Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

Modeling and Analysis of Big Data for Covid-19 Epidemic

Georgi Simeonov

Thesis for awarding
the educational and scientific degree "Doctor"

Field of Higher Education:
4. Natural Sciences, Mathematics and Informatics
Professional field:
4.6. Informatics and computer science
Doctoral program: Informatics

Scientific Adviser: Prof. Dr.Sci. Peter Stanchev

Scientific Consultant: Prof. Dr.Sci. Ognyan Kounchev

Contents

List of Figures

General Remarks to the Dissertation

We provide the main features of the present Dissertation.

Timeliness and relevance of scientific problems addressed by the Dissertation

Infectious diseases account for many millions of deaths every year and are the main reason for human mortality. The well know deadly diseases are pneumonia, tuberculosis, diarrheal diseases, malaria, measles, HIV/AIDS, and more recently, the COVID-19. The control of epidemic outbreaks may be considered as the main driver for the development of the modern Epidemiology, in particular of the development of the new area of **Digital and Computational Epidemiology**.

During the last decade, an intensive worldwide effort is speeding up the developments in the establishment of a global surveillance network providing epidemiological Big Data, for combating pandemics of emergent and re-emergent infectious diseases. This trend has developed recently very strongly due to the Covid-19 pandemic.

Scientists from different fields extending from medicine, virology, immunology, genetics, molecular biology to computer science, statistics and applied mathematics have teamed up for rapid assessment of potentially urgent situations. Towards this aim mathematical modeling plays a central role in efforts that focus on forecasting, assessing, and controlling potential outbreaks. To better understand and model the spread of the contagious diseases the impact of numerous variables ranging from the micro host–pathogen level to host-to-host interactions, as well as prevailing ecological, social, economic, and demographic factors across the globe have to be analyzed and thoroughly studied. For that reason it is important to be aware of the main approaches that are used for the surveillance and modeling of infectious disease dynamics, and the basic concepts underpinning their implementation and practice in the area of available Big Data. See Liu et al. [2020], Kuhl [2021], Siettos and Russo [2013].

The present Dissertation represents a contribution to the newly emergent area of Digital and Computational Epidemiology. The last is a "Big Data"-Driven Modeling of contagious diseases, in particular of COVID-19, and may be defined as the building and study of new mathematical models for analyzing of epidemic outbreaks including their computer implementations and simulations in the framework of Big Data.

State of the Art of the problem studied

One of the main objectives of the present Dissertation is to build models for the analysis of Covid-19 disease spread as well as to provide their Web-based implementations, as instruments for interactive real time data analysis. There exists a list of popular world wide instruments for analysis of the COVID-19 spread, supplied with Web-based (free) online Analytical Tools. To name a few of the most popular:

1. The tool "Projections" by the IHME Institute is available at the link `https://covid19.healthdata.org/bulgaria?view=total-deaths tab=trend`
2. Tool Delphi from the Massachussets Institute of Technology, `https://www.covidanalytics.io/home`
3. Tool at Imperial College London, `https://www.covidsim.org`
4. The Los Alamos National Laboratory (LANL), `https://covid-19.bsvgateway.org/`
5. The SI-KJalpha model from the University of Southern California, `https://github.com/scc-usc/ReCOVER-COVID-19)`

Objectives of the present work

The main objective of the Dissertation is to develop new mathematical models and to provide their computer implementations in the form of Web-based instruments. The tasks which have been solved for achieving the above objective are:

1. The building of a model for short-term forecast of the epidemiological curves, TVBG-SEIR, based on splines with four nodes representing the transmission and the removal coefficients of the SEIR model. It analyses the short-term global evolution of the epidemics controlled by the introduction of lockdown/open up measures by the authorities. The incorporation of different lockdown prediction scenarios varying in time permits to analyze not only the primary epidemic wave but

also the arising secondary wave and any further waves. See Chapter 2.

2. Developed a web-based Scenario Building Tool for COVID-19 (shortly, SBT-COVID-19), based on the software platforms Jupyter and Bokeh; it uses Jupyter Notebooks Architecture: Jupyter Notebooks work with a two-process model based on a kernel- client infrastructure. The tool may be used as a decision support software by (health) policy makers to explore various scenarios, by controlling/changing the scale of the containment measures (home and social isolation/quarantine, travel restrictions and other) and to assess their effectiveness. The SBT-COVID-19 Tool permits to assess how long the lockdown measures should be maintained. See Chapter 2 for the description of the functionalities of the tool and Chapter 4 for description of the software solution.

3. The building of a model for long-term forecast, ATVBG-SEIR, based on estimation of the duration of the Epidemic of COVID-19 in a single country, accounting for different scenarios. We included vaccinations in the model which are carried out according to a vaccination plan provided on a monthly basis. We have modeled the seasonal effect as well. The algorithm takes into account the main constraint of the health system which is the number of Intensive Care Units (ICU) intended for COVID-19 patients. See Chapter 3.

4. We have developed a web-based Lockdown Scenarios Tool based on the software platforms Jupyter and Bokeh; the tool is available online at `http://atvbg-seir.eu`. It is using the algorithm implementing the methodology of the ATVBG-SEIR model. Results are demonstrating the effciency of the tool by applying it to COVID-19 data from Austria, Bulgaria, Germany, Italy, UK and USA. See Chapter 3 for the description of the functionalities of the tool and Chapter 4 for description of the software solution.

Acknowledgments

I express my thanks and gratitude to Prof. Dr.Sci. Petar Stanchev for being working together for years on Open Access projects and to all my colleagues from IMI-BAS who helped me with opinions, suggestions, comments and feedback.

My sincere gratitude to scientific consultant Prof. Dr.Sci. Ognyan Kunchev for the numerous discussions on various topics for mathematical modelling of the dissertation and the invaluable advice support and encouragements during the working on dissertation.

This dissertation received financial support from KP-06-N32-8 and KP-06N42-2 of the Bulgarian Research Fund (NSF) and support with a grant from the European Open Science Cloud (EOSC) within the EOSC Secretariat.

Chapter 1

Introduction

In the present Chapter we discuss the role of Big data in Medicine and in particular in Epidemiology related to Covid-19. We also introduce some basic concepts of modeling in Mathematical Epidemiology.

Historically, mathematical modeling in Epidemiology has appeared already in the works of Bernoulli about 1766, Straif-Bourgeois et al. [2014], but until recently, it was not much used as a broadly accepted tool for public health policies, but was considered a specialized research area for applied mathematicians and theoretical biologists. Things have changed with the advent of the HIV pandemic, when mathematical models were first used to predict future epidemic spread, and to analyze the impact of behavior change on HIV incidence.

The real breakthrough for mathematical modeling as a public health tool came with the concerns that smallpox virus could be used in a deliberate release and lead to devastating outbreaks in the only partially immune populations of present societies.

How can public health policy be developed against threats with pathogens that are *not circulating at present*? There is no way to conduct epidemiological investigations, and the only available data in the case of smallpox were from before the eradication era. Therefore, to design policy, knowledge from *historical smallpox* outbreaks had to be combined with *data about present-day society*, and possible interventions had to be tested on the basis of this available information. Mathematical modeling provided a flexible tool to solve the problem, and was used to analyze possible vaccination strategies and other interventions. Mathematical modeling has shaped the present paradigms of infectious disease epidemiology. In modern terms, for analyzing the smallpox spread, the epidemiologists and the modellers have solved the epidemiological problem by applying an approach which belongs to the Big Data paradigm and Machine learning – massive use of historical data from various countries, generation of projections (or, prediction scenarios) by different alternative models, and by "Machine learning" from the histor-

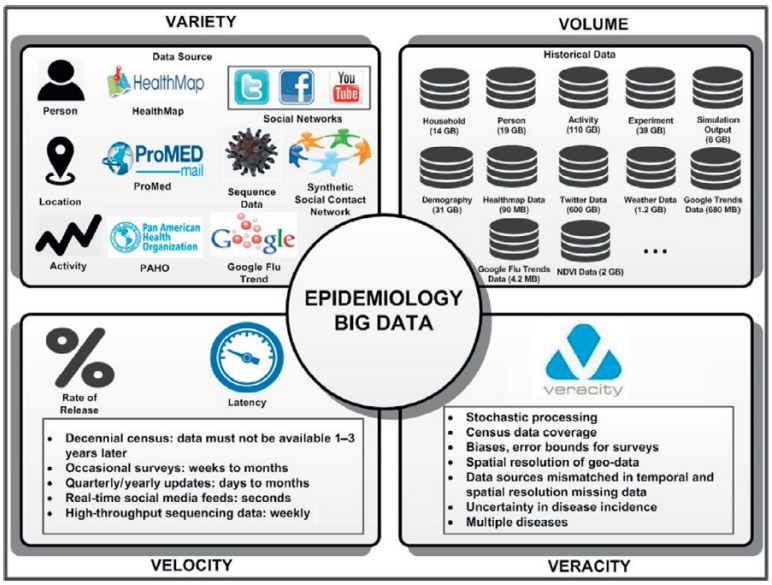


Figure 1.1: Image courtesy: S.M. Shamimul Hasan, Virginia Tech.

ical data.

