AMA4951 Capstone Project

Topic: Mechanistic Model for Infecticous Diseases

Title: Unexpected positive correlation between human development index and risk of infections and deaths of COVID-19 in Italy

Kai Liu1, Mu He2, Zian Zhuang3, Daihai He3 and Huaichen Li4,5

1 Department of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance.

2 Department of Mathematics and Statistics, McMaster University, Canada.

3 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR China

4 Department of Respiratory and Critical Care Medicine, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, 250021, China. lihuaichen@163.com.

5 School of Medicine and Life Sciences, University of Jinan-Shandong Academy of Medical Sciences, Jinan, 250062, Shandong, China. lihuaichen@163.com.

Email of authors

Kai Liu [liuk0809@lixin.edu.cn](mailto:liuk0809@lixin.edu.cn)

Mu He [hem11@math.mcmaster.ca](mailto:hem11@math.mcmaster.ca)

Zian Zhuang [larry.zhuang@polyu.edu.hk](mailto:larry.zhuang@polyu.edu.hk)

Daihai He [daihai.he@polyu.edu.hk](mailto:daihai.he@polyu.edu.hk)

huaichen Li lihuaichen@163.com

Correspondence to

Daihai He, Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong SAR China, [daihai.he@polyu.edu.hk](mailto:daihai.he@polyu.edu.hk)

# ABSTRACT

The Covid-19 virus is a global pandemic virus that is caused by the virus called severe acute respiratory syndrome coronavirus 2 or (SARS-CoV-2). The virus was first reported in the Wuhan District of China and since then the various has spread to almost every corner of the globe due its unique mode of transmission. It is a known fact that flue like diseases spread like wildfire.

Over the centuries that mankind has walked the earth several pandemics have threatened his existence such as the Prehistoric Epidemic of Circa (300 B.C), the Spanish Influenza and even HIV and AIDs disease pandemic. But none have raged chaos such as the Covid-19 pandemic that has devastated and looted the world of its peace from and to all ends. Peoples livelihoods were disrupted, many were infected and died and even some people were deported back to their home countries.

Much of research was conducted into the causes and effects of the disease. An intrigue was taken into the dynamic causal of the disease as many were unsure as to whether the disease propagated itself or mostly the propagation was at relationship with other factors such as the human development index.

The aim of this research paper is to study the correlation between human development index and the risks of infections and death in Italy through a mechanistic model that is grounded on raw data that serves to prove the significance of the assumed hypothesis.

# 1.CHAPTER 1

## Introduction

In this analysis, I am going to observe that how human development index, including age, education and life expectation, correlates with number of cases and death rate among the population of COVID-19, especially in Hong Kong. After the analysis, it has shown that HDI is positively correlated with annual income and chronical ill, and is negatively correlated with smoking habit. I will explain why these factors can be observed between human development index and infection risks and mortality of COVID-19 in Hong Kong.

## Historical Background

The dynamic of infectious disease epidemics is divided into individuals with differing disease status. Starting from December 2019, there was a new virus called SARS-CoV-2. It first appeared in China and expanded globally quickly. Many patients infected with COVID-19 have suffered from high fever, a cough and being weird on sense of smell, which were classified as mild symptoms, while others might not have any symptoms. However, some would evolve into more severe and fatal symptoms affecting lungs, heart and brain.

COVID-19 used ACE2, which is a human enzyme, to submerge the cells of its host. ACE2 decreases blood pressure in human body and has opposite function with another enzyme called ACE. Thus, human body has to balance ACE and ACE2 level to ensure a normal standard of blood pressure. And it is not sure how COVID-19 affects ACE and ACE2 work.

When COVID-19 first discovered in China, their researchers collected cells and fluid from sufferers’ lungs, in order to identify the virus. The researchers analyzed the data collected former work to study if changes in body regulation of blood pressure would cause severe symptoms of COVID-19. They found that the ACE level would decrease and ACE2 level in lung cells would increase due to COVID-19. This will lead to rise in molecule level called bradykinin in cells, which will prompt pain and cause expansion of blood vessels while this will cause swelling and agnail of surrounding cell. Moreover, it was found that it would produce a substance called hyaluronic acid and the enzymes that could demote it greatly decreased. Hyaluronic acid can form a hydrogel from the absorption of more than 1,000 times its own weight in water. This will lead to fluid leakage into the lungs combined with the excess hyaluronic acid. This would result in the creation of Jelly-like substance to prevent COVID-19 patients suffering from severe symptoms from oxygen uptake and carbon dioxide release in the lungs. Therefore, the findings suggested that this will lead to more severe symptoms of COVID-19.

The COVID-19 has exploded quickly in Hong Kong since January 2020 while the pestilence in China was then controlled. Obvious differences in contagious diffusion and fatality rate due to COVID-19 across countries appear, however, factors and reasons about the differences are not clear, and is meaningful to study the relationship between the infectious disease and human development level and other factors.

It is assumed that people are independent in the same country or region, with same probability of infection and confirmation. A multiple logistic regression model was performed between human development index and confirmed cases. It is found that HDI is significantly large, p-value is less than 2 \* 10-16. If the HDI rose by 1, the number of a confirmed case would increase by exp(2.8648).

Death rate is defined as the scale of death among confirmed cases. Not all the patients of COVID-19 are counted into and diagnosed as confirmed cases. It is assumed the people in the same country or region have same probability to suffer from infection or death because of COVID-19 but the death risk is different in different regions. The regression model is to observe how the performance of relation between death and HDI is. And it is found that a rise in 0.1 HDI will increase 39 deaths.

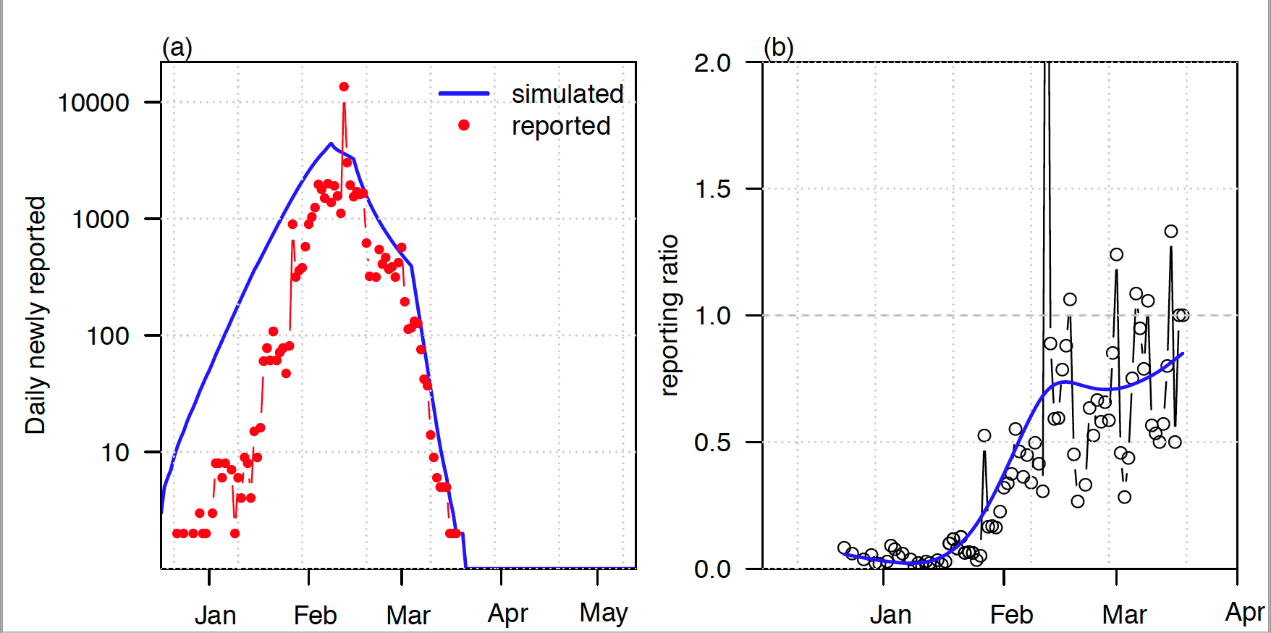
It is found that the higher the human development index in a region, the higher the COVID-19 case rate and death rate. HDI is a statistic index of life expectancy, education, living standard and gross income per capita.

