermore, we know, that the transition rates do not surpass the value of 1. Consequently, we force the value of both rates to

be in a range of $[-1, 1]$. Therefor, we regularize the parameters using the *tangens hyperbolicus*. This results in the terms,

$$\widehat{\beta} = \tanh(\tilde{\beta}), \quad \widehat{\alpha} = \tanh(\tilde{\alpha}), \tag{3.2}$$

where $\tilde{\beta}$ and $\tilde{\alpha}$ are the predicted values of the model and $\widehat{\beta}$ and $\widehat{\alpha}$ are regularized model predictions.

The input data must include the time point $\boldsymbol{t}^{(i)}$ and its corresponding measured true values of $(\boldsymbol{S}^{(i)}, \boldsymbol{I}^{(i)}, \boldsymbol{R}^{(i)})$. In its forward path, the PINN receives the time point $\boldsymbol{t}^{(i)}$ as its input, from which it calculates its model prediction $(\hat{\boldsymbol{S}}^{(i)}, \hat{\boldsymbol{I}}^{(i)}, \hat{\boldsymbol{R}}^{(i)})$ based on its model parameters θ. Subsequently, the model computes the loss function. It calculates the observation loss by taking the mean squared error of Equation (3.1) over all N_t training samples. Therefore, the term for the observation loss is,

$$\mathcal{L}_{\text{obs}}(\boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R}, \hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \left\| \hat{\boldsymbol{S}}^{(i)} - \boldsymbol{S}^{(i)} \right\|^2 + \left\| \hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)} \right\|^2 + \left\| \hat{\boldsymbol{R}}^{(i)} - \boldsymbol{R}^{(i)} \right\|^2, \tag{3.3}$$

is the term for the observation loss. Given superior performance in practical applications relative to the ODEs of Equation (2.8), we utilize the ODEs of Equation (2.9) in our physics loss. In order for the model to learn the system of differential, it is necessary to obtain the residual of each ODE. The mean square error of the residuals constitutes the physics loss $\mathcal{L}_{\text{physiks}}(\boldsymbol{t}, \boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R}, \hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}})$. The residuals are calculated using the model predictions $(\hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}})$ and the regularized model predictions of the parameters $\widehat{\beta}$ and $\widehat{\alpha}$. The residuals are given by,

$$0 = \frac{d\hat{\boldsymbol{S}}}{d\boldsymbol{t}} + \widehat{\beta}\frac{\hat{\boldsymbol{S}}\hat{\boldsymbol{I}}}{N}, \quad 0 = \frac{d\hat{\boldsymbol{I}}}{d\boldsymbol{t}} - \widehat{\beta}\frac{\hat{\boldsymbol{S}}\hat{\boldsymbol{I}}}{N} + \widehat{\alpha}\hat{\boldsymbol{I}}, \quad 0 = \frac{d\hat{\boldsymbol{R}}}{d\boldsymbol{t}} + \widehat{\alpha}\hat{\boldsymbol{I}}. \tag{3.4}$$

Thus,

$$\begin{aligned} \mathcal{L}_{\text{SIR}}(\boldsymbol{t}, \boldsymbol{S}, \boldsymbol{I}, \boldsymbol{R}, \hat{\boldsymbol{S}}, \hat{\boldsymbol{I}}, \hat{\boldsymbol{R}}) = &\left\| \frac{d\hat{\boldsymbol{S}}}{d\boldsymbol{t}} + \widehat{\beta}\frac{\hat{\boldsymbol{S}}\hat{\boldsymbol{I}}}{N} \right\|^2 + \left\| \frac{d\hat{\boldsymbol{I}}}{d\boldsymbol{t}} - \widehat{\beta}\frac{\hat{\boldsymbol{S}}\hat{\boldsymbol{I}}}{N} + \widehat{\alpha}\hat{\boldsymbol{I}} \right\|^2 + \left\| \frac{d\hat{\boldsymbol{R}}}{d\boldsymbol{t}} + \widehat{\alpha}\hat{\boldsymbol{I}} \right\|^2 \\ &+ \frac{1}{N_t} \sum_{i=1}^{N_t} \left\| \hat{\boldsymbol{S}}^{(i)} - \boldsymbol{S}^{(i)} \right\|^2 + \left\| \hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)} \right\|^2 + \left\| \hat{\boldsymbol{R}}^{(i)} - \boldsymbol{R}^{(i)} \right\|^2, \end{aligned} \tag{3.5}$$

is the equation of the total loss for our approach. This loss value is then back-propagated through our network, while the model predictions of the parameters β

and α are optimized using the loss as well.

As this section concentrates on the finding of the time constant parameters β and α, the next section will show our approach of finding the reproduction number $\mathcal{R}_t$ on the German data of the RKI.

3.3 Estimating the Reproduction Number using PINNs

The previous section illustrates the methodology we employ to detemine the constant transmission and recovery rates from a data set obtained from the COVID-19 pandemic in Germany. In this section, we utilize PINNs to identify the time-dependent reproduction number, $\mathcal{R}_t$, while reducing the number of state variables and the reliance on assumptions, by reducing the system of ODEs comprising the SIR model. The methodology presented in this section is based on the approach developed by Millevoi *et al.* [18].

In real-world pandemics, the rate of infection is influenced by a multitude of factors. Events such as the growing awareness for the disease among the general population, the introduction of non-pharmaceutical mitigations such as social distancing policies, and the emergence of a new variants have an impact on the transmission rate β. Accordingly, a transmission rate that is not time-dependent and constant across the entire duration of the pandemic may not accurately reflect the dynamics of the spread of a real-world disease correctly. Although we set the transmission rate to be time-dependent, the recovery time is assumed to be relatively constant over time. The Robert Koch Institute[5] posits that the typical recovery period for the illness under normal conditions is 14 days, while those individuals with severe cases require approximately 28 days to recover. In the light of the negligible number of severe cases in comparison to the number of normal cases, we can set the recovery time to $D = 14$, which yields $\alpha = 1/14$. The reproduction number, $\mathcal{R}_t$ (see Section 2.3.2), represents the number of new infections that occur as a result of one infectious individual. It indicates whether a pandemic is emerging or if it is spreading rapidly through the susceptible population. By inserting the definition of Equation (2.13), into the system of ODEs of the SIR model, we can derive one Equation (2.16). In order to solve this, we must identify a function that maps a time point to the size of

[5] `https://github.com/robert-koch-institut/SARS-CoV-2-Infektionen_in_Deutschland.git`

the infectious compartment and the specific reproduction number.

As with the constant transition rates, we employ a data-driven approach for identifying the time-dependent reproduction number $\mathcal{R}_t$. The PINN approximates the size $\boldsymbol{I}$ with its model prediction $\hat{\boldsymbol{I}}$ by minimizing the term,

$$\left\| \hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)} \right\|^2, \tag{3.6}$$

for each $i \in \{1, ..., N_t\}$. In order to identify the reproduction number, the PINN minimizes the residuals of the ODE during the training process. The training process is analogous to that of the PINN, which identifies β and α (see Section 3.2). However, there are two key differences. Firstly, the absence of trainable parameters. Secondly, the inclusion of an additional state variable that fluctuates in response to the input. While the state variable $\boldsymbol{I}$ is approximated using the error between the training data and the predicted values, the state variable $\mathcal{R}_t$ is approximated exclusively based on the residual of the ODE.