In the Appendix, section 7.1 we provide a more detailed account of the role of the modern mathematical modeling in Epidemiology following Straif-Bourgeois et al. [2014].

In view of the above observations the modern Epidemiology supplied with the appropriate mathematical modeling may be considered as a science based on Big Data in all aspects of this notion.[1] We refer to Pyne et al. [2015] for a detailed discussion of the role of Big Data in solving the urgent problems of Epidemiology related to the modern societies' preparedness for controlling the spread of pandemics, ironically published in 2015. We provide the following Figure visualizing the Big data issues in Epidemiology: variety, volume, velocity, and veracity.

In order to give an idea about how big the Big data are in the range of problems which we deal with, we mention the following facts about the formats and volume of the daily data on Covid-19: The data provided by majority of sources and World Health Organization (WHO) are prevailing **Time series** in CSV format and are in a raw format. However they are usually transformed in some Relational Database but this normally happens after some time, not immediately. The popular Institute for Health Metrics and Evaluation (IHME) which maintain the popular Projections Tool, has a detailed description of their datasets containing data for all countries in

[1] About the recent concepts of Open Data and Big Data we refer to Stanchev et al. [2020] and Srebrov et al. [2020].

the world: COVID-19 estimate downloads are available at the link `http://www.healthdata.org/covid/data-downloads` . The decompressed daily dataset is about 600 MB, which means that the size of a three month period (normally used for the models TVBG-SEIR and ATVBG-SEIR) is about 54 GB of data.

1.1 Big Data Applications in Health Sciences and Epidemiology

Infectious diseases account for more than 13 million deaths a year and are the main reason for human mortality. The well know deadly diseases are pneumonia, tuberculosis, diarrheal diseases, malaria, measles, HIV/AIDS, and more recently, the COVID-19.

An *epidemic* is an occurrence in a region or a country, of large number of cases of a disease, which exceeds a certain measure of normality. A *pandemic* is an epidemic that spans a large portion of the world, such as the H1N1 outbreak in 2009.

Epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the prevention and control of health problems, Last [2001]. Unlike the medical sciences, the main concern of Epidemiology is on population-level issues. Epidemiologists are primarily concerned with public health, which includes the design of studies, evaluation and interpretation of public health data, and the maintenance of *Big Data* collection systems.

The *control of epidemic outbreaks* may be considered as the main driver for the development of the modern Epidemiology. It is remarkable that the collection of Big Data in Epidemiology was one of the main sources for the new science, when the medical Dr. John Snow's has discovered in 1854 the causes for the *Broad Street cholera outbreak* (currently, Broadwick Street), in the Soho district of the City of Westminster, London, England. Earnestly, and risking his life, Dr. Snow, walking from door to door in the Soho district, has collected the data about the fatalities. He put on the map all locations (houses) where death cases have occurred, which may be considered as the very first model with Big Data in Epidemiology. Dr. Snow proved using his map that germ-contaminated water was the source of cholera, rather than particles in the air (referred to as "miasma"), see Wiki page `https://en.wikipedia.org/wiki/John_Snow` for more details.

With the advent of modern science, pharmaceutical measures have been widely used to control and prevent outbreaks. For example, **vaccines** have become a critical method of controlling, preventing, and possibly eradicating infectious diseases in host populations. Despite their success, nonpharmaceutical methods, such as **quarantining** and **social distancing**, continue to play a central role in **controlling** infectious disease outbreaks; they are

especially important during an outbreak caused by an **emerging pathogen**, when pharmaceutical methods are still not available.

Public health authorities have made significant strides in reducing the burden of infectious diseases. Nevertheless, infectious diseases continue to be an important source of concern. A number of global trends further amplify these concerns: increased urbanization, increased global travel, denser urban regions, climate change, and increased older and immunocompromised population.

Many of the changes that we see around us are, to a large extent, anthropogenic and are happening at a scale wider and faster than ever before in human history. Further, **new pathogens** are emerging regularly, which raises the importance of societies' need to **understand** and be prepared to **systematically address** the challenge of **emerging infectious diseases** at different levels. In particular, it is necessary to develop adequate **Mathematical models** for the spread of the new infectious diseases, which are able to coop with the large size of the data, arising from the potential pandemic character of the disease spread.

There is a growing concern about our preparedness for controlling the spread of pandemics such as COVID-19 and similar viral infections. The dynamics of epidemic spread in large-scale populations is very complex. On the other hand, human behavior, social contact networks, and pandemics are closely intertwined and evolve as the epidemic spreads. Very often the normal social interactions are changing drastically in response to the public policies undertaken by the health authorities which makes the modeling of the pandemics even more complicated. The planning and response strategies by the authorities must take these complicated interactions into account. **Mathematical models** are key to understanding the spread of epidemics. In the present work we consider newly developed mathematical models for studying the complex dynamics of epidemics in large-scale populations, based on the classical SIR/SEIR models in Epidemiology.

Using these models for creating scenario predictions (called sometimes nowadays *projections*) for the epidemic spread and developing public health policies leads to problems that are typical for the Big Data applications. Thus our methods are applicable to Big Data in Epidemiology.

The role of epidemiologists is central in the pandemic and the **mathematical models** they produce are instrumental in understanding how the virus might impact populations, helping to inform *government policy* around the world, Czyzewski [2020]. The modern epidemiologists view themselves as **mathematical modellers** working in isolation from the wider public health community; when actually they are epidemiologists first, who happen to have a more quantitative background and they work quite closely across disciplines, as Immunology, and Clinical medicine. Therefore **mathematical modelling** isn't just a case of producing these projections, but it's a way of formalising a lot of the information exchange that the modern

epidemiologists have with their public health colleagues.

The importance of the concept of Big Data in Epidemiology may be demonstrated by examples. The attempts to explain the phenomenon of epidemics peak out of season of the COVID-19 in India shows that Epidemiology has to be a science based on Big Data if a reasonable explanation has to be found. Indeed, in order to give a scientifically argumented explanation, one needs to extrapolate from the data for Flu season in previous years since no genuine summer COVID-19 summer season existed in 2020. In a similar way, the attempts to model the summer season low of the COVID-19 in Bulgaria has to be based on data about the flu summer season in other countries.

1.2 Compartmental and Mathematical SIR/SEIR Models in Epidemiology

The main focus of the present work are the models which belong to the family of the so-called Compartmental models. Before providing the State-of-the-Art models we will introduce the concept of Compartmental models as well as the basic notions and notations which will be used lateron.

Compartmental models are a framework used to model in an adequate way the dynamics of infectious disease. The population is divided into compartments, with the assumption that every individual in the same compartment has the same characteristics. This framework has been developed for the first time in the paper of Kermack and McKendrick in 1927 Kermack and Mckendrick [1927]. One may use a deterministic approach using a system of ODEs or the framework of the stochastic models. In the present work we devote our studies to the deterministic approach, and we will be focused primarily on the so-called SEIR model. The main reason for choosing SEIR model against its simpler relative SIR model, is the *long incubation period* of COVID-19, hence, the large "exposed cases" compartment to be defined properly below.

For a detailed and excellent and concise introduction to the compartmental models we refer to the monograph Keeling-Rohani, Keeling and Rohani [2008], and to Straif-Bourgeois et al. [2014]. For completeness sake, we provide a description of the deterministic SEIR model which will be the main approach in our research. Let us remark that during the modeling process one needs to consider several (four) different settings of the disease spread. The setting based on the direct reference to the physical / medical side is the model which we describe below – we call it *Realistic model*. Then, we have the official sources of data, which represent the so-called *Official model*. Finally, we have the *Ideal model* which is meant to provide the best Mathematical approximation to the Realistic model.

We start with the description of the *Realistic* (Compartment) SEIR

model.

The classical SEIR model considers the society of individuums subdivided into four compartments: C_S, C_E, C_I and C_R which are described as follows:

1. Compartment C_S : it has size $Sr(t)$, and contains the number of "susceptible" people at time t. Usually, at the start, $S(0) = N$ is the whole population of the country under consideration. It is supposed that nobody has automatic immunity against the virus, i.e. everybody is susceptible. However, it is also possible to assume that a part of the population is not susceptible and this is a serious assumption in some models which needs to consider an additional parameter as the percentage of the susceptible proportion of the whole population.