Number of persons suffering from at least 1 chronic diseases per 50 persons and number of cigarettes per 50 persons with smokers aged at least 10 years are used to calculate life expectancy, since they may have health issues. And the median of life expectancy is 41.The gross income of a region is to measure a region’s living standard. And it median is USD27962. As a result, it shows that the higher the HDI in a region, the higher case rate and death rate of COVID-19.

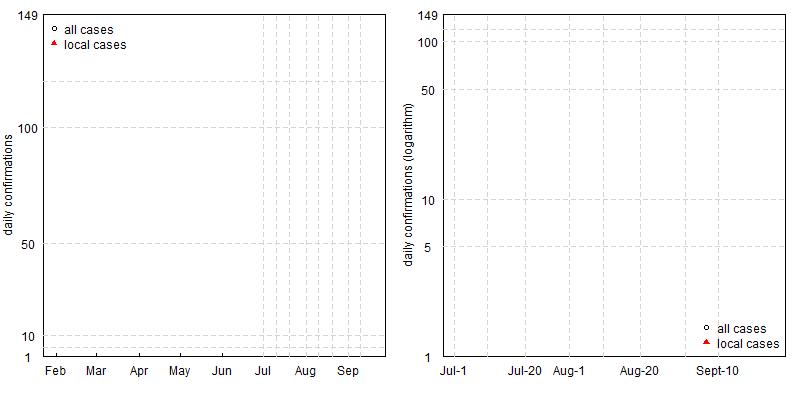
Simple regression model is to measure the effect HDI on case rate and fatal rate of COVID-19 directly. The results can be generated by standard errors.

Multiple regression model is to study the further effect of HDI on case rate and fatal rate after other factors’ adjustments. While more people consume cigarettes and suffer from chronic diseases, they will positively be correlated with case rate and fatal rate. And these factors are more significant than measuring gross income of a region. When HDI increases by 0.1, the number of confirmed cases will increase by 415.72 and the death rate will increase by more than 10000.

In conclusion, the higher the HDI means the case rate and death rate of COVID-19 are higher. Although higher HDI means improvement of living standards and better life expectancy with high gross income, it is found that there will be more persons suffering from more than 1 chronic diseases and smokers in a region.



Trend Graphs



# CHAPTER 2 THEORETICAL BACKGROUND

## 2. Related work

Mathematical modeling concedes rapid computation and estimation of pandemic outbreaks and plays a valuable role in decision making. Simulation techniques, on the other hand, are used when the data collection involves a large number of conditions to test, which leads to increased cost ([Siettos & Russo, 2013](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib52)). Several mathematical and statistical models have been derived recently such as the Multivariate linear regression ([Thomson et al., 2006](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib54)), time series models ([Kurbalija et al., 2014](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib31)), grey forecasting models ([Wang et al., 2018a](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib58); [Zhang et al., 2017](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib66)), back-propagation neural networks ([Liu et al., 2019](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib33); [Ren et al., 2013](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib45); [Zhang et al., 2013](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib65)), and simulation models ([Nsoesie et al., 2013](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib37); [Orbann et al., 2017](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib38)). The spread of an epidemic is unpredictable and random. Due to this reason, it becomes difficult to build mathematical models to analyze epidemic randomness. [Hethcote (1989](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib24), pp. 119–144) identifies three basic types of deterministic models – SIS endemic, SIR epidemic, SIR endemic for the mathematical modeling, and predicting the spread of infectious disease. Theorems that consist of “reproduction number R0, contact number σ and replacement number R″ are presented for mathematical models like SEIR and MSEIRS ([Hethcote, 1989](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib24), pp. 119–144). Compared with statistics methods, mathematical modeling based on dynamical equations receive relatively less attention ([Adam et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib1)), though they can provide a more detailed mechanism for the epidemic dynamics.

The classical susceptible exposed infectious recovered model (SEIR) is one of the most widely adopted methods for characterizing the epidemic of COVID-19 outbreak in both China and other countries ([Hethcote, 1989](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib24), pp. 119–144). The SEIR model replicates the “time-history” of any epidemic or pandemic outbreak, and it presents the model of dynamic interaction between people with four different health conditions or phases of the pandemic, namely the susceptible (S), exposed (E), infective (I), and recovered (R). SEIRD model contains the 4 basic containers in the SEIR model: **S**usceptible, **E**xposed, **I**nfective, **R**ecovered, along with an added container - **D**ead. A “Formal Characterization and Model Comparison Validation” based on the SEIRD model which uses the data from Korea and Spain is proposed by Casas et al. ([Fonseca i Casas et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib15)). The proposed model showed the predicted parameterization with empirical evidence and a decision support system (DSS) is implemented to study the nature of the pandemic in Catalonia ([Fonseca i Casas et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032#bib15)).

A data-driven model to predict the spread of Covid-19 for an upcoming week using the SEIRD model is studied and tested for datasets obtained from Italy, India, and Russia ([Rapolu et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib44)). The proposed model ([Rapolu et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib44)) produces results in which the parameters are calculated from the data, to plan for the future requirement of PPEs for hospital staff and healthcare devices. Contrarily, the transmission dynamics of Covid-19 were evaluated based on a SEIRD compartmental modeling approach by Mukaddes et al. ([Mukaddes et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib35)). This model was based on the “kinematic parameters” that describe the transmission, recovery, and death rate in Bangladesh. The study also highlights the dynamic factors and two parameters that refer to infection which derives the reproduction number R0. This study is groundbreaking for research work carried out in developing nations to reopen the businesses and to boost the economy back to the pre-covid period. However, external influences such as weather, herd immunity were not considered as a part of the study ([Mukaddes et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib35)).

The COVID-19 pandemic is very dynamic and spreads rapidly, and hence there is a need to create robust modeling solutions to curb the outbreak. A forced SEIRD model with two different infection rate functions of the Covid-19 spread in Italy was investigated by Piccolomini et al. ([LoliPiccolomini & Zama, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib34)), in which the “integration time was distributed into sub-intervals” to estimate the model parameters. The study was based on data collected from two regions in Italy, Lombardia and Emilia-Romagna. This model ([LoliPiccolomini & Zama, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib34)) will be efficient to make predictions about different stages of the epidemic outbreak across various regions in Italy and Europe. Another popular and widely used statistical method for time-series forecasting is the **A**utomatic **R**egressive **I**ntegrated **M**oving **A**verage (ARIMA) model, which studies the series of temporal structures in time series data. Earlier study on disease management techniques with time series using ARIMA models is proposed by [Sato (2013)](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib47). The author [Sato (2013)](https://www.sciencedirect.com/science/article/pii/S2468042720301032#bib47) emphasizes the fact that the options to follow up and spot the difference in data patterns should be given importance in healthcare practices.

Forecasting a disease is essential for the healthcare department and policymakers to strengthen their vigilance and reallocate their resources. ARIMA time series model is a widely accepted method for the pandemic forecasting because of its simplicity and systematic structure ([Ceylan, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib8); [Wang et al., 2018b](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib59)). The extent of the outbreak of the Covid-19 pandemic in Italy, Spain, and France was examined with the ARIMA model ([Ceylan, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib8)). The proposed model consists of 4 steps in modeling which include “assessment, prediction of parameters, characteristic checking, and forecasting”. The outcome of the study can guide policymakers and healthcare authorities in the European nations to effectively allocate resources and plan for the future flare-up of the current situation ([Ceylan, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib8)).

Alzahrani et al. ([Alzahrani et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib3)) studied the spread of pandemic outbreak using the ARIMA prediction model in Saudi Arabia. The authors have used the “linear parametric model prediction approach”, in which the parameters of the ARIMA model were chosen based on the value of the “Akaike information criterion” ([Akaike, 1974](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib2); [Alzahrani et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib3)). The dataset was divided into training and testing datasets and four statistical models were employed to predict the spread of Covid-19, by comparing the performance of each model with the evaluation metrics from which good fit is derived ([Alzahrani et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib3)). Other related studies and research work on disease prediction using ARIMA and hybrid ARIMA modeling techniques found in the literature are presented in [Table 1](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "tbl1).