The PINN receives the input of $\boldsymbol{t}^{(i)}$ and generates a prediction of $(\hat{\boldsymbol{I}}^{(i)}, \mathcal{R}_t^{(i)})$. As previously stated, the PINN minimizes the distance between the true values of $\boldsymbol{I}$ and the model predictions $\hat{\boldsymbol{I}}$ by minimizing the mean squared error. Consequently, the observation loss function is defined by,

$$\mathcal{L}_{\text{rSIR}}(\boldsymbol{I}, \hat{\boldsymbol{I}}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \left\| \hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)} \right\|^2. \tag{3.7}$$

The physics loss function is defined as the squared error of the residual of the ODE. The residual of the reduced SIR model is given by,

$$0 = \frac{dI_s}{dt_s} - \alpha(t_f - t_0)(\mathcal{R}_t - 1)I_s(t_s). \tag{3.8}$$

By combining the observation loss with the physics loss, we arrive at the total loss for the PINN that solves the reduced SIR model, which is given by,

$$\mathcal{L}_{\text{rSIR}}(\boldsymbol{t}, \boldsymbol{I}, \hat{\boldsymbol{I}}) = \left\| \frac{dI_s}{dt_s} - \alpha(t_f - t_0)(\mathcal{R}_t - 1)I_s(t_s) \right\|^2 + \frac{1}{N_t} \sum_{i=1}^{N_t} \left\| \hat{\boldsymbol{I}}^{(i)} - \boldsymbol{I}^{(i)} \right\|^2. \tag{3.9}$$

The process of determining the reproduction number, along with the other techniques, that this chapter presents find application in the following chapter.

Chapter 4

Experiments

In Chapter 3, we explain the methods based the theoretical background, that we established in Chapter 2. In this chapter, we present the setups and results from the experiments and simulations. First, we discuss the experiments dedicated to identify the epidemiological transition rates of β and α in synthetic and real-world data. Second, we examine the reproduction number in synthetic and real-world data of Germany.

4.1 Identifying the Transition Rates

In this section, we aim to identify the transmission rate β and the recovery rate α from either synthetic or preprocessed real-world data. The methodology that we employ to identify the transition rates is described in Section 3.2. Meanwhile, the methods we utilize to preprocess the real-world data are detailed in Section 3.1.2. In the first part we present the setup of our experiments, then we provide the results including a discussion.

4.1.1 Setup

Synthetic Data: In order to validate our method, we first generate a dataset of synthetic data. We achieve this by solving Equation (2.9) for a given set of parameters. The parameters are set to $\alpha = 1/3$ and $\beta = 1/2$. The size of the population is $N = 7.6\text{e}6$ and the initial amount of infectious individuals is $I_0 = 10$. We conduct the simulation over 150 days, resulting in a dataset of the form of Section 4.1.1.

Real-World Data: In order to process the real-world RKI data, it is necessary to preprocess the raw data for each state and Germany separately. This is achieved by utilizing a recovery queue with a recovery period of 14 days. With regard to population size of each state, we set it to the respective value counted at the end of 2019[6]. The initial number of infectious individuals is set to the number of infected people on March 09. 2020 from the dataset. The data we extract spans from March 09. 2020 to June 22. 2023, encompassing a period of 1200 days and representing the time span during which the COVID-19 disease was the most active and severe.

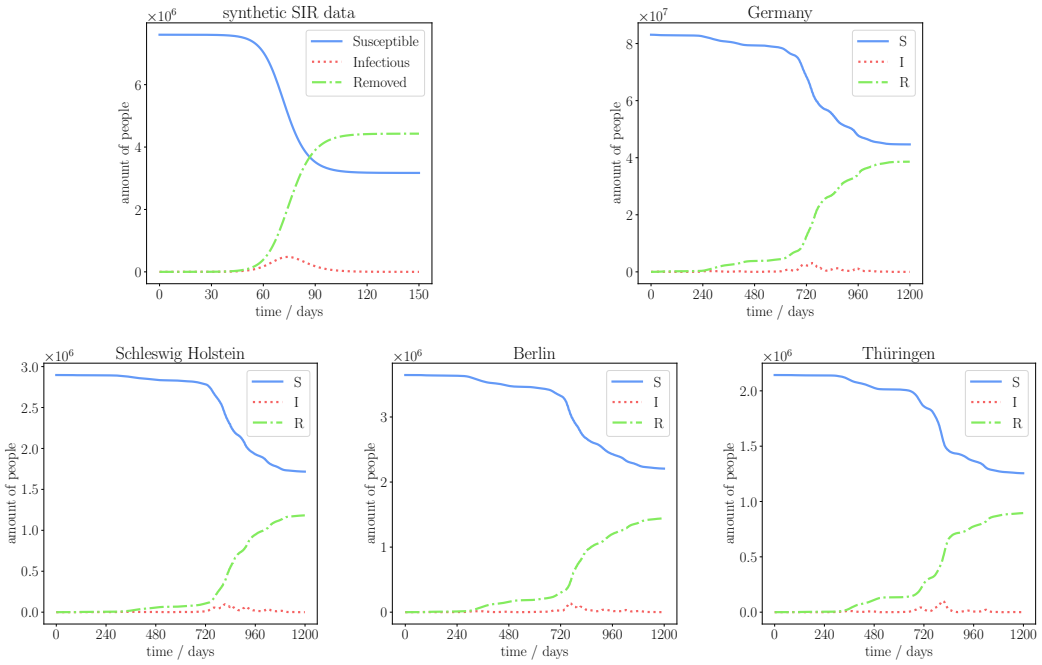


Figure 4.1: Synthetic and real-world training data. The synthetic data is generated with $\alpha = 1/3$ and $\beta = 1/2$ and Equation (2.9). The Germany data is taken from the death case data set. Exemplatory we show illustrations of the datasets of Schleswig Holstein, Berlin, and Thuringia. For the other states see Chapter 6

Training Parameters: The PINN that we utilize comprises of seven hidden layers with twenty neurons each, and an activation function of ReLU. We follow the hyperparameter setting in [32] but change the base learning rate to 1e−3. And employ a polynomial scheduler implementation from the PyTorch library [22] instead. We

[6] https://de.statista.com/statistik/kategorien/kategorie/8/themen/63/branche/demographie/#overview

train the model for 10000 epochs to extract the parameters. For each set of parameters, we conduct five iterations to demonstrate stability of the values. For measuring the accuracy, we calculate the error e, using the 2-Norm. Let G be the set of compartment training data the SIR model with $\boldsymbol{g} \in G$ and $\hat{\boldsymbol{g}}$ be the corresponding model prediction, then,

$$e_G = \frac{1}{|G|} \sum_{g \in G} \frac{\left\| \hat{\boldsymbol{g}} - \boldsymbol{g} \right\|_2}{\left\| \boldsymbol{g} \right\|_2}, \tag{4.1}$$

is the average error across all three groups.

4.1.2 Results

In this section, we start by examining the results for the synthetic dataset, focusing the accuracy and reproducibility. We then proceed to present and discuss the results for the German states and Germany.

The results of the experiment regarding the synthetic data can be seen in Table 4.1. The error and the standard variation for both parameters are negligible small. Taking the mean of the parameters across the five iterations yields more accurate results.

The results demonstrate that the model is capable of approximating the correct

Table 4.1: Simulation results for the synthetic data. The true values and the respective mean parameter is given.

α		β		
true	μ	true	μ	e_{SIR}
0.333	$0.333_{\pm 0.001}$	0.500	$0.500_{\pm 0.002}$	0.004

parameters for the small, synthetic dataset in each of the five iterations. The mean of the predicted values results in values with a sufficiently small error. Thus, we argue that our selected method is well suited to analyze real world pandemic data collected in Germany.

In Table 4.2 we present the results of the training for the real-world data. The results are presented from top to bottom, in the order of the community identification number, with the last entry being Germany. Both the mean μ and the standard de-

viation σ are calculated across all five iterations of our experiment. We can observe that the error e_{SIR} is the highest for *Saxony* and the lowest for *Lower Saxony*. Furthermore, we include the distance $\Delta\beta_{\text{Germany}} = \beta_{\text{state}} - \beta_{\text{Germany}}$ and the percentage of people that have a basic immunity through vaccination ν for each state provided by the Robert Koch Institute[7].