2. Compartment C_E : its has size $Er(t)$, which is the number of "exposed" people at time t - these are the people who have come in a contact with virus spreaders, and are "virus carriers" but are not "virus spreaders"; the virus is in a latent form, and usually they do not show symptoms of sickness. For different viruses the incubation (latent) period is very different – for the coronavirus it was recently statistically estimated from empirical data, that for 99% of the cases, the mean incubation period is 5.9 days, while the minimum is 3 days and the maximum is 14 days, (see the references Mcaloon et al. [2020], [Daley et al., 2020, Bibliography to chapter 2]). Not everybody in C_E may become "virus spreader", i.e. move to the next compartment C_I. Practically, the compartment C_E does not enter the official statistics since it is not observable, but it is very important for the modeling of the dynamics of the virus spread. This compartment is missing in the simpler SIR model.

   Here it is important to distinguish three categories: persons who never becomes sick, persons who are in C_E and become sick (virus spreaders) but without symptoms (asymptomatic), and the usual persons who become sick with symptoms. The two last categories become immediately members of the compartment C_I. Let us remark that there are various studies recently, which show that even if the person is still not showing any clinical symptoms (i.e, he/she is presymptomatic), he/she could be a source of infection spread, and lateron he/she gets symptoms.

3. Compartment C_I: its size is $Ir(t)$, which is the number of infectious cases at time t - these are the people who are "virus spreaders", majority of them show some symptoms, although they may not show any symptoms (asymptomatic). It is important to understand in the modeling that many people who are diagnosed positively are almost

immediately hospitalized or quarantined. Hence, after quarantine they go immediately to compartment C_R, but they have stayed in C_I only until they have been diagnosed. The paradox is that these are the official data which we obtain – the number of those who are officially registered is denoted by $Idata\,(t)$.

4. Compartment C_R : its size is $Rr\,(t)$ – the number of recovered, quarantined, immune, or deceased individuals, which are all called "**removed**". Normally they come from compartment C_I after becoming healthy (or dead) and no more virus spreaders. Officially these data are provided in a cumulative way.

Some models distinguish the fatalities in a variable D, hence, SEIRD model; others distinguish symptomatic and asymptomatic, many models also consider the natural birth and mortality rates incorporated in the equations, etc.

The simplest Compartment models are the SIR models (with missing C_E compartment) and SEIR models which are based on some simplified assumptions:

1. Everybody in the population of size N is susceptible; those who have an inborn immunity are considered to be out of consideration since they do not get infected and infection spreaders. This group contains also people who are completely isolated in the society.

2. We assume that the population of size N is perfectly and evenly (uniformly) mixed.

3. It is assumed that everybody who becomes infectious has the same transmission capacity which is summarized by the transmission coefficient β of the models. Obviously, this is in practice not true since some people spend the disease heavily but others spend it even without symptoms; hence, the last spread the viruses much less.

4. The coefficient β which will be discussed below includes also *the intensity* of the contacts within the population, which is also assumed to be equal for everybody in the population.

There are many Compartment models which are a lot more complicated than the usual SIR and SEIR models, and which enter in a lot more details, by subdividing the population according to its demographic and health structure: age, sex, health status, employment, number of contacts, etc. Further, one may use more detailed information on the population density, age structure, transport links, the size of social networks and health-care provision, etc. For completeness sake, we mention some Compartmental Epidemic Models with their diagram representation:

Figure 1.2: SI/SIS model

Figure 1.3: SIR model

SI / SIS Model: Some infections like those from the common cold and influenza, do not confer any long-lasting immunity. Such infections do not give immunity upon recovery from infection, and individuals become susceptible again. The diagram of the model is provided in Figure 1.2.

SIR Model: SIR is an epidemic model as illustrated in Figure 1.3, the SIR model characterizes the dynamics among the susceptible individuals (S), infectious individuals (I) and removed individuals (R) or recovered/deceased - individuals that are assumed cannot become susceptible again.

SEIRD Model: SEIRD stands for Susceptible Exposed Infectious Recovered and Dead. The dead individuals are in a separate compartment. There are some diseases for example Ebola, where dead can still be infectious and SEIRD model may be also used to study the interactions between D and E individuals (Weitz et al.). For more details we refer to `https://www.nature.com/articles/srep08751`. The diagram of the model is provided in Figure 1.4.

Let us focus on the description of the SEIR model used as a basis for the spline models which we develop further here: Important ingredient of the SEIR model is the proportion of the infected people who die, i.e. the **fatality rate**. This is also very dependent on the age structure of the population, but we consider it again as an averaged percent.

Insights into the dynamics of the SIR model have led to the definition of a number of important concepts that are universal for all models and all infectious diseases. The most important of these concepts is the basic reproduction number $\mathcal{R}_0$. It is very important indicator for an epidemic. The number $\mathcal{R}_0$ is intimately related to the Compartmental models and may

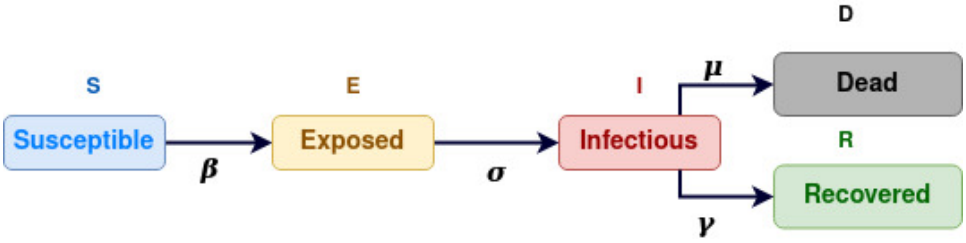


Figure 1.4: SEIRD model

be very effectively determined by them. The basic reproduction number $\mathcal{R}_0$ is defined as the number of secondary cases caused by one index case during his/her entire infectious period in a susceptible population. In other words, the basic reproduction number is given by the product of the transmission rate (number of new infections *per time unit*) and the *duration* of the infectious period, Straif-Bourgeois et al. [2014].

Further peculiarities of the Coronavirus spread have become clear, especially after the Chinese experience (He et al. [2021], Nature [2021]), in particular, that 80% of those who have met the virus have spent the disease in an asymptomatic way. But these 80% still actively have spread the disease. Hence, the official data which are only available need to be considered as no more than 20% of the whole compartment C_I. These issues will be discussed below as an important component of the modeling.

It has been a subject of numerous investigations that even people in the incubation period before the appearance of symptoms, spread the virus for several days. Many investigations are devoted to the spreading of the virus several days after the disappearance of the symptoms by people who have become sick. It has been estimated that the asymptomatic virus spreaders are responsible for about 33% to 59% of the total spread of the Coronavirus, Greenhalgh et al. [2021].

For the more detailed modeling of the COVID-19 infection spread, it is important to summarize some facts about the immunity of those who have spent the disease: the Chinese experience has shown (He et al. [2021], Nature [2021]) that after 9 months almost 99% of the tested persons are losing gradually their natural immunity. Until the present moment there is no certain data about how long the immunity lasts after vaccination; it is not clear whether the particular vaccine, the racial, sex, or other differences really matter.[2]

[2]From the above mentioned references He et al. [2021], Nature [2021]:
"The neutralizing antibodies that the immune system produces to disable the virus SARS-CoV-2 can last for at least nine months after infection, but not everyone makes them in detectable quantities."

There exist other alternative approaches to the deterministic SIR/SEIR models: Let us mention that there is a class of Agent-based models which however need an enormous amount of data but the final results do not differ essentially, Czyzewski [2020]. The Stochastic models using branching processes also need more complicated implementations but are very useful in the initial onset of the disease spread.

The SIR/SEIR type models alow for a simple incorporation of the health policy measures, e.g. when large parts of the population start to work from home (online). Unlike the SIR/SEIR models, the Stochastic models and the Agent-based models one needs to incorporate enormous amount of information for tracking each individual separately, but the final effect will be the same.

Remark 1 *Due to the lack of sufficient data for the COVID-19 epidemic, in particular for modeling the so-called seasonal effects, it is very important that influenza epidemics may be used as suitable for use as a model for the COVID-19 epidemic, given that they are respiratory diseases with similar modes of transmission. However, data directly comparing the two diseases are scarce, Piroth [2020].*

1.2.1 Further assumptions and characteristics of the Comparment and SIR/SEIR models

1. The simplest SIR/SEIR models make basic assumptions, such as that everyone has the same chance of catching the virus from an infected person because the population is perfectly and evenly mixed, and that people with the disease are all equally infectious until they die or recover. More-advanced models, which make the quantitative predictions policymakers need during an emerging pandemic, subdivide people into smaller groups — by age, sex, health status, employment, number of contacts, and so on — to set who meets whom, when and in which places (see section 'Measuring social mixing' in Adam [2020]).