Table 1. Research work found in the literature for disease prediction using ARIMA/hybrid models.

| **Method (s)/Type of Modeling** | **Pandemic/Epidemic/Endemic** | **The research found in the literature** |
| --- | --- | --- |
| ARIMA | Malaria | Gaudart et al., ([Gaudart et al., 2009](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib16)) |
| ARIMA, Artificial Neural Networks (ANN) | HAV | Guan et al., ([Guan et al., 2004](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib20)) |
| ARIMA | SARS | Earnest et al., ([Earnest et al., 2005](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib13)) |
| ARIMA, Seasonal Autoregressive Integrated Moving Average (SARIMA) | Influenza | He et al., ([He & Tao, 2018](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib23)); Chen et al., ([Chen et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib9)) |
| Multivariate Poisson Regression (MPR), ARIMA, and ANN | Dengue Fever | Polwiang ([Polwiang, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib41)) |
| Random Forest (RF), ARIMA/X Models | Infectious Diarrhea | Fang et al., ([Fang et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib14)) |
| Elman Recurrent Neural Networks (ERNN), ARIMA, and Jordan Neural Networks (JNN) | Brucellosis | Wu et al., ([Wu et al., 2019](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib60)) |
| ARIMA (or) SARIMA-NAR (Nonlinear Autoregressive Network) (or) hybrid model | Covid −19 | Ceylan ([Ceylan, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib8)); Alzahrani et al. ([Alzahrani et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib3)); Perone ([Perone, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib40)); Kumar et al. ([Kumar et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib30)); Sato ([Sato, 2013](https://www.sciencedirect.com/science/article/pii/S2468042720301032#bib47)); Wang et al. ([Wang et al., 2018b](https://www.sciencedirect.com/science/article/pii/S2468042720301032#bib59)); Benvenuto et al. ([Benvenuto et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib6)); Hernandez-Matamoros et al. ([Hernandez-Matamoros et al., 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib22)); Kufel ([Kufel, 2020](https://www.sciencedirect.com/science/article/pii/S2468042720301032" \l "bib29)) |

This paper aims to build a hybrid model that will allow us to estimate the short and long-term dynamics of COVID disease and build confidence intervals of predictions by using a hybrid dynamic model based on SEIRD with ARIMA corrections. This model will help officials to be prepared for the waves of pandemic and reserve hospital beds in advance.

# Stochastic SEIRD model

The model is SEIRD model, where infected individuals can survive or die at different rates, with waning immunity. The full model specification is:

* SS: susceptibles
* EE: exposed, i.e. infected but not yet contagious
* IRIR: infectious who will survive
* IDID: infectious who will die
* RR: recovered
* DD: dead

There are no birth of natural death processes in this model. Parameters are:

* ββ: rate of infection
* δδ: rate at which symptoms appear (i.e inverse of mean incubation period)
* γRγR: recovery rate
* γDγD: death rate
* μμ: case fatality ratio (proportion of cases who die)
* ϵϵ: import rate of infected individuals (applies to EE and II)
* ωω: rate waning immunity

The model will be written as:

St+1=St−βSt (IR,t+ID,t)Nt+ωRtSt+1=St−βSt(IR,t+ID,t)Nt+ωRt

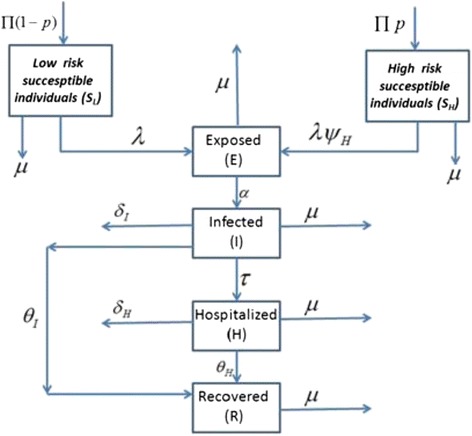
Et+1=Et+βSt(IR,t+ID,t)Nt−δEt+ϵEt+1=Et+βSt(IR,t+ID,t)Nt−δEt+ϵ

IR,t+1=IR,t+δ(1−μ)Et−γRIR,t+ϵIR,t+1=IR,t+δ(1−μ)Et−γRIR,t+ϵ

ID,t+1=ID,t+δμEt−γDID,t+ϵID,t+1=ID,t+δμEt−γDID,t+ϵ

Rt+1=Rt+γRIR,t−ωRtRt+1=Rt+γRIR,t−ωRt

Dt+1=Dt+γDID,t



* [Khan et al. (2015)](https://dx.doi.org/10.1186%2Fs40249-015-0043-3)

4 Time-Dependent SEIRD Model

In order to optimize the model, different approaches to the SEIR model are considered. This leads to the implementation of a time-dependent SEIRD model [4]. This approach resulted in accurate forecasts of Infected, Recovered, and Deceased rates for a week. Run time is exceptionally low, one minute at the most. This model can be represented as follows:

*β*(*t*)*S*(*t*)*I*(*t*)

*∆S* = − (1)

*N*

*β*(*t*)*S*(*t*)*I*(*t*)

*∆E* = − *α*(*t*)*E*(*t*) (2)

*N*

*∆I* =*α*(*t*)*E*(*t*) − *γ*(*t*)*I*(*t*) − *δ*(*t*)*I*(*t*) (3)

*∆R* =*γ*(*t*)*I*(*t*) (4)

*∆D* =*δ*(*t*)*I*(*t*) (5)

From (4), we have

*∆R*

*γ*(*t*)= (6)

*I*(*t*)

From (5), we have

*∆D*

*δ*(*t*)= (7)

*I*(*t*)

Using (6) and (7) in (3) yields

*∆I* +*∆R*+*∆D*

*α*(*t*)= (8)

*E*(*t*)

Using (8) in (2) yields

(*∆E* +*∆I* +*∆R*+*∆D*)*N*

*β*(*t*)= (9)

*S*(*t*)*I*(*t*)

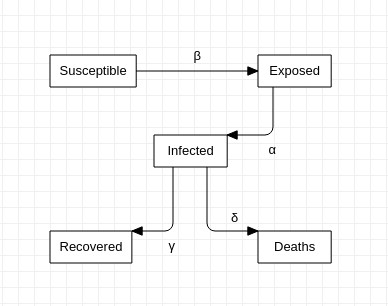


Fig. 3: Fig.3 SEIRD Model

# CHAPTER 3 RESEARCH METHODLOGY

here.

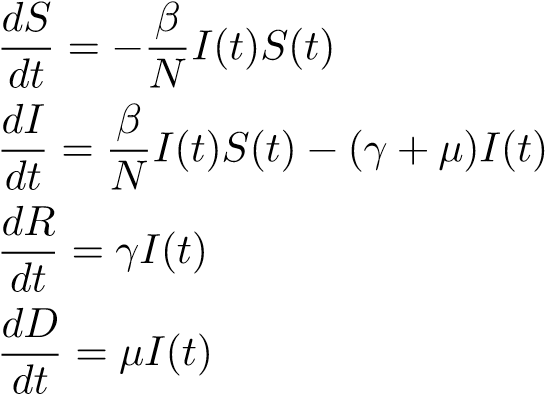
# Theory

Mathematical models for disease epidemic are either deterministic or stochastic [17], where the first may be considered some sort of thermodynamic limit of the second. An analogy made with thermodynamics, where given a big enough number of particles in a gas, for example, you find deterministic equations to describe the behavior of the gas given by the laws of thermodynamics without the need to know the exact behavior of each particle. Otherwise, when your number of molecules is low or you try to compute too many interactions between particles of your system, the random behavior and fluctuations start taking place and you get to a stochastic model.

We describe here a simple extensions of models constantly used on literature [18, 19, 10] and with it show some possible behaviors of a disease outbreak.

2.1 SIRD

A simple mathematical model for disease epidemic can been built dividing the population in 3 groups: susceptible individuals (S), infected individuals (I) and recovered individuals (R). The model composed by these groups is called the *SIR model*. In this article, however, we consider also individuals who have died by the disease, denote by D. Following the same arguments of the *SIR* model, the *SIRD* model can be described by the set of four differential equations:

(1)

(2)

(3)

(4)

Last equation is easily understood by thinking that the variation of the number of deaths may be proportional to the infected individuals, where the proportionality constant is denoted by *µ*. The constants *γ* and *β* are, respectively, the recovery rate and the number of infected, where *µ* and *γ* are given in terms of the infection fatality rate (IFR) or the case fatality rate (CFR); that is, the number of people who contracted the disease and died according to the total number of infections (IFR) or the registered number of infections (CFR), and the average time taken from symptoms onset to recovery, *τr*, or death *τd*, formally *µ* = *PCFR/τd* and *γ* = (1 − *PCFR*)*/τr*. The equations are simply a mathematical way to describe how individuals passes from one group to the other according to the following chain of events: A susceptible individual becomes infected by the virus, and from this point, it either dies or recovers (Figure 1).