Table 4.2: Mean and standard deviation, error e_{SIR} and the distance $\Delta\beta_{\text{Germany}} = \beta_{\text{state}} - \beta_{\text{Germany}}$ across the 5 iterations, that we conducted for each German state and Germany as the whole country. Furthermore we include the vaccination percentage ν provided from the RKI.

state name	α	β	e_{SIR}	$\Delta\beta_{\text{Germany}}$	ν [%]
Schleswig Holstein	$0.076_{\pm 0.001}$	$0.095_{\pm 0.001}$	0.085	-0.013	79.5
Hamburg	$0.082_{\pm 0.001}$	$0.104_{\pm 0.001}$	0.095	-0.004	84.5
Lower Saxony	$0.075_{\pm 0.002}$	$0.097_{\pm 0.002}$	0.077	-0.011	77.6
Bremen	$0.058_{\pm 0.002}$	$0.078_{\pm 0.002}$	0.093	-0.030	88.3
NRW	$0.079_{\pm 0.001}$	$0.101_{\pm 0.001}$	0.078	-0.007	79.5
Hesse	$0.065_{\pm 0.001}$	$0.085_{\pm 0.001}$	0.102	-0.023	75.8
Rhineland-Palatinate	$0.085_{\pm 0.004}$	$0.108_{\pm 0.004}$	0.090	0.001	75.6
Baden-Württemberg	$0.091_{\pm 0.002}$	$0.118_{\pm 0.003}$	0.080	0.010	74.5
Bavaria	$0.085_{\pm 0.004}$	$0.116_{\pm 0.005}$	0.095	0.008	75.1
Saarland	$0.075_{\pm 0.002}$	$0.099_{\pm 0.003}$	0.108	-0.009	82.4
Berlin	$0.087_{\pm 0.001}$	$0.109_{\pm 0.001}$	0.067	0.001	78.1
Brandenburg	$0.087_{\pm 0.003}$	$0.110_{\pm 0.003}$	0.072	0.002	68.1
MV	$0.089_{\pm 0.002}$	$0.114_{\pm 0.002}$	0.054	0.006	74.7
Saxony	$0.075_{\pm 0.002}$	$0.099_{\pm 0.002}$	0.111	-0.009	65.1
Saxony-Anhalt	$0.092_{\pm 0.003}$	$0.119_{\pm 0.005}$	0.079	0.011	74.1
Thuringia	$0.091_{\pm 0.002}$	$0.119_{\pm 0.003}$	0.084	0.011	70.3
Germany	$0.083_{\pm 0.001}$	$0.108_{\pm 0.002}$	0.080	0.000	76.4

In Figure 4.2, we present a visual representation of the means and standard deviations in comparison to the national values. It is noteworthy that the states of Saxony-Anhalt and Thuringia have the highest transmission rates of all states, while Bremen and Hesse have the lowest values for β. The transmission rates of Hamburg, Baden Württemberg, Bavaria, and all eastern states lay above the national rate of transmission. Similarly, the recovery rate yields comparable outcomes. For the recovery rate, the same states that exhibit a transmission rate exceeding the national

[7] https://impfdashboard.de/

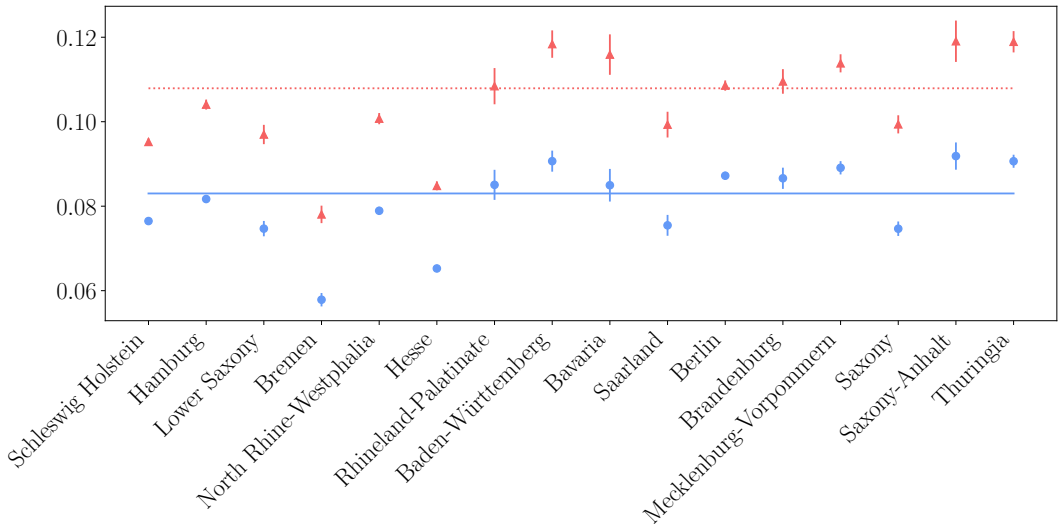


Figure 4.2: Visualization of the mean and standard deviation of the transition rates α and β for each state compared to the mean values of α and β for Germany.

value, have a higher recovery rate than the national standard, with the exception of Saxony. It is noteworthy that the recovery rates of all states exhibit a tendency to align with the recovery rate of $\alpha = 1/14$, which is equivalent to a recovery period of $D = 1/\alpha = 14$ days. When calculating the correlation coefficient between the predicted transmission rate and the vaccination ratio, we get a value of -0.5134. The strong negative correlation indicates that the transmission rate is high when the vaccination ratio is low, and vice versa. This shows that the impact of the vaccines can be witnessed in our results.

It is evident that there is a correlation between the values of α and β for each state. States with a high transmission rate tend to have a high recovery rate, and vice versa. The correlation between α and β can be explained by the implicate definition of α using a recovery queue with a constant recovery period of 14 days. This might result to the PINN not learning α as a standalone parameter but rather as a function of the transmission rate β. This phenomenon occurs because the transmission rate determines the number of individuals that get infected per day, and the recovery queue moves a proportional number of people to the removed compartment. Consequently, a number of people defined by β move to the R compartment 14 days after they were infected.

This issue can be addressed by reducing the SIR model, thereby eliminating the significance of the R compartment size. In the following section, we present our experiments for the reduced SIR model with time-dependent parameters.

4.2 Identifying the Reproduction Number

In this section we describe the experiments we conduct to identify the time-dependent reproduction number for both synthetic and real-world data. Similar to the previous section, we first describe the setup of our experiments and afterwards present the results and a discussion. The methods we employ for the preprocessing are described in Section 3.1.2 and for the PINN, that we use, are described in Section 3.3.

4.2.1 Setup

Synthetic Data: For the purposes of validation, we create a synthetic dataset, by setting the parameter of α and the reproduction value each to a specific values, and solving Equation (2.16) for a given time interval. As in the synthetic data for the aforementioned experiments, we set $\alpha = 1/3$ and $\mathcal{R}_t$ to the values as can be seen in Figure 4.3 as well as the population size $N = 7.6\text{e}6$ and the initial amount of infected people to $I_0 = 10$. Furthermore, we set our simulated time span to 150 days. We use this dataset to demonstrate, that our method is working on a simple and minimal dataset.

Real-World Data: To obtain a dataset of the infectious group, consisting of the real-world data, we processed the data of the dataset *COVID-19-Todesfälle in Deutschland* [25] to extract the number of infections in Germany as a whole. For the German states, we use the data of *SARS-CoV-2 Infektionen in Deutschland* [26]. In the preprocessing stage, we employ a constant rate for α to move individuals into the removed compartment. For each state we generate two datasets with a different recovery rate. First, we choose $\alpha = 1/14$, which aligns with the time of recovery [26]. Second, we use $\alpha = 1/5$, as 5 days into the infection is the point at which the infectiousness is at its peak [7]. As in Section 4.1, we set the population size N of each state and Germany to the corresponding size at the end of 2019. Furthermore, for the same reason we restrict the data points to an interval of 1200 days, beginning on March 09. 2020.