2. Another assumption is that there is no natural immunity to COVID-19 — so the entire population starts out in the susceptible group — and that people who recover from COVID-19 are immune to reinfection in the short term.

3. An alternative to the SIR/SEIR models are the *Agent-based models* which build the same kinds of model, but each person can behave differently on a given day or in an identical situation. However these very specific models are extremely data hungry. One needs to collect

"The researchers found that only 7% of the population had been infected with the virus, of whom more than 80% had had no symptoms. Around 40% of the infected people produced neutralizing antibodies that could be detected for the entire study period."

information on households, how individuals travel to work and what they do at the weekend.

4. Let us note that in favour of our choice of the SEIR model speaks the fact that the team at Imperial College which has a very long experience in the application of mathematical models in Epidemiology, has switched from the agent based model to the SEIR model in March 2020, due to its simplicity and the fact that both type of models "give broadly similar overall numbers", Adam [2020].

5. An important feature of the algorithms which we consider, is a parameter which is the percentage of the severe cases (hospitalized) which need Intensive Care Units (ICUs). A curious story about how this percentage has been adaptively found in UK is available in Adam [2020], and shortly retold in Appendix **??**.

6. The range of the main epidemiological parameter, the basic reproduction number $\mathcal{R}_0$, was not clear at the beginning of the COVID-19 pandemic; initially it was estimated upwards to between 2.4 and 3.3; at the end of March 2020, in a report on the spread of the virus in 11 European countries, the researchers put it somewhere in the range of 3 to 4.7.

   In the practice of the SIR and SEIR models the value of $\mathcal{R}_0$ is found after fitting of the model to the data.

7. Another important feature, are the unreported cases. In Germany these are called *Dark numbers*. The example of China has shown that in a very representative sample of 9500 persons, the percentage of those who have been sick with COVID-19 and have not shown symptoms are 80%. Hence, a rough estimate shows that we have to multiply the number of officially infectied data by 5. On the other hand, for finding the Dark numbers, the Robert Koch institute in Berlin is using a coefficient between 4 and 6; cf. also the website of an independent organization in Germany studying the dark numbers, using a specific model: https://covid19.dunkelzifferradar.de/, and an academic publications, Anguelov et al. [2020].

8. Further curious facts about the containment measures and the *Efficacy of Masks wearing*: study in 401 regions in Germany (see Peeples [2021]) has shown that requiring people to wear face masks decreases the daily growth rate of reported COVID-19 cases by more than 40% . In USA, Canada: In a similar study published in January, 2021, researchers found that a national mandate for employees to wear face masks early in the pandemic could have reduced the weekly growth rate of cases and deaths by more than 10% in late April 2020. This could have

reduced deaths by as much as 47% (or by nearly 50,000) across the country by the end of May 2020.

9. A more sophisticated approach to the definition of the basic reproduction number is provided in Van Den Driessche and Watmough [2002], the basic reproduction number $\mathcal{R}_0$ is the spectral radius of the next-generation matrix, see also Feng et al. [2021].

10. In some versions of the SEIR model (Radulescu et al. [2020]) researchers account separately for presymptomatic and asymptomatic transmission, by introducing two new compartments.

1.3 Official data available for modeling purposes

In practice, we do not have the "reality data" $Sr(t)$, $Er(t)$, $Ir(t)$, $Rr(t)$ which correspond to the Compartments of the SEIR model, described in section 1.2. It will be very helpful for somebody willing to fit a model, to know precisely what are the official data on the day t which are available, and how reliable they are. The daily data available are as follows:

1. The official (observed) data $Idata(t)$ which represent the *daily* "new infected cases" with COVID-19, and these are normally people with serious symptoms. These are the cases which have been tested and registered officially at the hospitals. The majority of them are almost immediately hospitalized or quarantined, hence, they are almost immediately moved from compartment C_I to compartment C_R. However it is well known that the number of the asymptomatic cases having COVID-19 represents about 80% of all infectious cases (see He et al. [2021], Nature [2021]), and they are in fact the most active infection spreaders. Thus, the size of the compartment C_I is much bigger than that indicated by the official data $Idata(t)$, and we have for that reason the inequality

$$Idata(t) \leq Ir(t). \tag{1.1}$$

   We denote by $TotalInf(t)$ the total number of reported infectious cases, which is a cumulative sum of the above daily data $Idata(t)$.

2. The officially announced data of recovered data, $Rdata(t)$, contains the cumulative number of recovered cases. It is important to remark that not all health authorities provide these data regularly. For example, in USA they have stopped to provide them since December 2020, and the main reason was explained to be the lack of proper precise definition of the meaning of "recovered". Also, in UK there is a lack of data about the recovered cases.

3. Data on the total number of fatalities due to Covid-19 - the precise diagnose of those who have died really by the Covid-19 infection is not well defined. We denote these data by $Ddata\,(t)$ which represents the cumulative number of fatalities until the date t.

4. Although not from the very start of the epidemy, the authorities provide the data about the cases in the Intensive Care Units (ICUs), but again not in all countries. The number of severe cases accommodated in Intensive Care Units (ICU) seems to be the easiest to categorize of all the data listed. We denote these data as $ICU\,(t)$.

A main point of the modeling paradigm for COVID-19 (and similar virus infections) is that, for a certain segment of the society (in this case, the younger people), the infection symptoms do not differ essentially from a seasonal flu, hence the number of unreported cases (those which are in compartment C_I but not in $Idata\,(t)$ for every time t) may be much bigger, thus in the above inequality (1.1) more appropriate is to use the symbol "$\ll$", which denotes roughly speaking "much less". In the case of the seasonal flu it may be even 100 times less.

One has to know that the large number of quarantined at home, many mild cases (also asymptomatic) do not reach the hospitals, and are not duly reported. On Figure 1.5, we provide an example of a typical list of data for Bulgaria in the period *29.11.2020-13.12.2020*: with red we have indicated the columns with the **Infected** (the total number of infectious persons until *Today*), **ICU_critical** (the current number of people on ICUs *Today*), **Recovered** (the total number of recovered until *Today*), and **Deaths** (the total number of fatalities until *Today*).

date	tests_tota	tests_daily	infected	active_cases	new_infected	hospitalized	ICU_critical	recovered	recovered	deaths	deaths_da
2020-11-29	964461	4928	141747	90219	1792	6830	431	47779	1039	3749	69
2020-11-30	966426	1965	142486	90078	739	6869	430	48594	815	3814	65
2020-12-01	973251	6825	145300	90700	2814	6783	457	50565	1971	4035	221
2020-12-02	982307	9056	148775	91587	3475	6635	493	53000	2435	4188	153
2020-12-03	990785	8478	151913	92360	3138	6635	523	55206	2206	4347	159
2020-12-04	999867	9082	155193	93549	3280	6766	523	57141	1935	4503	156
2020-12-05	1009397	9530	158807	94480	3614	6744	504	59677	2536	4650	147
2020-12-06	1015383	5986	160844	95442	2037	6959	516	60673	996	4729	79
2020-12-07	1016807	1424	161421	94378	577	7000	516	62246	1573	4797	68
2020-12-08	1023067	6260	164185	93559	2764	6821	523	65616	3370	5010	213
2020-12-09	1032782	9715	168165	93981	3980	6839	514	69028	3412	5156	146
2020-12-10	1041360	8578	171493	94132	3328	6998	544	72078	3050	5283	127
2020-12-11	1049830	8470	174568	93931	3075	7084	542	75232	3154	5405	122
2020-12-12	1058417	8587	177665	92581	3097	7151	588	79522	4290	5562	157
2020-12-13	1062824	4407	178952	91569	1287	7224	595	81757	2235	5626	64

Figure 1.5: Example of Bulgarian dataset.

1.4 Section: Data sources for COVID-19, their reliability and quality

One of the most important problems of analyzing the data for COVID-19 is to find reliable source of the data. Although there is a lot of web-sites which pretend to offer data, very often their quality is questionable; also, many of them are not sufficiently easy to access.

1. A very good example is the Bulgarian National Portal for COVID-19 data, which is called Open Data BG, which provides access to Bulgarian public data in open and machine-readable format, `https://data.egov.bg/data/resourceView/e59f95dd-afde-43af-83c8-ea2916badd19`. The official National COVID-19 Portal of Bulgaria: `https://coronavirus.bg/bg/statistika` provides only current data in graphical visualization, and no digital format.