Susceptible S(t)

Infected I(t)

Recovered R(t)

Dead D(t)

*βI*

(

*t*

)

*S*

(

*t*

)

*γI*

(

*t*

)

*µI*

(

*t*

)

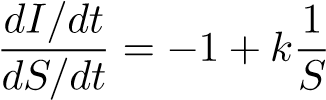
Figure 1: Representation of a SIRD model, a susceptible person gets infected and either dies or recovers from the disease.

Summing the four equations we get

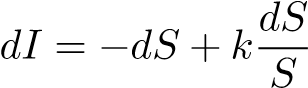
*S*(*t*) + *I*(*t*) + *R*(*t*) + *D*(*t*) = const*,* (5)

where the constant may represent the total number of individuals, *N*. Before proceeding, we propose the initial condition that, when *t* goes to zero, *I*(*t*) = *I*0, *R*(*t*) = *D*(*t*) = 0, and, therefore, *S*(*t*) = *S*0 = *N* − *I*0 ≈ *N*. Such an assumption is based on the fact that the entire population is susceptible to the SARS-CoV-2 virus.

Since *R*(*t*) and *D*(*t*) are both data updated day by day in Germany and Korea, it would be helpful to write *I*(*t*) as function of them so as to predict its behavior, obtaining, for example, the maximum number of infected individuals. For reasons that may be clear soon, we first write *I*(*t*) in terms of *S*(*t*). An intuitive step is to divide equation (2) by (1). Thus,

*,* (6)

where *k* = (*γ* + *µ*)*/β*. Eliminating the temporal dependence, we get a separable differential equation, that is,

*,* (7)

which the solution is easily verified to be

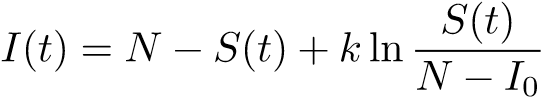
*I*(*t*) = −*S*(*t*) + *k* ln*S*(*t*) + const*.* (8)

Applying the initial condition, we obtain

*I*0 = −(*N* − *I*0) + *k* ln(*N* − *I*0) + const (9)

→ const = *N* − *k* ln(*N* − *I*0)*.*

Hence, equation (8) may be written as

*.* (10)

We can visualize here, that depending on the combination of *γ* and *µ*, *I* reaches 0 before the entire population *S* becomes infected (Figure 2).

0

2000

4000

6000

8000

10000

Susceptible

0

2000

4000

6000

8000

10000

I

n

f

e

c

t

e

d

*I*

*max*

*⇒*

*S*

=

*γ*

+

*µ*

*β*

*⇐*

*γ*

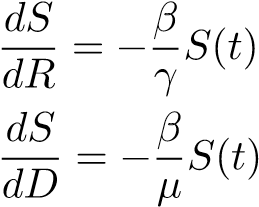
=

*µ*

= 0

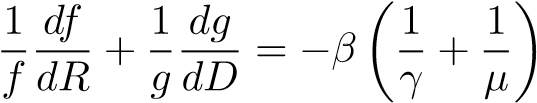
Figure 2: Plot of equation (10) with different conbinations of *γ* and *µ*.

Next, we may write *S*(*t*) in terms of *R*(*t*) and *D*(*t*). For this purpose, we begin by dividing equation (1) by (3) and (1) by (4),

 and (11)

*.* (12)

Adding these two equations and writing *S*(*R,D*) as *S*(*R,D*) = *f*(*R*)*g*(*D*), we get

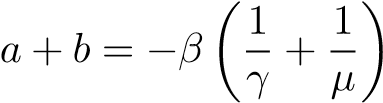
 (13)

|  |  |
| --- | --- |
| Since (13) is a separable equation, the well-known solution is given by |  |
| *f*(*R*) = *AeaR* and | (14) |
| *g*(*D*) = *BebD.*  Therefore, *S*(*t*) can be written as | (15) |
| *S*(*t*) = *CeaR*(*t*)+*bD*(*t*)*,* | (16) |

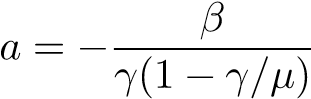
where we absorbed both constants *A* and *B* into *C*. By the initial condition, we find that *C* = *N* − *I*0. To find *a* and *b*, we must derive (16) in time under the condition that it may return to equation (1). In this way, we see that

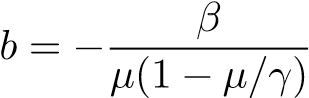
*aγ* + *bµ* = −*β.* (17)

By the other hand, substituting equations (14) and (15) in (13), we get

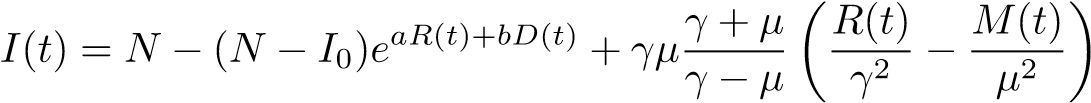
 *.* (18)

Solving this system,

 (19)

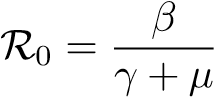
 (20)

Hence, *I*(*t*) can be finally written as

 (21)

With this equation, see that as *t* → ∞, *I* does not approaches *N* necessarily, depending on the recovery and death rates, *I* does not reach *N*.

The last important quantity extracted from this model is the basic reproduction number R0, given by [14]:

*.* (22)

This quantity, is of vital importance of the study of a disease outbreak.

2.2 SEIRD

Another deterministic mathematical model possible is the SEIRD model, in which we consider the population *N* of a given region as divided in 5 groups. At time *t*, there are those who are susceptible to get infected *S*(*t*), the ones who have already been exposed the virus but does not present symptoms yet *E*(*t*), people who are already infected and present the symptoms *I*(*t*), the ones that have already recovered from the disease *R*(*t*) and those who are dead due to the infection *D*(*t*). This model is a good approximation to a short epidemic, so the population of a region is roughly constant throughout the epidemic period. Also, since this is a deterministic model, we assume *N* to be a big number compared to the number of people associated with the infection of a single person. The final consideration is that we also assume that people that are recovered from the disease acquire immensity and does not become susceptible to become infected again.

The rate of infection *λ* is proportional to the number of people infected, *λ*(*t*) = *βI*(*t*), where the constant *β* represents the effectiveness of the infection, the rate of cure *γ* = *P*:)*τr*−1, where *P*:) is the probability of recovery and *τr* is the average time taken for an infected person to recover. Similarly the rate of death is , where *PCFR* = 1−*P*:) is the probability of death, given by the CFR and *τd* is the average time taken for an infected person to die. Figure 3 carries an visual representation of the SEIRD model.

Susceptible S(t)

Infected I(t)

Recovered R(t)

Dead D(t)

Exposed E(t)

*βI*

(

*t*

)

*S*

(

*t*

)+

*kE*

(

*t*

)

*S*

(

*t*

)

*cE*

(

*t*

)

*γI*

(

*t*

)

*µI*

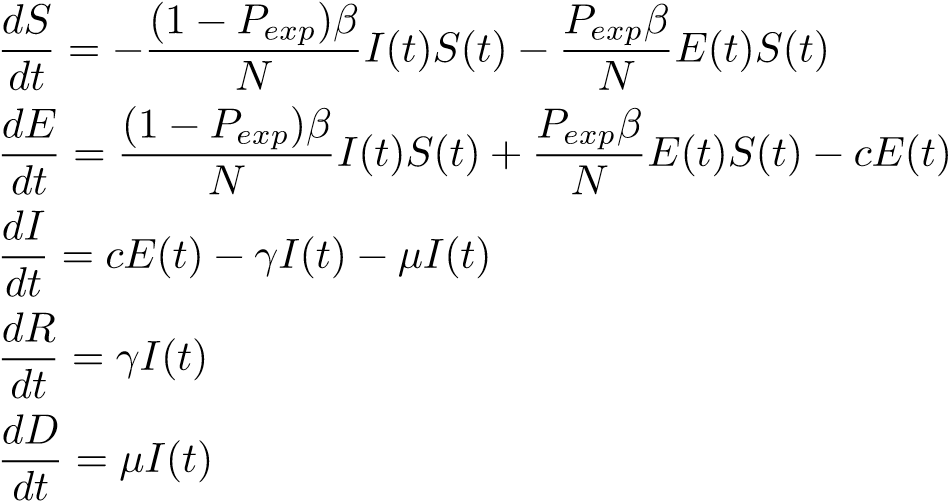
(

*t*

)

Figure 3: Representation of a SEIRD model, a susceptible person gets exposed to the virus, being infected afterwards and either dies or recovers from the disease.