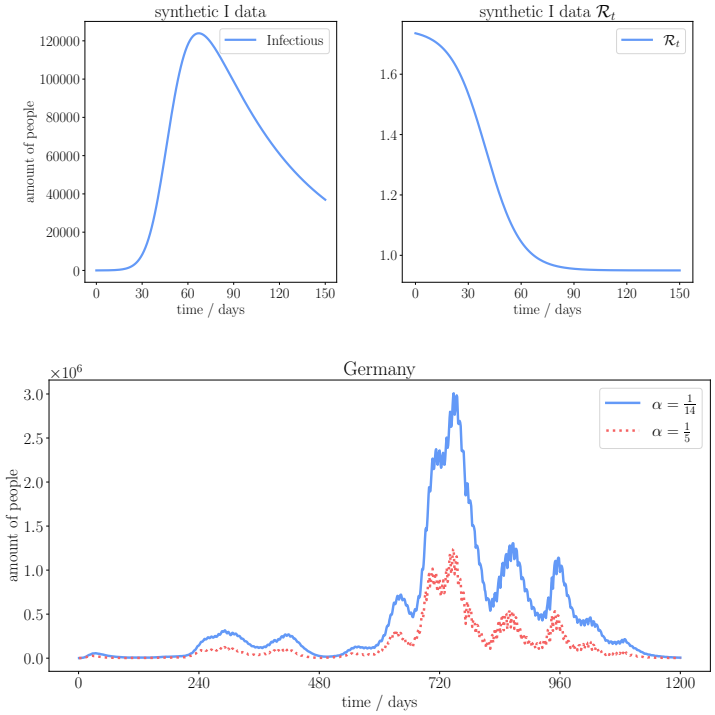


Figure 4.3: The upper two graphics show the curve of the size of the infectious group (left) and the corresponding true reproduction value $\mathcal{R}_t$ (right) for the synthetic data. The lower graphic exemplary illustrates the different curves for Germany.

Training Parameters: In order to achieve the desired output, the selected neural network architecture comprises of four hidden layers, each containing 100 neurons. The activation function is the tangens hyperbolicus function. For both the federal state and Germany, the physics loss is weighted by a factor of 1e−6, whereas the data loss belonging to Germany is also weighted with a high factor of 1e4, relative to the total loss. We found this approach to yield the best results. The model is trained using a base learning rate of 1e−3, with the same scheduler and optimizer as we describe in Section 4.1.1. We train the model for the states 20000 epochs and start the physics training after 10000 epochs, while we train for Germany for 25000 and start the physics training after 15000 epochs. To reduce the standard deviation, each experiment is conducted 15 times. For evaluation, we use the error e_G as we do in the subsequent section.

4.2.2 Results

Section 4.2.2 illustrates the results of our experiments conducted on the synthetic dataset, which can be seen in Figure 4.3. It is evident that the model is capable of learning the infection data across all data points. The error for this is, $e_I = 0.0016$, which is of a negligible magnitude.

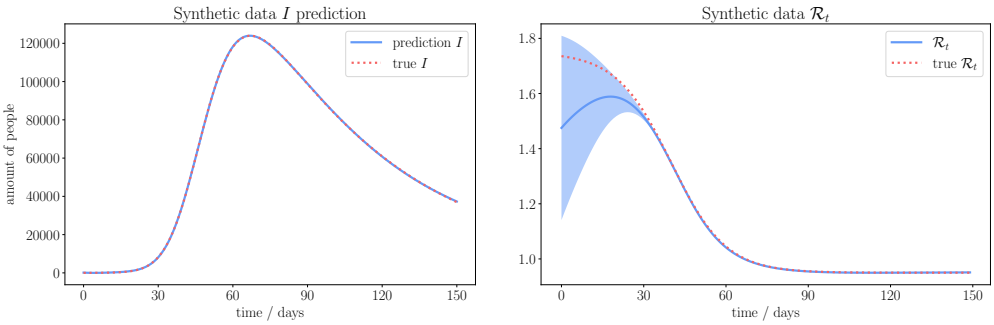


Figure 4.4: Results for the reproduction rate $\mathcal{R}_t$ on synthetic data. The left graphic show the prediction of the model regarding the I group. The right graphic presents the predicted $\mathcal{R}_t$ against the true value, with the standard deviation.

An examination of the predictions for the representation value $\mathcal{R}_t$ reveals that here as well, the model is capable of accurately delineating the value at each time point. However, during the first 30 days, the standard deviation is exhibits an upward trend, while during the final 120 days, the predictions demonstrate remarkable precision.

In Section 4.2.2, we present the graphs of $\mathcal{R}_t$ for the state with the highest value of β, namely Thuringia, and for the state with the lowest transmission rate β, namely Bremen. Further visualizations of the results can be found in Chapter 6. In all datasets, the graphs with $\alpha = 1/5$ are of a smaller size than those with $\alpha = 1/14$. This is due to the fact that the individuals are being moved to the removed compartment at a faster rate. Resulting, it can be observed that the value of $\mathcal{R}_t$ is constantly remaining closer to the threshold of $\mathcal{R}_t = 1$, while the reproduction number for datasets with $\alpha = 1/14$ reaches values of up to 1.6. In states with higher values of β, the period during which the value of $\mathcal{R}_t$ is above the threshold of one 1 is longer, but the peak is lower. In states with a lower transmission rate, the period above 1

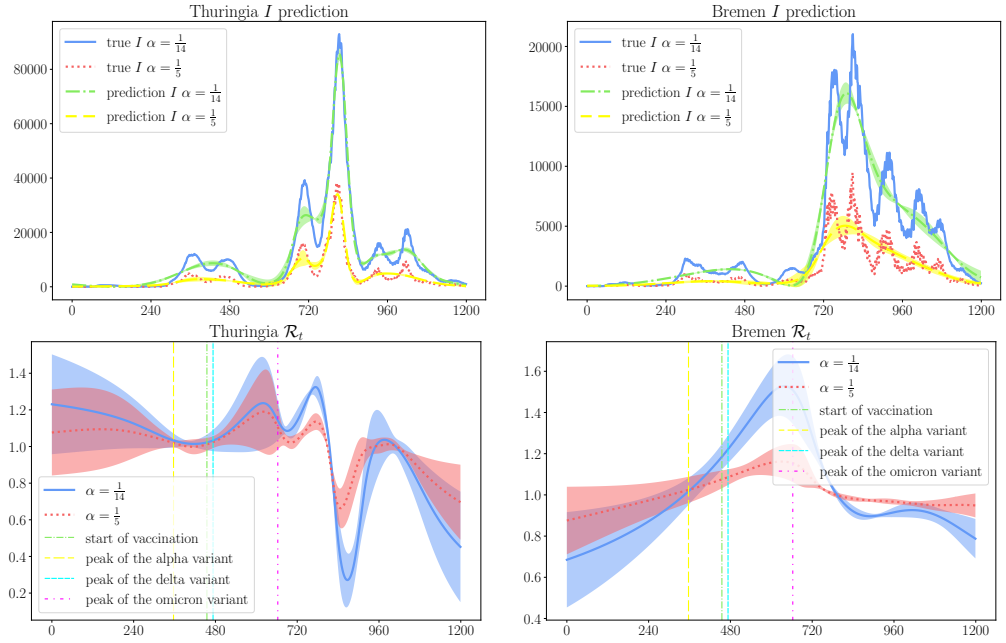


Figure 4.5: Visualization of the prediction of the training and the graphs of $\mathcal{R}_t$ for Thuringia (left) and Bremen (right) with both $\alpha = 1/14$ and $\alpha = 1/5$. Events [8] like the peak of an influential variant or the start of the vaccination of the public are marked horizontally. Further visualizations can be found in Chapter 6.

is shorter, but the peak value is higher.