2. The Wikipedia page contains a very comprehensive information about the real state of the development of the COVID-19 pandemic in Bulgaria.

3. It contains the official data announced by the Ministry of Health of Bulgaria, and promises Open Data formats and access. However, it is not directly accessible but one needs a special code to access the

source data – one needs to write a special code and one needs to get acquainted with the REST-API interface, which contains a special key. They do not have various types of file formats but only JSON, and one needs a special script to convert it to more usual formats as CSV. The query to access the data is uselessly complicated and requires POST method instead of GET method. Hence, the Open access, which is widely advertised, is practically missing.

4. This explains why European and World portals contain very often mismatched and missing additional (e.g. the ICU data) Bulgarian data. One of the very few exceptions is the Worldometer.info site which obviously have succeeded to struggle with the peculiarities not only of the Bulgarian portal but with similar national portals of more than 221 countries. However they provide a paid service if you want to embed their site in another site. Also, they refer to the so-called Projections Tool which is a project of the IHME institute.

5. Many European portals did not manage to get the data from the above portal, in particular the *European Centre for Disease Prevention and Control,* `https://www.ecdc.europa.eu/en`. Apparently, they are working with synthetic data obtained via models, which represents often a very big mismatch with the official data provided in the site `https://data.egov.bg/data/resourceView/e59f95dd-afde-43af-83c8-ea2916badd19`.

6. Another very reliable source of data for the whole world (similar to the Worldometer but with really open source access) is available at `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`, which is *Novel Coronavirus (COVID-19) Cases Data* which is compiled by *Johns Hopkins University.* The data is compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHU CCSE) from various sources including the World Health Organization (WHO), DXY.cn, BNO News, National Health Commission of the People's Republic of China (NHC), China CDC (CCDC), Hong Kong Department of Health, Macau Government, Taiwan CDC, US CDC, Government of Canada, Australia Government Department of Health, European Centre for Disease Prevention and Control (ECDC), Ministry of Health Singapore (MOH), and others. JHU CCSE maintains the data on the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository on Github.

   Fields available in the data include Province/State, Country/Region, Last Update, Confirmed, Suspected, Recovered, Deaths.

7. It is worth mentioning also other sources:

   COVID-19 Maps & visuals

Coronavirus COVID-19 global cases (**Johns Hopkins**)

COVID-19 event risk assessment planning tool (**Georgia** Tech)

US spread of COVID-19 maps and analytics (**SharedGe0**)

Novel coronavirus (COVID-19) outbreak timeline map (**HealthMap**)

Novel coronavirus infection map (University of **Washington**)

COVID-19 surveillance dashboard (University of **Virginia**)

Novel coronavirus (COVID-19) situation dashboard (**WHO**)

Coronavirus disease 2019 (COVID-19) in the US (**CDC**)

Geographical distribution of COVID-19 cases worldwide (**ECDC**)

COVID-19 coronavirus tracker (**Kaiser** Family Foundation)

COVID-19 coronavirus outbreak (**Worldometer**)

The COVID tracking project (**COVID Tracking**)

Data visualization (**CDC**)

Coronavirus: the disease COVID-19 explained (**South China Morning** Post)

Mapping the novel coronavirus pandemic (**Esri** StoryMaps)

8. Last but not least we mention the Kaggle and Ourworlddata web-sites devoted to the Coronavirus.

9. **Data on COVID-19 (coronavirus) by the project "Our World in Data":** The datasets provided by the project are available in the basic formats: CSV, XLSX, JSON. The complete COVID-19 dataset is a collection of the COVID-19 data maintained by Our World in Data. They update it daily throughout the duration of the COVID-19 pandemic. It includes a dataset with the following characteristics provided in Figure 1.6.

   The complete Our World in Data COVID-19 dataset is available at the link `https://github.com/owid/covid-19-data/tree/master/public/data`

1.4.1 State–of–the–art models with Projections and Analytical Tools

We provide a list of some of the most popular world wide instruments for analysis of the COVID-19 spread, supplied with Web-based (free) online Analytical Tools. We provide some comments.

They are as follows:

Metrics	Source	Updated	Countries
Vaccinations	Official data collated by the Our World in Data team	Daily	217
Tests & positivity	Official data collated by the Our World in Data team	Weekly	136
Hospital & ICU	Official data collated by the Our World in Data team	Weekly	35
Confirmed cases	JHU CSSE COVID-19 Data	Daily	194
Confirmed deaths	JHU CSSE COVID-19 Data	Daily	194
Reproduction rate	Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C	Daily	185
Policy responses	Oxford COVID-19 Government Response Tracker	Daily	186
Other variables of interest	International organizations (UN, World Bank, OECD, IHME...)	Fixed	240

Figure 1.6: Table of Our World in Data

1. The tool "Projections" by the IHME Institute is available at the link `https://covid19.healthdata.org/bulgaria?view=total-deat hstab=trend`. It is embedded in the Worldometers.info web-site. It does not model properly the vaccinations, however it is claimed that they use "vaccination distributions" which is not clear neither from the graphs nor from the academic publications to which they refer. In principle, their models use too many parameters. They do not show the real (observed) data but some smoothed data, in particular for the infectious and fatalities curves.

2. Tool Delphi from the Massachussets Institute of Technology, `https: //www.covidanalytics.io/home`

3. Tool at Imperial College London, `https://www.covidsim.org`

4. The Los Alamos National Laboratory (LANL), `https://covid-19.b svgateway.org/`

5. the SI-KJalpha model from the University of Southern California, `ht tps://github.com/scc-usc/ReCOVER-COVID-19)`

Daily deaths from other modeling groups are smoothed to remove inconsistencies with rounding. Regional values are aggregates from availble locations in that region.

In Figure 1.7, (from the site https://covid19.healthdata.org/bulgaria?view=total-deaths&tab=trend) we show projections of daily COVID-19 deaths from other modeling groups which we mention below.

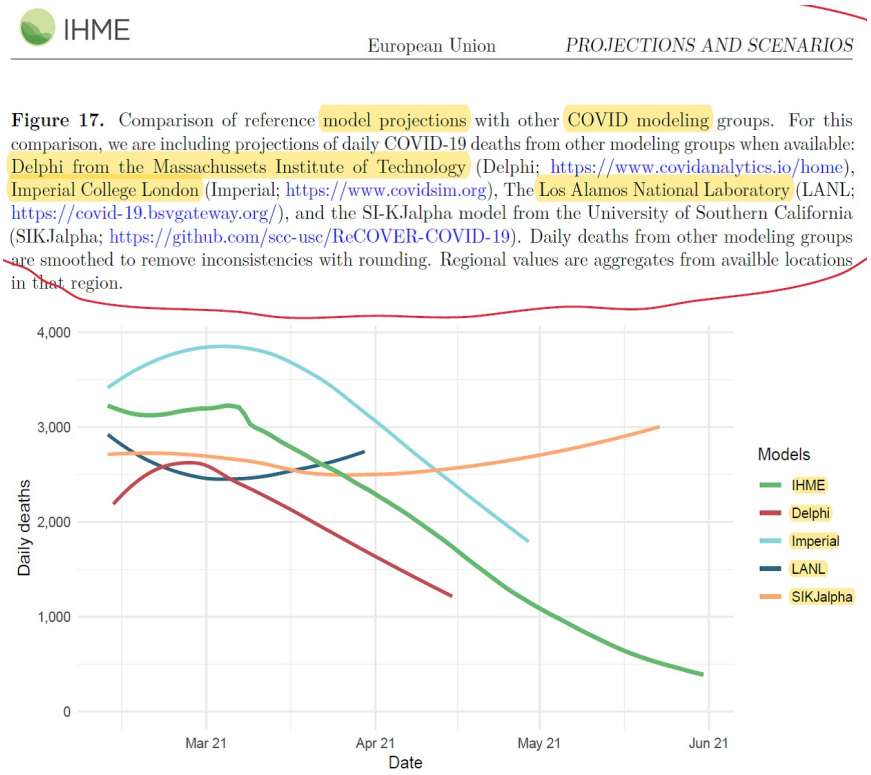


Figure 1.7: Projections of daily COVID-19 deaths by different modeling groups.

1.5 Multidisciplinarity of the Epidemiological Models for COVID-19

Multidisciplinarity is very important for the success of solving the nowaday challenges in modern Epidemiology. To cite an highly reputable source, Nature, Editorial [2021], "Theorists and experimentalists must join forces". Namely, epidemiologists and disease modeling experts have been working together to build **mathematical models** and run simulations to better understand how SARS-CoV-2 impacts populations. This, in turn, has helped governments with devising policies and *non-pharmaceutical* interventions to help slow the spread of the virus1. *Physics*- and *machine learning*-based models have also been used by the research community to study the virus and to find potential *drug-related* solutions to the disease, as described in the issue of the same journal.