The differential equations representing the evolution of the populations are given by

(23)

(24)

(25)

(26)

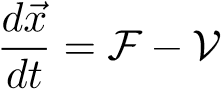
(27)

We first turn our attention to the construction of an appropriate formula for calculating R0 with this model. For that we follow the method derived on [20]. The study develops a mathematical generalization for writing R0 depending on the type of epidemiological model. R0 is defined as

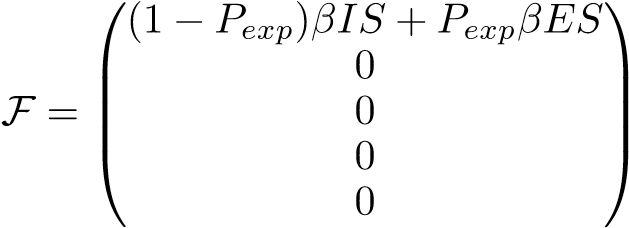
R0 = *ρ*(*FV* −1) (28)

where *ρ*(*X*) means the spectral radius of the matrix X, that is, the largest absolute eigenvalue. Both *F* and *V* are the matrices of the derivatives of the functions defining the behavior of the disease population, with respect to each population compartment.

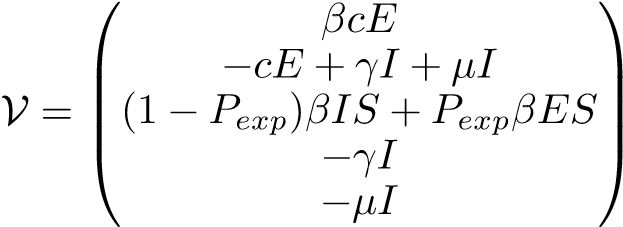
To get to these matrices, we first note that the set of equations regarding the dynamics of the SEIRD model can be expressed as follow: Consider *~x* the vector of populations, that is *~x* = (*x*1*,x*2*,x*3*,x*4*,x*5) where *x*1 = *E*, *x*2 = *I*, *x*3 = *S*, *x*4 = *R* and *x*5 = *D*. Analogously, *d~x/dt* is the vector of the first derivatives. Then, we can write the dynamics of the populations as

 (29)

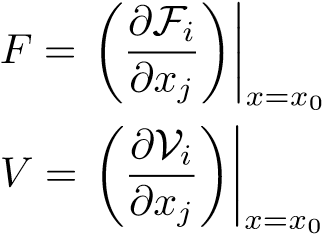
where F is the vector that relates the appearance of new infections on the disease populations due to contamination, and V is the input and output of members in all populations due to all other causes, such as recovery from the disease, development of symptoms after an incubation period, etc. In our case, since all newly infected members go to the *E* population

 (30)

while

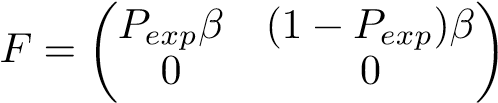
 *.* (31)

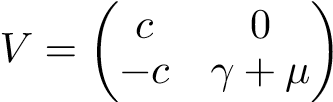
Now, we know that the situation of a disease free equilibrium (DFE), meaning no disease is happening, is achived by the vector *~x*0 = (0*,*0*,S*0*,*0*,*0), where *S*0 = *N*. According now to [20] we can calculate *F* and *V* as

 1 ≤ *i* (32)

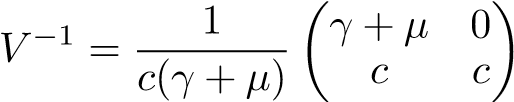
*j* ≤ *m* (33)

being *xj* the vector components of *~x* related to the populations with the disease, in our case *E* and *I*, and *m* is the number of populations related to infectious beings. Here *m* = 2. Performing the derivatives, we conclude

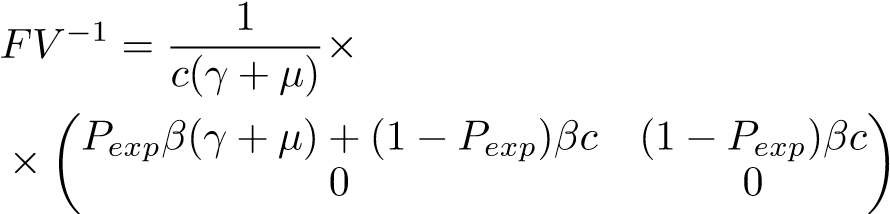
 (35)

 (36)

The next step is to find the inverse matrix of *V* , fortunately *V* is a 2x2 matrix and the formula for it’s inverse is straightforward

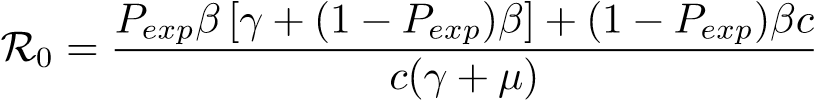
 *,* (37)

and we proceed to the last step of combining *FV* −1 in order to retrieve *ρ*(*FV* −1) and find R0.

(38)

*,*

therefore, by computing the eingenvalues of *FV* −1 we find

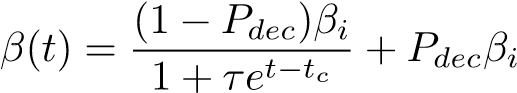
 (39)

Having R0 in our hands, we continue to the study of some behaviors of this model.

The set of equations describing the model is subjected to the initial conditions. When *t* → 0, *I*(*t*)− → *I*0, *S*(*t*) → *S* = *N* − *I*0, *R*(*t*) → 0, *D*(*t*) → 0 and *E*(*t*) → *E*0, where *I*0 is the initial number of infected, *E*0 is the initial number of exposed in the population and no deaths or recoveries are assumed at *t* = 0.

2.3 Non-pharmaceutical intervention

Without vaccines or efficient medicine against the disease, non-pharmaceutical interventions are the only effective way to prevent further increase of the pandemic [13]. These interventions take different approaches such as social distancing, social isolation and lockdown of the population. Despite the differences, they all carry the same objective, decreasing the infection rate *β*. It is convenient to implement the effect of these interventions on the model, when making predictions. Here, we model this effect by a logistic function, where *β* starts at a initial value *βi* and at some critical time *tc* a intervention is imposed and beta decreases to *βf* = *Pdecβi*, where *Pdec* is the fraction of *βi* decreased by the intervention. In France, studies estimate that the intervention decreased *βi* by 77% [21], therefore *Pdec* = 0*.*77 in France.

 (40)

where *τ* is a constant related to the time taken for the intervention to have the effect desired. Such model reconstruct the general behavior of interventions against the spread of the disease (Figure 4)