Table 4.3 presents data regarding the discrepancy between the predicted and actual values from the dataset for compartment I. It is evident, that the error for all experiments falls within a range of values that is not negligible and will have an influence on the resulting reproduction values that are learned while fitting the data. A comparison of the results for the various values of α reveals that the errors associated with $\alpha = 1/14$ are consistently smaller, with the exception of Saxony and Germany. This can be attributed to the differing sizes of infection counts, particularly in relation to the normalization factor C. The model is unable to learn effectively if the values of the data loss $\mathcal{L}_{\text{data}}$ are too large or too small at the beginning.

As illustrated in Section 4.2.2, the training data is overlaid with the corresponding prediction of the model. We can observe that the prediction, though an exact reconstruction, accurately captures the general trajectory of the pandemic. The model's

Table 4.3: This table displays all average values of the error e_{I} for all German states and Germany. The average is formed across all 10 iteration.

	e_I		days with $\mathcal{R}_t > 1$		peak $\mathcal{R}_t$	
state name	$\alpha = \frac{1}{14}$	$\alpha = \frac{1}{5}$	$\alpha = \frac{1}{14}$	$\alpha = \frac{1}{5}$	$\alpha = \frac{1}{14}$	$\alpha = \frac{1}{5}$
Schleswig Holstein	0.228	0.258	467.5	458.5	1.475	1.166
Hamburg	0.265	0.330	424.3	409.8	1.500	1.297
Lower Saxony	0.224	0.340	413.1	430.3	1.662	1.223
Bremen	0.246	0.380	468.6	539.1	1.582	1.179
NRW	0.185	0.252	486.3	602.0	1.573	1.205
Hesse	0.302	0.346	553.0	511.2	1.409	1.157
Rhineland-Palatinate	0.256	0.277	484.7	404.7	1.534	1.175
Baden-Württemberg	0.198	0.284	469.2	590.0	1.457	1.180
Bavaria	0.225	0.318	490.5	486.1	1.428	1.199
Saarland	0.284	0.408	500.2	564.7	1.515	1.180
Berlin	0.201	0.240	591.9	514.4	1.721	1.262
Brandenburg	0.237	0.242	555.9	596.3	1.447	1.159
MV	0.170	0.257	537.5	544.3	1.563	1.135
Saxony	0.292	0.256	722.3	695.4	1.790	1.407
Saxony-Anhalt	0.213	0.268	572.0	631.9	1.387	1.165
Thuringia	0.180	0.222	732.1	730.6	1.586	1.249
Germany	0.284	0.239	587.7	430.7	1.561	1.219

prediction demonstrates an ability to capture larger peaks, exhibiting a tendency to ignore smaller changes. This suggests that the prediction of the model is capable show the rough outline of the progression of COVID-19. In the beginning, the majority of predictions below $\mathcal{R}_t = 1$, indicating an outbreak. As we observed in the synthetic data, the model exhibits a higher standard deviation at the boundaries. In the graphs, we mark the peaks of the most severe COVID-19 variants in Germany [8]. While the peaks of the Alpha and Delta variants are clearly visible in the data, the model does not learn these, and thus they are not reflected in the results. The peak of the Omicron variant represents the culmination of the COVID-19 pandemic in Germany and can be identified as the most prominent peak in the dataset. Immediately preceding this peak, we observe the highest value of the reproduction number across all states. This phenomenon can be explained, by number of individuals infected by one infectious person reaching its peak. In some states the peaks of other Omicron variants after the maximum peak are visible (see Thuringia).

The experiments demonstrate, that our model encounters difficulties in learning the data for the states and Germany and consequently in predicting the reproduction values for each dataset. Nonetheless, the predictions illustrate the general trends of the most impactful events of the COVID-19 pandemic.

Chapter 5

Conclusions

The objective of this thesis is to identify quantifying measures for the COVID-19 pandemic in Germany and its 16 federal states. We use the SIR model to describe the dynamics of the disease over time, offering an approximation of the reality. In this model, the transmission rate β and recovery rate α describe the infectiousness and resolution of the disease that the respective population experience. These rates serve as constant evaluation measures throughout the entire duration of the pandemic. The time-dependent reproduction number indicates the number of individuals infected by a single infectious individual. The SIR model is defined on a system of differential equations that elucidates the relations between these rates. In order to obtain these values for Germany, it is necessary to solve the ordinary differential equations (ODEs) for the data pertaining to the pandemic in each state and in Germany as a whole. We employ a physics-informed neural network in our approach to solve the ODE's. The data on which we train is collected by the Robert Koch Institute and made publicly available on GitHub, where they can be accessed for download. We preprocess the data to fit have the required format for the PINNs to reconstruct it, and at the same time predicts the transition rates and the reproduction number for the given data. Using this we conduct experiments on synthetic data and on the data for the German states and Germany itself. The results for the synthetic data demonstrate the efficacy of our data on small datasets.

The results of our work regarding the real-world data are divided into two groups. First we have the constant transmission rates, which provide insight into the overall trajectory of the pandemic in a given region. A high transmission rate indicates that, on average, the significant number of individuals were infected during the pandemic. Conversely, a high recovery rate indicates that individuals either recovered or died from the disease at a faster rate. Due to this contradiction in positive or negative meaning in α paired with the uncertainty of a possible dependency on β during train-

ing, we want to shift the focus on our results of β. The states with the highest transmission rate values are Thuringia, Saxony-Anhalt and Mecklenburg-Vorpommern. Furthermore, it is evident the six eastern states exhibit a higher transmission rate than the overall German rate (see Figure 4.2). These results align with the ongoing narrative of the COVID-19 pandemic in Germany, which has highlighted a perceived discrepancy in vaccination rates between the eastern and western federal states. This assertion which can be substantiated by a comparison of the vaccination ratios ν of each state and our findings. We find a strong negative correlation between ν and β. The results from our second experiments, underscore these findings. Here, we approximate the reproduction number $\mathcal{R}_t$ from the data. When $\mathcal{R}_t > 1$, the disease spreads rapidly through the population. Our results indicate a tendency for states with a high β to experience longer periods with $\mathcal{R}_t > 1$. Furthermore, we can identify the time point on which the most impactful events happened during the pandemic in Germany.

Although larger events are visible, smaller, less impactful events that are still visible on the raw data, do not appear in our results. This discrepancy can be attributed to the less precise reconstruction of the input data. The predicted version is smooth and does not contain any smaller peaks. To address these implementational limitations of our method, we intend to conduct comprehensive hyperparameter search to find the best configuration of our models to fit the data. Further optimizations can be applied to the epidemiological model that we employ, for which we present options in the subsequent section.

5.1 Further Work

Our findings demonstrate that with our methods enable the quantification of the course of the COVID-19 pandemic in Germany using the data provided by the Robert Koch Institute. Additionally, we present the limitations of our work. The SIR model is subject to numerous limitations. For instance, it does not account for individuals, who may be immune due to the vaccination status or those who are not infectious due to quarantine. In this section, we explore epidemiological models that illustrate these dynamics observed in real-world pandemics and recommend further investigation for Germany. First, we examine extensions of the SIR models, then we focus on agent-based models (ABMs).