In order to get an idea about the components of the epidemiological models one needs to take into account different factors.

1. First of all, the **demographic** factors play and important role in solving the real problems of Epidemiology – in particular, for understanding the details of the COVID-19 spread in Bulgaria, one needs to take into account that in Bulgaria 41% of the population live **three generations** under the same roof. Similar is the situation in other Mediterranean countries, as Greece, Italy, Spain. One may find more information about the demographic status at the link of EUROSTAT, "Is your home too crowded? ": `https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210105-1`. About the current state of the living conditions in Bulgaria and other countries, see Figure 1.8 and Figure 1.9 [3]

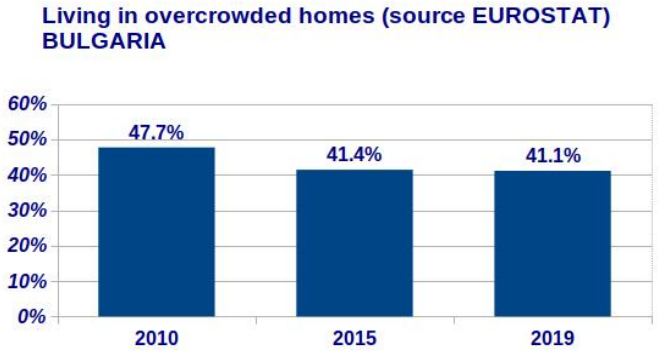


Figure 1.8: Percentage of population in Bulgaria living in overpopulated homes.

In the case of the COVID-19 pandemic this means that the school vacations play important role for the regulation of the epidemic spread. Indeed, if the pupils visit school then they become "superspreaders" (due to their intensive contacts with their classmates) and coming home they spread the disease to the oldest generation, their grandparents which are beyond 60 year old. The picture is completely different in the countries where no more than two generations live together – Germany, UK, France, Switzerland, Scandinavian countries, etc. In these countries the closure of the schools do not play that big role for the containment of the spread of COVID-19. Hence, the effectiveness of the Containment measures depends strongly on the so-called "communicability index" within the nation.

2. The **climate/weather** conditions seem to be important for the so-called "seasonal effect". The specific **climate**/weather conditions in a

[3]The Figure is available in the publication of national TV channel BTV at the link `https://btvnovinite.bg/bulgaria/pod-edin-pokriv-pokolenija-balgari-zhivejat-zaedno.html`.

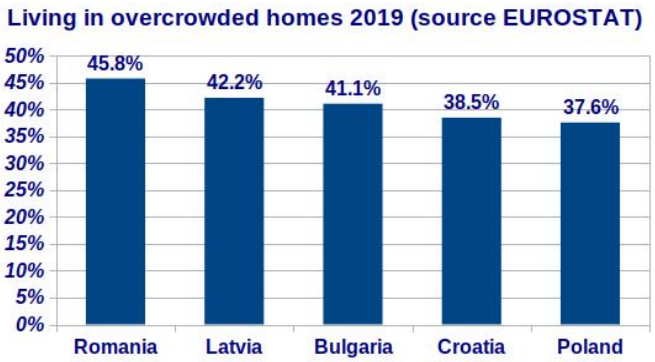


Figure 1.9: Percentage of population in several countries in EU, living in overpopulated homes.

particular year play also important role for the spread of the disease. Since the Flus and the Coronavirus diseases are a ,,cold disease“ (one becomes sick easier if heshe catches cold) the weather conditions in the particular year reduce or increase the disease spread. Apparently, this has a strong impact on the seasonal effect – sometimes it starts earlier in the spring at the same geographical latitude as Bulgaria.

3. In order to understand what is the role of the seasonal effects, it seems to be also important to measure the "variation" of the epidemiological curves for the Flus and for the Coronavirus during the year, in our latitutes and in the India latitude.

1.6 Seasonal effect of Covid-19 spread - the big challenge

One of the main challenges for the present research was the design of predictive models for the seasonal effect of COVID-19 spread in the summer of 2021 and 2022. The main issue with the modeling of the seasonal effect in Bulgaria is the lack of unpolluted data for the summer of 2020 - the huge influx of Bulgarian guest workers have completely spoiled the curve of infectious cases with COVID-19. The only possibility was to use "Big data" from other countries for an extrapolation, or in other words, "to teach the models". Another possibility, also related to Big data paradigm would be to use data about the Flu season (which is the winter season in Bulgaria and the majority of the European countries, however not in India). Again the lack of user-friendly access to open data is the problem. Due to the small number of vaccinations until March 2021, we could use data not influenced by the vaccines, although a Lockdown was introduced already on Novem-

ber 28, 2020. This Lockdown was raised gradually: on February 1, 2021 the schools and kindergardens have opened; from March 1 the public facilities (restaurants, bars) were opened.

1.7 Seasonal effect: the Indian seasonality phenomenon

The unusual peak in the infectious curve in April-May 2021 in India may be explained only if one deepens in the understanding of the seasonal effect – it is not just summer and winter time - it is strongly dependent on the Flu season in every particular country. One also needs to apply the hypothesis that the Flu seasons coincide with the Coronavirus seasons. We will see below that the Flu (peak) season in the majority of India's states is in the range of June-September, see Figure 1.10. Though physically located in northern hemisphere, India has distinct seasonality that might be related to latitude and environmental factors. Thus, the unusual spike in the infectious curve in April-May 2021 was just at the end of the Flu low season, and was explained by the drastic neglecting of social distancing measures during election meetings, religious gatherings, celebations, etc. We cite the Wikipedia (`https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India`) about the multiple factors that caused it:

1. highly-infectious variants of concern such as Lineage B.1.617
2. a lack of preparations as temporary hospitals were often dismantled after cases started to decline, and new facilities were not built
3. health and safety precautions being poorly-implemented or enforced during weddings, festivals (such as Holi on 29 March, and the Haridwar Kumbh Mela in April), sporting events (such as IPL)
4. state and local elections in several states, and in public places
5. an economic slowdown put pressure on the government to lift restrictions
6. there had been a feeling of exceptionalism based on the hope that India's young population and childhood immunisation scheme would blunt the impact of the virus
7. Last but not least, models may have underestimated projected cases and deaths due to the under-reporting of cases in the country

We have two references, Koul et al. [2014], Chadha and Potdar [2015], where the peak and the low seasons of the Flu infection for several major states in India are well described. We provide the main figure from