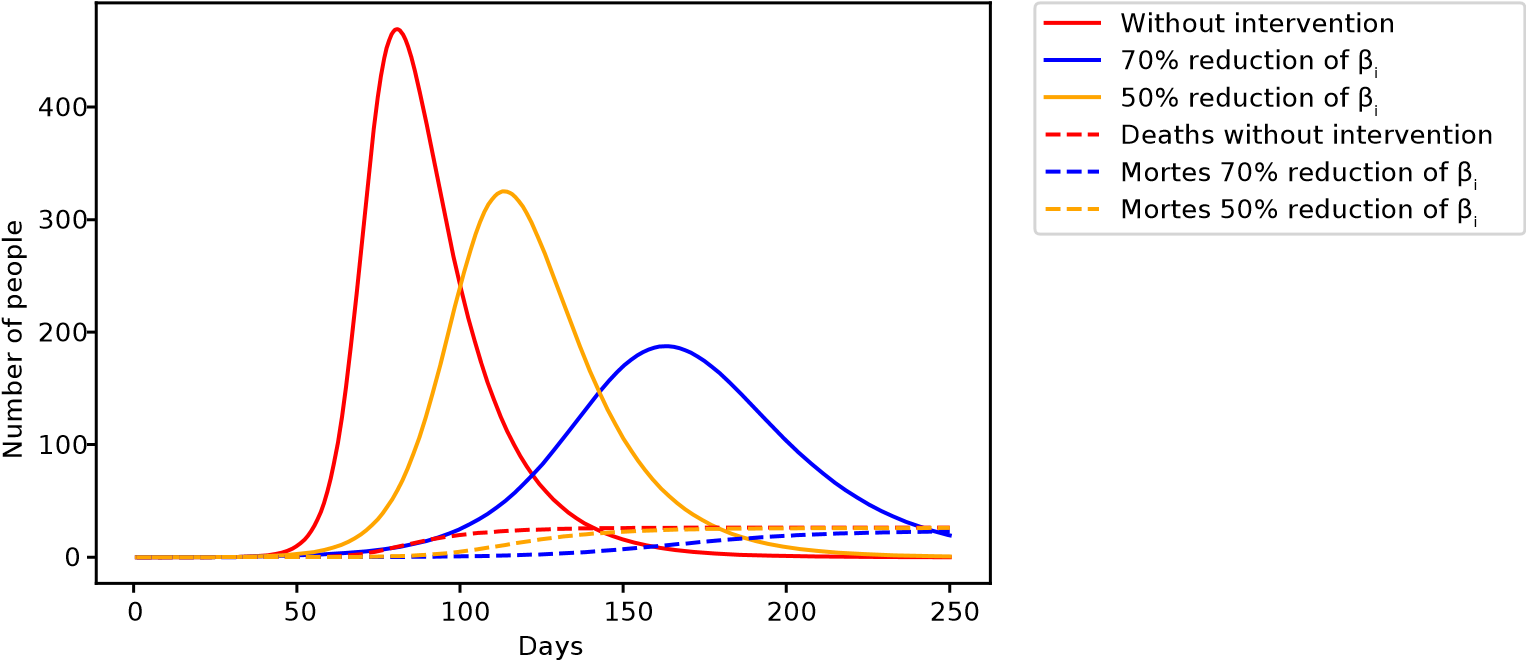


Figure 4: Visual representation of the effect of non-pharmaceutical interventions on the infection curve, depending on the efficiency of the intervention, given by *Pdec*.

2.4 Age division

Since the case fatality rate (CFR) of COVID-19 is different among age groups [6, 7, 22], we propose here a modification on both models, including the age distribution of the population and the social aspects of close contact between members of the population. The modification is describe as follow: Each compartment is divided into *M* age groups, where each *i*-th group has a *PCFRi* associated to it, that is, the probability of death associated to the *i*-th age group. The *β* parameter is now described as the average number of daily contacts between a member belonging to the *i*-th age group to the *j*-th age group, multiplied by the infection probability *Pinfc*

|  |  |
| --- | --- |
| *N βiI* = X*CijIjPinfc,*  *j*=1 | (41) |
| *N βiE* = X*CijEjPinfc,* | (42) |

*j*=1

where *Cij* is called the social contact matrix and we included *Ij* and *Ej* inside *β* now to place everything on the same sum. The age distribution among the population is retrieved from the UN prospects [23] and the social contact matrix for those countries was measured on previous studies [24, 25]. The specific contact matrix for the Republic of Korea was not found, however, [26] finds evidences of cultural clusters in the world, where countries belonging to the same cluster share cultural similarities; thus, we use this fact to justify the use of Hong Kong’s social contact matrix to describe the Republic of Korea. That way, we include cultural and population aspects for each of those countries, increasing the odds of a successful prediction. This type of model was used recently to describe the coronavirus outbreak on large cities in Brazil [27].

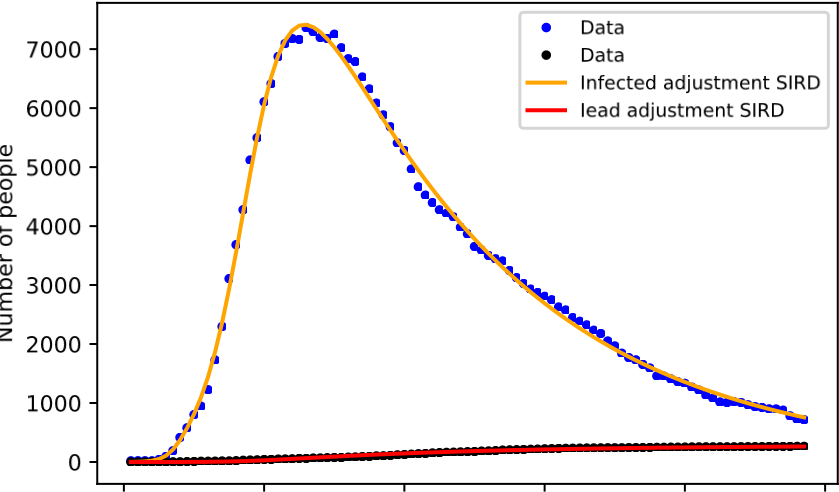
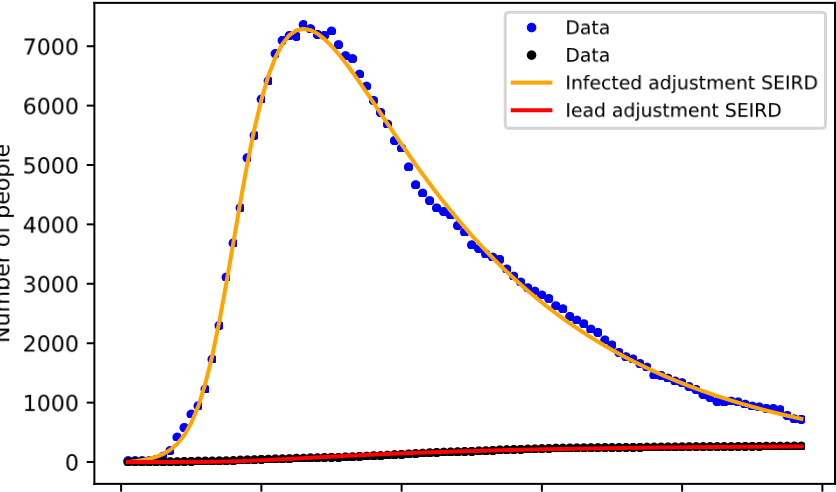
# Comparing Adjustments

To test the SIRD and SEIRD model we first compare them to the pandemic crisis on the Republic of Korea, running a numerical solution for the differential equations (23) - (27) we adjust the general behavior of the populations to Korean data acquired from [28] since 15/02/2020. The data from the Republic of Korea consists of the infection curve and death curve. To prevent problems with initial guess on the fitting process, both models used the same values for the initial guess, except *E*0, which is found only on the SEIRD model. *S*0 = *N* was also left as a free parameter of the adjustment instead of set to the total population of the country, which is justified by a limitation in both models, where the population is assumed homogeneously spread, which does not correspond to reality. Thus, *N* does not represent the total population, instead it represents an effective population smaller than the total population, due to non-homogeneous distribution throughout the territory, the interpretation of *N* as the disease evolves is discussed on the discussion session. The parameter was chosen to be *k* = 0*.*44*β*, we considered a study which estimated that presyntomatic cases caused 44% of infections [29], while for *c* we used an average of several clinical studies shown on table 1.

|  |  |  |
| --- | --- | --- |
| incubation time | 95% confidence | Reference |
| 6.4 days | 5.6-7.7 | [30] |
| 5.2 days | 4.1-7 | [3] |
| 5 days | – | [31] |
| 4 days | – | [32] |
| 5.1 days | 4.5 - 5.8 | [33] |

Table 1: Incubation time of the disease according to other studies.

Since the Korean government did not impose a lockdown or social isolation, we set *Pdec* = 0 in both models. Figures 5 and 6 show the result of the fitting process and table 2 includes the acquired values for all parameters for each model.