5.1.1 Further Compartmental Models

As our results demonstrate, the SIR model is capable of approximating the dynamics of real-world pandemics. However, the model is not without limitations. As previously stated, the SIR model assumes that recovered individuals remain immune and does not account for the reduction of exposure of susceptible individuals through the introduction of non-pharmaceutical mitigation policies, such as social distancing policies. These shortcomings can be addressed by incorporating additional compartments and transmission rates into the model. For example, the SEIRD model incorporates an *Exposed* group and subdivides the *Removed* group into *Dead* and *Recovered* compartments. Furthermore, this adds four additional rates to the model: the contact rate, representing the average number of contacts between infectious and susceptible people with a high probability of infection; the manifestation index, indicating the proportion of individuals exposed to the disease who will become infectious; the incubation rate, measuring the time required for exposed individuals to become infectious; and the infection fatality rate, quantifying the fraction of individuals who succumb to the disease. As Doerre and Doblhammer [5] show for Germany using a numerical approximation method, for an SIERD model that they specialize to be age- and gender-specific, that it shows the impact of non-pharmaceutical mitigation policies. In their work, Cooke and van den Driessche [3] propose the SEIRS model with two delays. This is model is capable of approximating diseases, that have an immune period, after which the recovered individual becomes susceptible again. These are just a few examples of the numerous modifications of the basic SIR model that can be used to approximate and consequently quantify a pandemic.

5.1.2 Agent based models

While compartmental models, such as the SIR model, look at the population as a divided group, with each group representing a specific characterization that all inhabitants of that group share, an *Agent-Based Model* (ABM) sets its focus on the individual. Each individual, or agent, has specific attributes that determine its behavior and interactions with other agents during the simulation. As Gilbert [9] states, ABMs simulate the behavior of large groups, with each individual following simple rules. Kerr *et al.* [13] put forth a simulation tool, *Covasim*, which they base on an ABM. The ABM employs local data, including demographic data, disease incidence data from the region, and contact data for household, schools and workplaces, to define its simulation for a specific region. In their work, Maziarz and Zach [17]

address the criticism levied against ABMs for simplifying the dynamics and lacking the empirical support for the assumptions it they make. The authors utilize an ABM and the data specific to Australia to demonstrate the efficacy of ABMs in portraying the dynamics of the COVID-19 pandemic. They further state that ABMs can serve as serve as a tool for assessing the impact of non-pharmaceutical mitigation policies. This illustrates that ABMs play a distinct role in analyzing the COVID-19 pandemic. As the data situation has evolved, it is imperative to investigate the potential of utilizing ABMs as a tool to assess the pandemic's course.