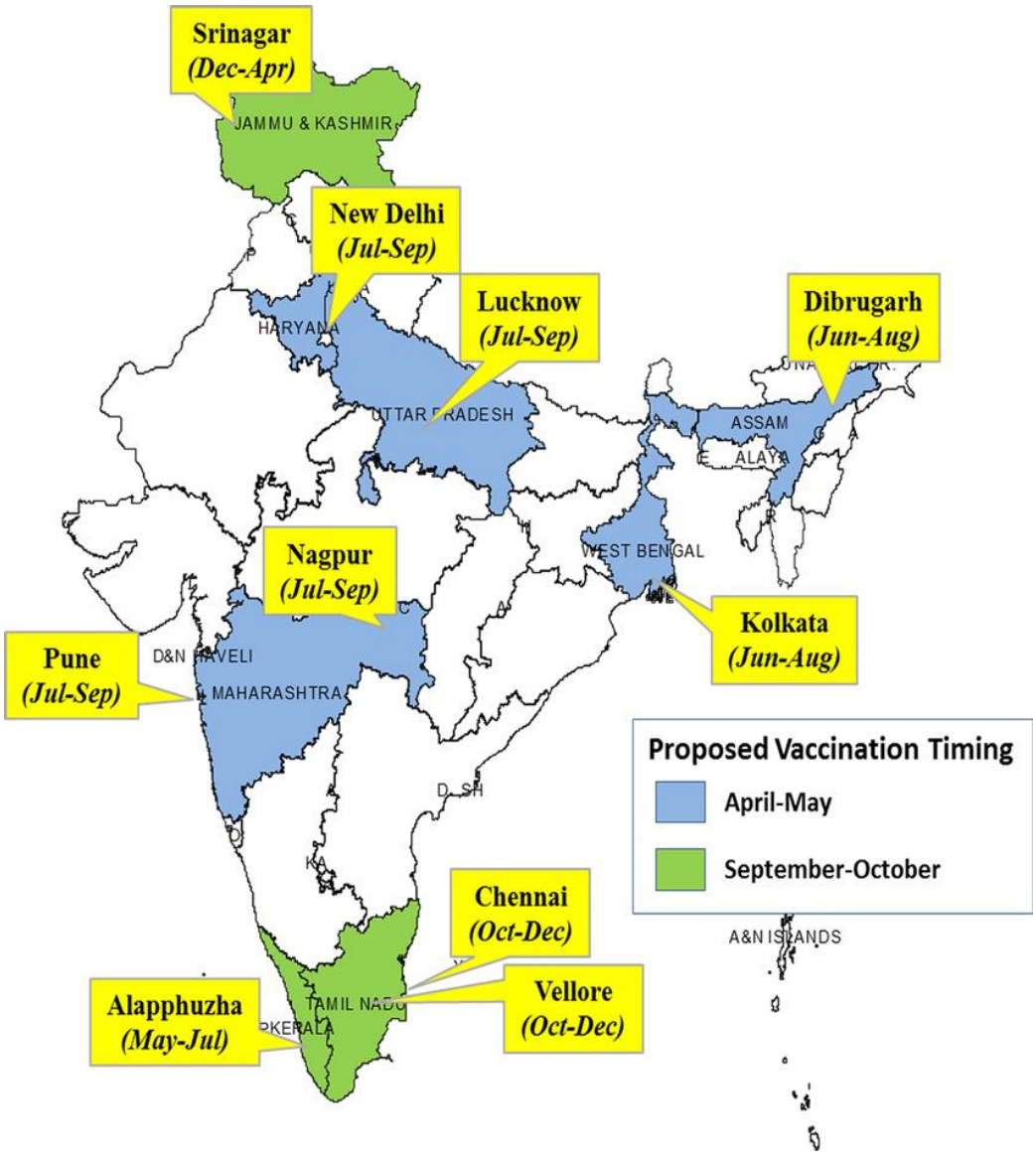


Figure 1.10: Peak seasons of the Flu for some states in India.

Chadha and Potdar [2015], see Figure 1.10. One sees how unusual the peak seasons are, although these states are all in the Northern hemisphere. Especially paradoxical seems the case of the very south locations Allapphuzha and Chennai which are on the same latitude and are quite close, but have complete different peak seasons.

It shows that the big infection in April-May 2021 is during the low season

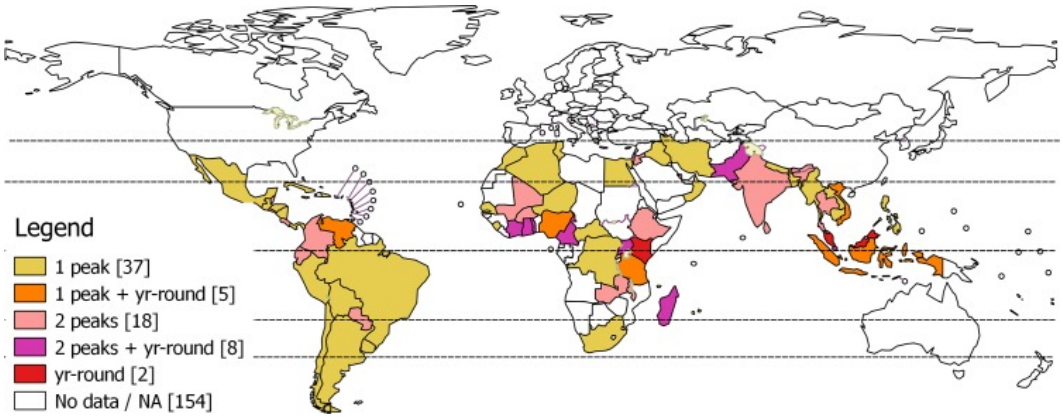


Figure 1.11: Peak seasons of the Flu for some countries.

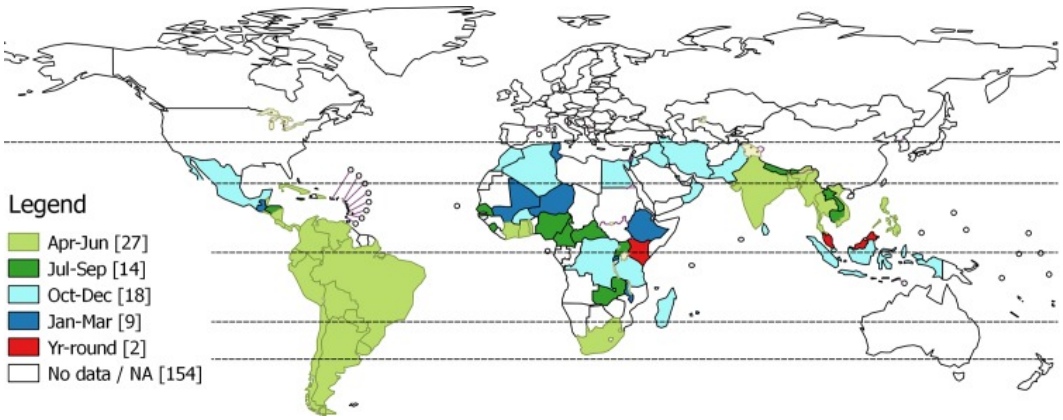


Figure 1.12: Proposed vaccination timing.

in the majority of the states. In view of the wide similarities between the ways of spread, one may conjecture that the peak seasons of the Flu and the Covid-19 are very similar. This is the basis for modeling of the seasonal effect of Covid-19 spread.

It is very instructive to see how different are the Flu peak seasons in the whole world, Hirve et al. [2016], see Figure 1.11 and Figure 1.12:

However, in order to be able to draw some conclusions, it is important to know the precise form of the infection curves of the Flu disease, and in particular their "infectivity variation around the mean", i.e. what is the ratio of the average number of infections during the low season and the same average during the peak season. We expect that this "infectivity

variation" is much stronger in the countries with stronger climatic variations (in the subtropics), and not that strong in the tropical areas (as India) where the climatic variations are not that strong. To make more precise conclusions one needs to carry out a detailed study of the infection curves for the seasonal Flu at the locations of interest.

1.8 Data Acknowledgements

We acknowledge the following data sources used in the present research of COVID-19:

1. **HDX Humanitarian Data Exchange V1.39.3**
2. time_series_covid19_confirmed_global.csv
3. time_series_covid19_deaths_global.csv
4. time_series_covid19_recovered_global.csv
5. `https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases`
6. Kaggle SRK Sudalairaj Kumar, `https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#covid_19_data.csv`
7. European Centre for Disease Prevention and Control: ICUs and Vaccinations
   https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-COVID-19,
   `https://ourworldindata.org/covid-hospitalizations`

1.9 Acknowledgements

We acknowledge the support by grants KP-06-N32-8 and KP-06N42-2 of Bulgarian NSF.

We acknowledge the support of a fast-track grant by the European Open Science Cloud (EOSC) in the framework of the EOSC-Secretariat.

Chapter 2

TVBG-SEIR – the short-term SEIR model based on splines, and a tool for COVID-19 prediction scenarios

In the present chapter we develop a novel TVBG-SEIR spline model for analysis of the coronavirus infection (COVID-19), aimed at analyzing the short-term global evolution of the epidemics controlled by the introduction of lockdown/open up measures by the authorities. The incorporation of different lockdown prediction scenarios varying in time permits to analyze not only the primary epidemic wave but also the arising secondary wave and any further waves. Let us note that what we call *"prediction scenarios"* is nowadays called very often *projections*.

The model is supplied by a web-based *Scenario Building Tool for COVID-19* (called shortly *SBT-COVID-19*) which may be used as a decision support software by (health) policy makers to explore various scenarios, by controlling/changing the scale of the containment measures (home and social isolation/quarantine, travel restrictions and other) and to assess their effectiveness. In particular, the SBT-COVID-19 Tool permits to assess how long the lockdown measures should be maintained.