0 20 40 60 80 100 0 20 40 60 80 100

Days since 15/02/2020 Days since 15/02/2020

Figure 5: Fit for the infected and deaths by SARS-CoV-2 Figure 6: Fit for the infected and deaths by SARS-CoV-2 on the Republic of Korea using the SEIRD model. on the Republic of Korea using the SIRD model.

|  |  |  |
| --- | --- | --- |
| Parameter | SIRD | SEIRD |
| *χ*2 | 0.9978 | 0.9978 |
| *τr* | 7.9 ± 0.3 | 8 ± 0.3 |
| *τd* | 29.6 ± 0.2 | 28.8 ± 0.2 |
| *I*0 | 2 ± 1 | 1 ± 4 |
| *E*0 | – | 62 ± 57 |
| *β* | 0.478 ± 0.004 | 0.513 ± 0.008 |
| *N* | 11035 ± 57 | 11218 ± 48 |

Table 2: Parameters found by the adjustment with both models

The recovery time on both models is close to 8 days, while other studies such as [34] found 10 days. The time from symptoms onset to death was in both models close to 30 days, being 1 day shorter with the SEIRD model, [35] and [36] found *τd* = 18 or 11 days.

Comparing the accuracy of the fitting with the data, both models resulted the same value of *χ*2. The parameter *E*0 presents a large margin of error, which is expected given the lack of real data concerning the exposed population.

Proceeding to the calculation of R0 for both models, using equations (39) and (22) we found

|  |  |
| --- | --- |
| R0*SEIRD* = 1*.*92 ± 0*.*07 | (43) |
| R0*SIRD* = 2*.*98 ± 0*.*09 | (44) |

The value for R0 according to other studies ranges from 2 to 3 [37, 38, 39, 40], therefore, both models yield acceptable values for R0 being the one predicted by the SEIRD model lower.

# Prediction Accuracy with Age Division

We now proceed to test the prediction accuracy of both models. We used the first third of the data for fitting both models and extracting parameters, after having the parameters, we compare the prediction for the next days with these parameters with the rest of the dataset.

4.2 Korea

The training set consisted of 20 days, corresponding to the infections from 15/02 to 06/03. The SEIRD model found *Nmin* = 0*.*03% and *Nmax* = 0*.*04% of the total Korean population, while *Pinfc* varied between 80 to 85%.

The simple SIRD model found *Nmin* = 0*.*02% and *Nmax* = 0*.*03%, *β* went from 0.345 to 0.436, and *I*0 was between 23 to 65. Figures 9 and 10 present the result for prediction of both models.

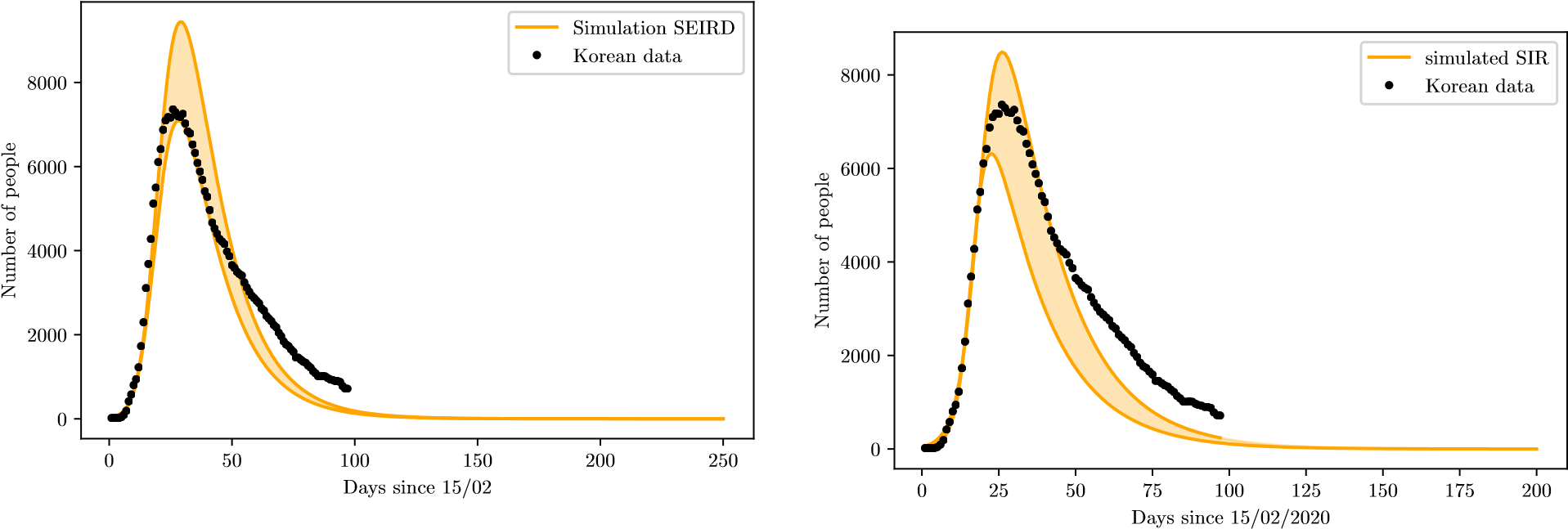


Figure 9: Prediction for the Republic of Korea in comparison with real data using the SEIRD model with age Figure 10: Prediction for the Republic of Korea in compardivision ison with real data using the simple SIRD model.

# Discussion

When concerning the adjustment process for acquisition of parameters with both models, there were no difference on the accuracy of the fit, and both models yielded very close values for the parameters. However, *τd* is super estimated in both models, being slightly lower on the SEIRD model. The value of *τr* is acceptable inside the variation of clinical measures.

The SEIRD model yields a slower growth rate than the SIRD model, that might happen due to the incubation period on the SEIRD model, which slows down the propagation of the virus towards other individuals. The main difference between the growth rate predicted by both models is better visualized by figure 11, the action of the incubation period slows down the rate of infection, as seen by the adjustments, but also decreases the peak of infections. However, the cumulative numbers of infection, deaths and recoveries are the same.

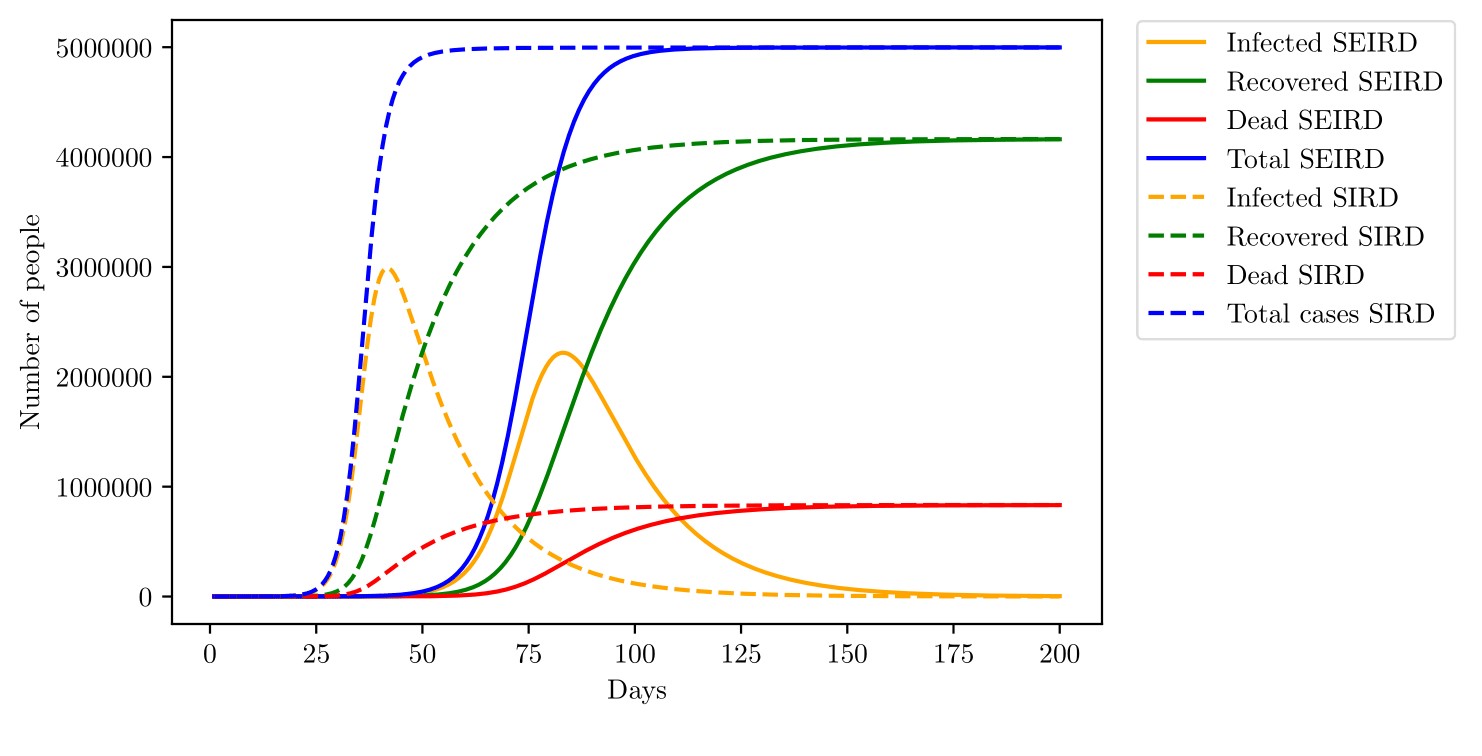


Figure 11: Comparison between SIRD and SEIRD models. The parameters chosen were the same for both models, except for *c* = 1*/*5*.*1 on the SEIRD. *β* = 0*.*45, *γ* = 0*.*054, *µ* = 0*.*0014, *k* = 0*.*44*β* and *N* = 6000000