Chapter 6

Appendix

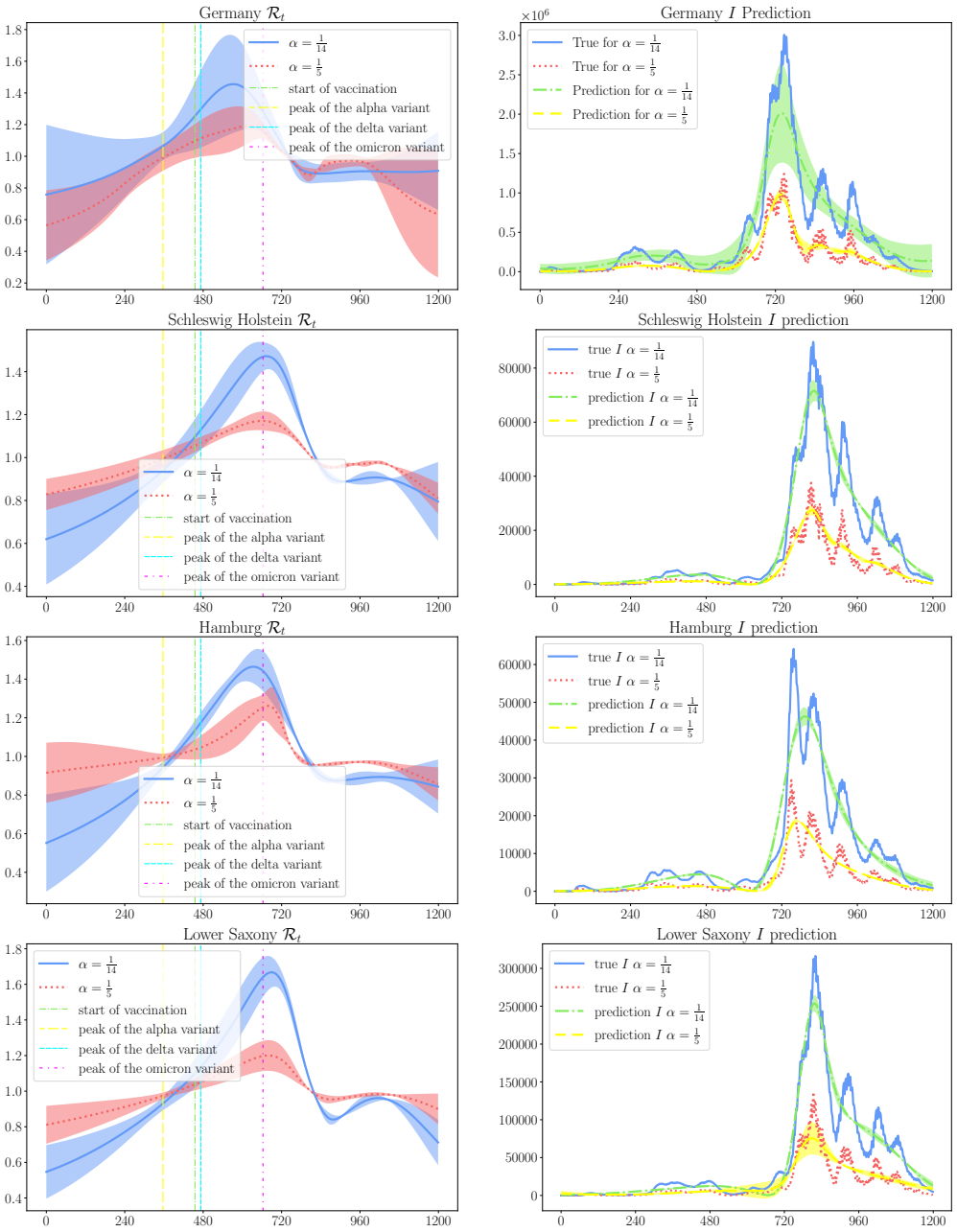


Figure 6.1: Part 1 of the results

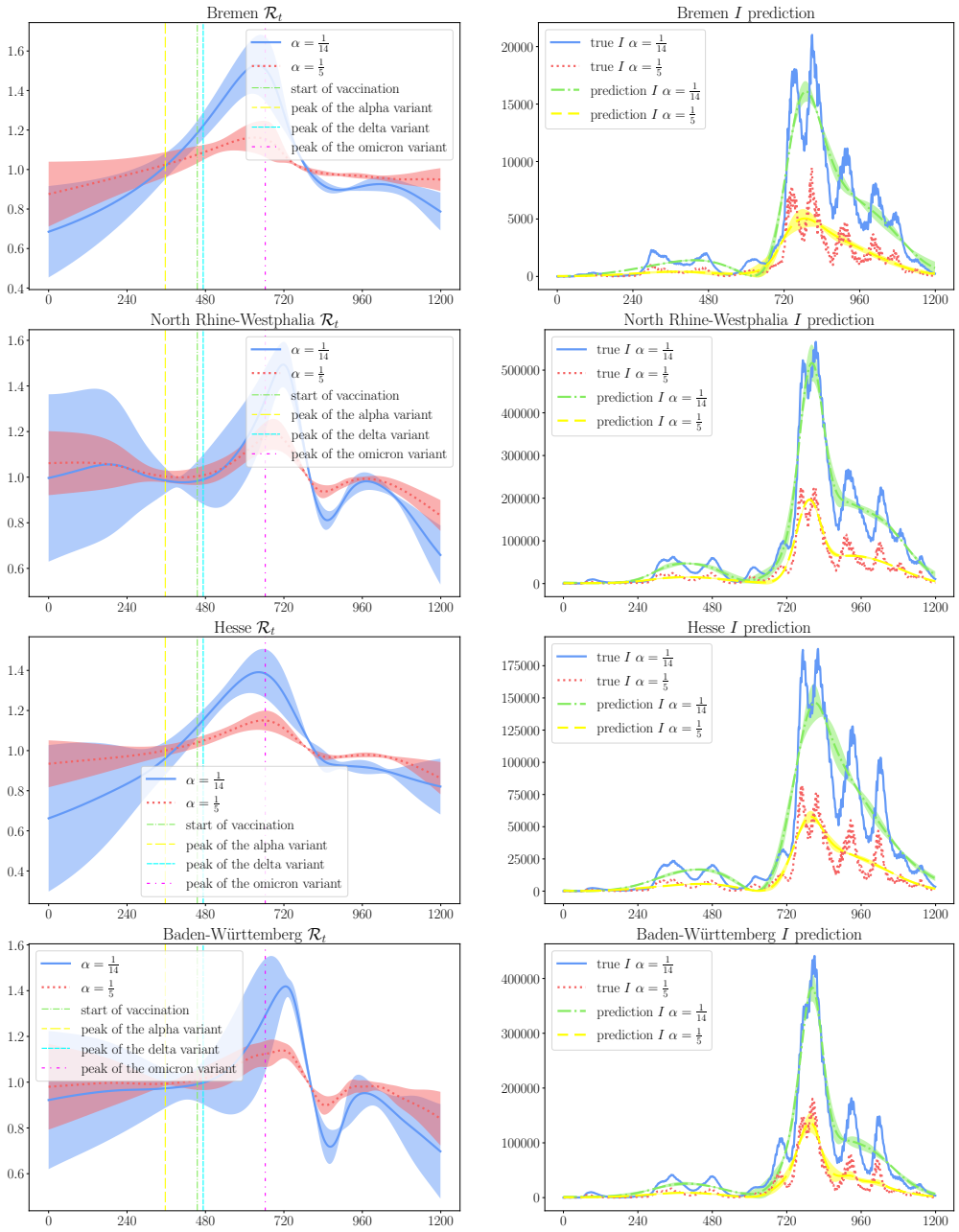


Figure 6.2: Part 2 of the results

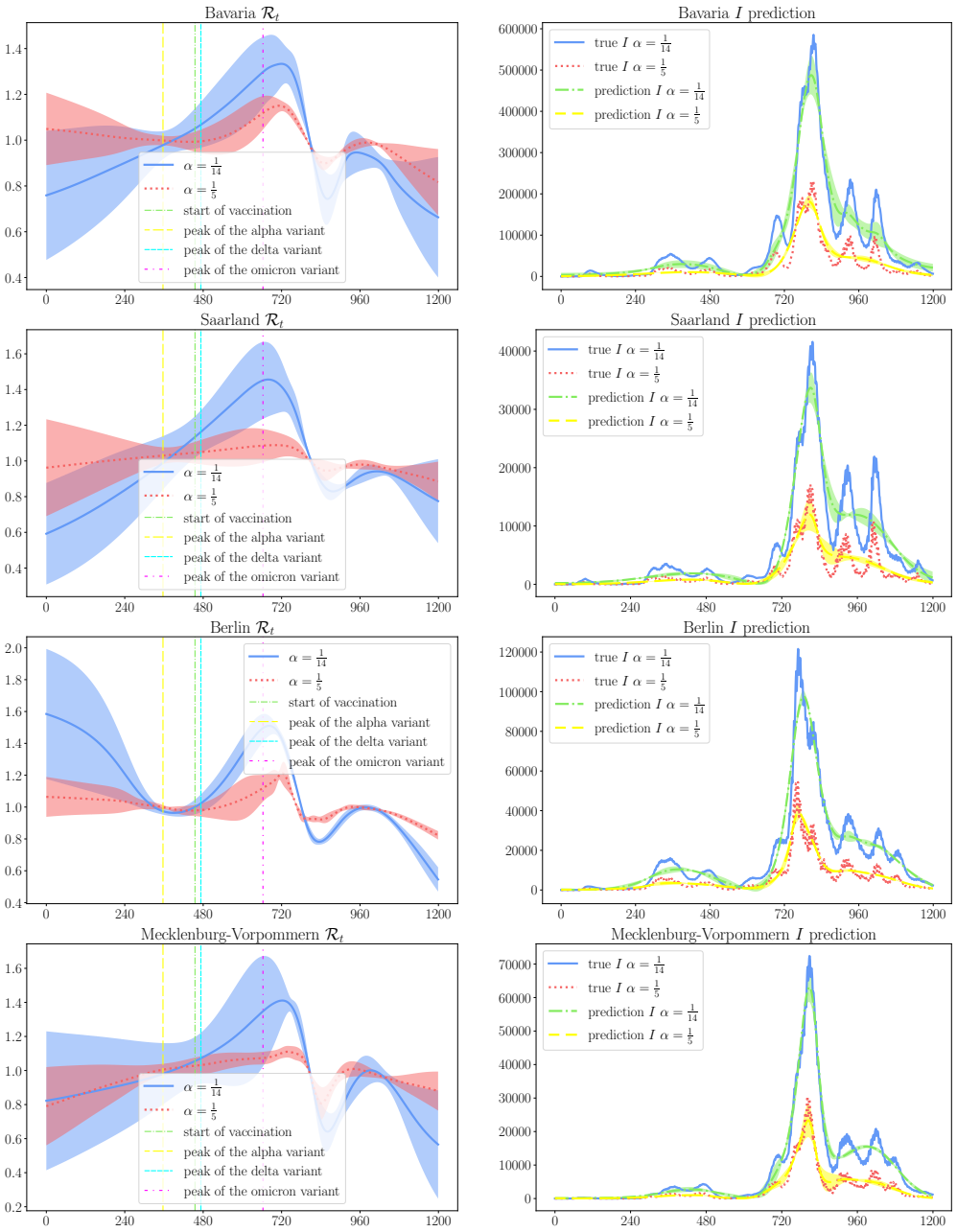


Figure 6.3: Part 3 of the results

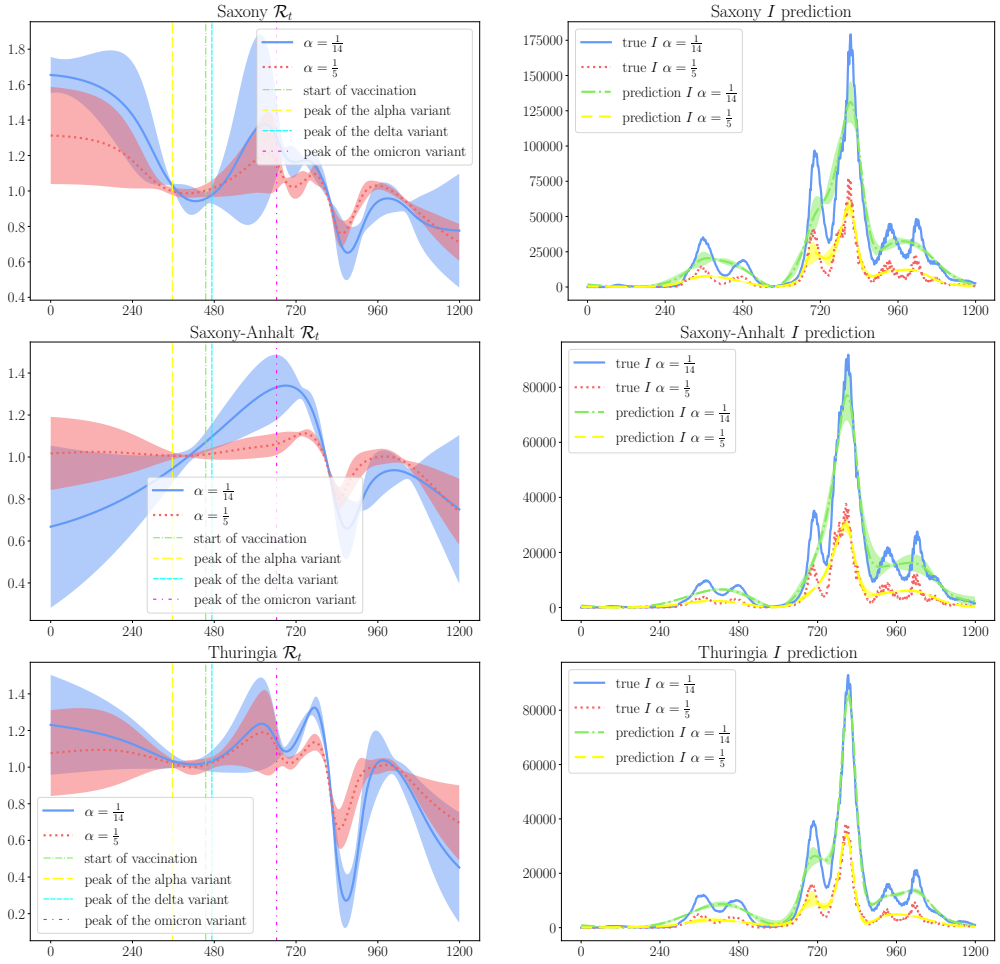


Figure 6.4: Part 4 of the results

Bibliography

[1] R. M. Anderson, Roy Malcolm; May. *Infectious diseases of humans : dynamics and control*. Oxford University Press, 1991.

[2] S. Berkhahn and M. Ehrhardt. A physics-informed neural network to model covid-19 infection and hospitalization scenarios. *Advances in Continuous and Discrete Models*, 2022(1), Oct. 2022.

[3] K. L. Cooke and P. van den Driessche. Analysis of an seirs epidemic model with two delays. *Journal of Mathematical Biology*, 35(2):240–260, Dec. 1996.

[4] W. Demtröder. *Experimentalphysik 1*, volume 1 of *Lehrbuch*. Springer Spektrum, Berlin, 9. auflage edition, 2021. Auf dem Umschlag: Mit über 2,5 h Lösungsvideos zu ausgewählten Aufgaben.

[5] A. Doerre and G. Doblhammer. The influence of gender on covid-19 infections and mortality in germany: Insights from age- and gender-specific modeling of contact rates, infections, and deaths in the early phase of the pandemic. *PLOS ONE*, 17(5):e0268119, May 2022.

[6] L. Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, 2005.

[7] Federal Centre for Health Education. Ansteckung, Übertragung und krankheitsverlauf. `https://www.infektionsschutz.de/coronavirus/fragen-und-antworten/ansteckung-uebertragung-und-krankheitsverlauf/`. Accessed: 2024-09-05.

[8] Federal Ministry of Health. Coronavirus-pandemie: Was geschah wann? `https://www.bundesgesundheitsministerium.de/coronavirus/chronik-coronavirus.html`. Accessed: 2024-09-05.

[9] G. N. Gilbert. *Agent-based models*. Number 153 in Quantitative applications in the social sciences. Sage Publ., Los Angeles [u.a.], 3. pr. edition, 2010.

[10] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[11] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, Jan. 1989.

[12] 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*, 115(772):700–721, Aug. 1927.

[13] C. C. Kerr, R. M. Stuart, D. Mistry, R. G. Abeysuriya, K. Rosenfeld, G. R. Hart, R. C. Núñez, J. A. Cohen, P. Selvaraj, B. Hagedorn, L. George, M. Jastrzębski, A. S. Izzo, G. Fowler, A. Palmer, D. Delport, N. Scott, S. L. Kelly, C. S. Bennette, B. G. Wagner, S. T. Chang, A. P. Oron, E. A. Wenger, J. Panovska-Griffiths, M. Famulare, and D. J. Klein. Covasim: An agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149, July 2021.

[14] S. Kirchhoff. Ueber den durchgang eines elektrischen stromes durch eine ebene, insbesondere durch eine kreisförmige. *Annalen der Physik*, 140(4):497–514, Jan. 1845.

[15] I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. 1997.

[16] X. Liu and P. Stechlinski. Infectious disease models with time-varying parameters and general nonlinear incidence rate. *Applied Mathematical Modelling*, 36(5):1974–1994, May 2012.