*N* could be understood as the population susceptible to the first pandemic wave, due to the non-homogeneous distribution of the population, not everyone is susceptible to the disease right at the start. With such an interpretation, *N* tends to increase with time and approach the total population, here. Comparing predictions generated by the SIRD model with the SEIRD model with age division, the SEIRD model becomes a little more precise, although both simulations fail to predict the slower decrease of Korean data, that might be explained by an increase on *N* as time passes, resulting in new cases registered and therefore, slowing down the rate of decrease. Such hypothesis is well acceptable since the Republic of Korea did not adopt any lockdown or social isolation measure, making the disease able to propagate towards other regions, increasing *N* with time. Even with better prediction, the SEIRD model is far more complicated than the SIRD model and the use of the later should probably not compromise any data analysis. The same must hold true for simple SIR and SEIR models, when deaths are not a population to be accounted for, instead are just represented with a rate of removal for individuals.

The social isolation model developed here shows good results on the predictions, indicating that the description of *β* should be close to reality. Here we find a huge advantage of the SEIRD model with age division in comparison with the SIRD model; by including age division, it is possible to simulate the effect of specific non-pharmaceutical interventions, such as school closure, which in principle would decrease *βiI* and *βiE* for the age groups between 0 to 19 years only. Another possibility is to include isolation of only elderly individuals. Several non-pharmaceutical measures have been already described in literature [42], other studies show how the total number of infected might be changed due to the efficiency of non-pharmaceutical measures [43].

Other models might present more complete analysis of the disease, including hospitalizations and even asymptomatic cases, which are difficult to track and seem to vary a lot from place to place, the Diamond Princess cruise ship found 17.9% of asymptomatic infections [44], while an airplane flight found 11.2% of cases being asymptomatic. An Italian village presented 50 to 70% of cases being asymptomatic [45]. There are yet the problem of assuring that the asymptomatic cases registered on studies are really asymptomatic and not presymtomatic, that is, are people still on the incubation period.

Of course, any mathematical model is only as good as the data allows, using mathematical models to describe the disease on countries with low testing rates might yield unrealistic predictions. For example, [46] estimates 86% of infections being undocumented on China, at the early stages of the outbreak.

Another consideration we did not take, was the possibility of reinfection, where individuals leave the recovered group and re-enter the susceptible compartment. However, since other coronaviruses belonging to the same genus *betacoronavirus* such as the SARS-CoV and the MERS-CoV does not present a high enough mutation rate to cause reinfection in short term [47], the only cause of reinfection would be the loss of antibodies to fight the virus; nevertheless, on both diseases, the infected person acquires antibodies enough to prevent reinfection for a period of 2 - 3 years [48]. With those considerations, we did not assume reinfection was probable on short-term. Future studies may be conducted to study the possibility of reinfection of individuals on the long-term.

# CHAPTER 4 EMPIRICAL DATA ANALYSIS

# CHAPTER 5 CONCLUSION

The proposed hybrid model consists of a dynamic SEIRD model with vital dynamics and decaying COVID mortality rate and three ARIMA models that cancel out dynamic model residuals and enhance prediction quality. Unlike pure dynamics models like SIR, SEIR, and SEIRD, this model allows us to make precise predictions for up to 2 months ahead.

The model was tested on US COVID statistic data. Obtained validation results allow us to conclude that the proposed hybrid model has good prediction ability and decent performance. Obtained long-term predictions reflect the general dynamic of the outbreak and are especially useful for the healthcare system workers and government officials. Obtained short-term predictions allows us not only to forecast the future number of infected, recovered, and deceased patients, but only estimate forecast error under adverse or optimistic circumstances. The proposed method can be used as an effective tool for prediction and analysis of the dynamics of the COVID-19 pandemic.

Here are some perspective ways of further development of the proposed method:

(a)

Parameter estimation with different algorithms and boundaries;

(b)

Testing the method of COVID statistics in other countries;

(c)

Develop alternative methods for residue prediction;

Enhancing of the proposed hybrid model depends on profound research results about COVID-19. That is why monitoring recent research in the field and quickly adjusting the model according to the new data is crucial.

## Declaration of competing interest

# REFERENCES

References

1. Sourish Das, Prediction of Covid-19 Disease Progression in India, arXiv:2004.03147 [q-bio.PE],2020.
2. *Covid19India* <https://api.covid19india.org/>
3. Worldometer, <https://www.worldometers.info/coronavirus/>
4. Yi-Cheng Chen, Ping-En Lu, Cheng-Shang Chang, Tzu-Husan Liu, A Time-dependent SIR Modelfor COVID-19 with Undetected Infected Persons, [http://gibbs1.ee.nthu.edu.tw/A\_TIME\_ DEPENDENT\_SIR\_MODEL\_FOR\_COVID\_19.PDF](http://gibbs1.ee.nthu.edu.tw/A_TIME_DEPENDENT_SIR_MODEL_FOR_COVID_19.PDF) , 2020.
5. *Github Link* <https://github.com/Taarak9/COVID-19_Predictive_Model_SEIRD/>
6. Samuel M. Jenness, Steven M. Goodreau, Martina Morris EpiModel: An R Package for MathematicalModeling of Infectious Disease over Networks, J Stat Softw. doi:10.18637/jss.v084.i08, (2018).
7. Chayu Yang and Jin Wang , A mathematical model for the novel coronavirus epidemic in Wuhan,China, Mathematical Biosciences and Engineering, Volume 17, Issue 3, 2708–2724, (2020).
8. Zifeng Yang, Zhiqi Zeng, Ke Wang et al. Modified SEIR and AI prediction of the epidemics trend ofCOVID-19 in China under public health interventions, J Thorac Dis, 12(3):165-174, (2020).
9. Biao Tang, Nicola Luigi Bragazzi, Qian Li et al. An updated estimation of the risk of transmission ofthe novel coronavirus (2019-nCov), Infectious Disease Modelling 5, 248-255, (2020).
10. Meher K. Prakash, Shaurya Kaushal, Soumyadeep Bhattacharya, Akshay Chandran, Aloke Kumarand Santosh Ansumali, A minimal and adaptive prediction strategy for critical resource planning in a pandemic, medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057414, April, (2020).
11. Gaurav Pandey, Poonam Chaudhary, Rajan Gupta, Saibal Pal, SEIR and Regression Model basedCOVID-19 outbreak predictions in India, arXiv:2004.00958 [q-bio.PE], (2020).
12. <http://c19.dakshas.org:5000/>
13. Andrea Remuzzi and Giuseppe Remuzzi. COVID-19 and Italy: what next? *The Lancet.* 2020, 395:1225-28.
14. Edward Livingston and Karen Bucher. Coronavirus Disease 2019 (COVID-19) in Italy. *JAMA.* 2020, 323(14):1335. https://jamanetwork.com/journals/jama/fullarticle/2763401

15.Graziano Onder, Giovanni Rezza and Silvio Brusaferro. Case-fatality rate and characteristics of patients Dying in relation to COVID-19 in Italy. 2020, 323(18):1775-1776.https://jamanetwork.com/journals/jama/fullarticle/2763667

16. Guangchuang YU, <https://github.com/GuangchuangYu/nCov2019>. Accessed on May 18, 2020.