[17] M. Maziarz and M. Zach. Agent-based modelling for sars-cov-2 epidemic prediction and intervention assessment: A methodological appraisal. *Journal of Evaluation in Clinical Practice*, 26(5):1352–1360, Aug. 2020.

[18] C. Millevoi, D. Pasetto, and M. Ferronato. A physics-informed neural network approach for compartmental epidemiological models. 2023.

[19] M. Minsky and S. A. Papert. *Perceptrons*. The MIT Press, Cambridge/Mass. [u.a.], 2. print. with corr edition, 1972. Literaturangaben.

[20] B. Oksendal. *Stochastic Differential Equations*. Universitext Ser. Springer Berlin / Heidelberg, Berlin, Heidelberg, 5th ed. edition, 2000. Description based on publisher supplied metadata and other sources.

[21] K. D. Olumoyin, A. Q. M. Khaliq, and K. M. Furati. Data-driven deep-learning algorithm for asymptomatic covid-19 model with varying mitigation measures and transmission rate. *Epidemiologia*, 2(4):471–489, Sept. 2021.

[22] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[23] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, 2017.

[24] RKI. Covid-19-strategiepapiere und nationaler pandemieplan. `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/ZS/Pandemieplan_Strategien.html`. Accessed: 2024-09-06.

[25] RKI. Github covid-19-todesfälle in deutschland. `https://github.com/robert-koch-institut/COVID-19-Todesfaelle_in_Deutschland`. Accessed: 2024-09-05.

[26] RKI. Github sars-cov-2 infektionen in deutschland. `https://github.com/robert-koch-institut/SARS-CoV-2-Infektionen_in_Deutschland`. Accessed: 2024-09-05.

[27] RKI. Sars-cov-2: Virologische basisdaten sowie virusvarianten im zeitraum von 2020 - 2022. `https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Virologische_Basisdaten.html?nn=13490888#doc14716546bodyText10`. Accessed: 2024-09-05.

[28] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[29] W. Rudin. *Analysis*. Oldenbourg Wissenschaftsverlag GmbH, 2007.

[30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct. 1986.

[31] S. Setianto and D. Hidayat. Modeling the time-dependent transmission rate using gaussian pulses for analyzing the covid-19 outbreaks in the world. *Scientific Reports*, 13(1), Mar. 2023.

[32] S. Shaier, M. Raissi, and P. Seshaiyer. Data-driven approaches for predicting spread of infectious diseases through dinns: Disease informed neural networks, 2021.

[33] A. Smirnova, L. deCamp, and G. Chowell. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the seir model. *Bulletin of Mathematical Biology*, 81(11):4343–4365, May 2017.

[34] Statista Research Department. Anzahl infektionen und todesfälle in zusammenhang mit dem coronavirus (covid-19) in deutschland seit februar 2020. https://de.statista.com/statistik/daten/studie/1102667/umfrage/erkrankungs-und-todesfaelle-aufgrund-des-coronavirus-in-deutschland/. Accessed: 2024-09-06.

[35] M. Tenenbaum and H. Pollard. *Ordinary Differential Equations*. Harper and Row, Publishers, Inc., 1985.

[36] WHO. Coronavirus disease (covid-19). `https://www.who.int/health-topics/coronavirus#tab=tab_1`. Accessed: 2024-09-06.

List of Figures

List of Tables