The *SBT-COVID-19* tool is based on the Jupyter Notebooks Architecture: Jupyter Notebooks work with a two-process model based on a kernel-client infrastructure. This model applies a similar concept to the Read-Evaluate-Print Loop (REPL) programming environment that takes a single user's inputs, evaluates them, and returns the result to the user. Based on the two-process model concept, the main components of Jupyter may be visualized in the following way:

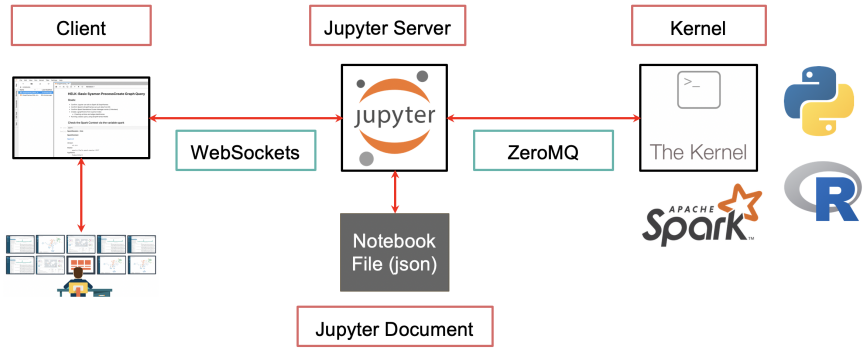


Figure 2.1: Jupyter Software Architecture

2.1 Introduction

The main point of our approach is the proper way to incorporate in our models the *a priori* known information about the *Containment measures*. The models which we present here did not come immediately in the best form, and they had to undergo a certain evolution:

1. When we first started to develop the TVBG-SEIR model we were sure that we have to use the dates on which Containment measures have been introduced by the Health authorities **as knots**, even the same for the two splines which represent the coefficients β and γ. This concept is also the one followed by the authors of other models, say published in Science, Dehning [2020].

2. Also, to keep it simple we have decided to train our models on real data where we have **no more than two** switches in the Containment measures – the two switches may be both strengthenings of the measures, or one of them may be relaxation. Hence, the most reasonable approach seems to be, to consider only splines for representing the parameters $\beta(t)$ and $\gamma\,(t)$ having just two break points in the interval of the historical data.

3. However, after we have fitted a lot of models to real data, we have realized that the knots that have resulted after the fitting may be quite far from the dates when the Containment measures have been introduced. Then we decided to let the two knots of the splines free in the Optimization process. This resulted in a quite computation consuming algorithms but there is no other way...

4. Perhaps, the most reasonable approach is to consider the dates of the two Containment measures as **initial guess** for the Optimization procedure. Then we will find some solutions for the global Optimization

which have close resulting RMS (root mean square). If there is no big discrepancy with the observed (official) data, we will take as satisfactory the solution with knots which are closest to the dates of the two Containment measures.

5. Not less important is to incorporate in a proper way the Monotonicity constraint as a Bayesian prior: we consider most naturally, in the case of two switching dates with strengthening measures a monotone decreasing $\beta(t)$ but monotone increasing $\gamma(t)$.

2.1.1 Context

The present model is designed with the main purpose to provide short-term planning of the containment measures. In particular, it is essential to assess in the short term, how the expensive, resource-intensive measures implemented by the authorities, as home and social isolation/quarantine, travel restrictions, etc., can contribute to the prevention and control of the COVID-19 infection, and how long they should be maintained.

However the classical SIR/SEIR models have been primarily studied in the case where the main parameters - the *transmission rate* β (reflecting the virus spread by infected individuals) and the *removed (removal) rate* γ (reflecting the hospitalization/isolation measures) - remain constant during the whole period of interest. This does not reflect in a proper way the extremely dynamic behavior of such measures during the COVID-19 and similar epidemics, resulting from the imposition of intensive containment measures by the authorities.

2.1.2 Aims and Methods Summary

It is important to extend the classical SIR/SEIR models by creating new models for the dynamics of the transmission rates $\beta(t)$ (sometimes referred to as **Beta**) and removed rates $\gamma(t)$ (sometimes referred to as **Gamma**). The main aim of the present research is to introduce a novel spline-based SEIR model with time-varying $\beta(t), \gamma(t)$ parameters, or abbreviated ***TVBG-SEIR*** model, which is used to estimate the practical implications of the public health interventions and containment measures. We have designed a **Scenario Building Tool for COVID-19 (SBT-COVID-19 Tool)** based on the TVBG-SEIR model, which may be used as a Decision Support Tool to assist the health decision- and policy-makers in creating *predictive scenarios* (*projections*). It may be used to assess the impact of previous public health interventions, and to plan quantitatively and qualitatively the introduction of future containment measures for achieving the necessary objectives.

For formulating our model, we use deterministic spline Ansatz: the transmission rates $\beta(t)$ and the removal rates $\gamma(t)$ are modeled by splines with

two nodes - *Node1, Node2* (the same nodes for both $\beta(t)$ and $\gamma(t)$) - within the time interval of interest – from *StartDate* until *TodayDate*. This Ansatz allows to properly model the dynamics due to the introduction of containment measures by the authorities in **two steps**. The purpose of fitting of the TVBG-SEIR model is to identify the nodes of the splines and the three values of $\beta(t)$ and $\gamma(t)$ on the intervals [*StartDate,Node1*] , [*Node1,Node2*] and [*Node2,TodayDate*]. It is assumed that $\beta(t)$ and $\gamma(t)$ are **constant** in the time interval *[Startdate, Node1],* and $\beta(t)$ is monotone **decreasing** while $\gamma(t)$ is monotone **increasing** function. There may be different data to which the TVBG-SEIR model may be fitted. In every situation it is preferable to choose those data which are the most reliable. In particular, we will fit the TVBG-SEIR model (with time-varying $\beta(t)$ and $\gamma(t)$) simultaneously to two sets of data: the **daily infected cases** (or their **cumulative** vector), and to the **removed** cases (which are all removed cases until a certain date, obtained as cumulative to the daily removed cases). Other possibility is to fit simultaneously the data to the number of daily reported cases in the Intensive Care Units (ICUs). (This approach is also used by some authors.)

The choice of just two nodes of the splines for the rates $\beta(t)$ and $\gamma(t)$ seems to be appropriate for models of historical data (until *Today*), but for not very long periods of time. These models are used as a basis for creation of "prediction scenarios" (projections) starting from *Today*, with a prediction perspective of about two months (**2m**) horizon. These models are updated every day (by the arrival of the official daily data), and the scenarios (projections) are renewed accordingly. In a more mathematical language "predictive scenario" (called very often today **projection**[1]) means choice of control parameters $\beta(t)$ and $\gamma(t)$ in the form of splines defined after *Today,* which determine the SEIR model to be defined in detail below. One may choose these scenarios in infinitely many ways. The main objective of our approach is to choose such scenarios $\beta(t)$ and $\gamma(t)$ for which the SEIR model generates curves which satisfy some reasonable restrictions, e.g. the number of infected daily cases does not explode too abruptly. Let us emphasize that these are just "possible prediction scenarios" but not extrapolations of the historical (observed) data in the classical sense of the word.

The web-based SBT-COVID-19 Tool was designed for visualization of the results of the fitted model (the daily infected cases), and for creating prediction scenarios (projections) for the daily infected cases during the next two month horizon, by controlling the future values of the coefficients $\beta(t)$ and $\gamma(t)$. It is described in detail in Section 2.8, and is available at the links:

[1]Projection is the process of moving forward in time through the imagining of future events, or by means of estimates based on certain assumptions or past trends, see `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Projection`.

The term projection is obviously synomimous to "prediction scenario" and we will use both, especially if comparison is needed with the recent literature.

Version 3 (with Matlab generated Figures)
`http://213.191.194.141:8888/notebooks/TVBG-SEIR-Spline-model_v3.ipynb`
Version 4 (with Python generated Figures)
`http://213.191.194.141:8889/notebooks/TVBG-SEIR-Spline-model_v4.ipynb`

2.1.3 Objectives

By analogy with the usual *seasonal flu* the main parameters of the spread of the viruses are the transmission rate β which reflects the rate of the transmission of the virus from infected people to susceptibles, the removed rate γ (which is the sum of "recovery to health cases" + "isolated sick cases" + "mortality due to the sickness cases"), and the parameter σ which is the reciprocal to the incubation period. Unlike the usual flu, Covid-19 has a long **incubation period** (at present, estimated to have a mean value of 5.9 days, Mcaloon et al. [2020]. Daley et al. [2020]). Due to the large number of asymptomatic or mild-symptomatic cases, COVID-19 has proved to be very insidious and requires intensive emergency measures from the authorities to reduce the transmission rate β and to increase the recovery rate γ. For comparison, in the case of the seasonal flu no intensive containment measures are necessary to be undertaken by the authorities, but just the usual two-week winter school vacations have proved to be sufficient to stop the epidemic.

For containing COVID-19, the authorities have introduced very strong measures which have essentially influenced the dynamics of the parameters β and γ. For the majority of states these measures have been introduced not only in one step but most often in two steps. It depends on every society how fast these measures will be implemented in life. There are two types of measures: for example, closing schools, pubs, restaurants, traveling national or international routes, social meetings, wearing masks, reduce directly the transmission rate of the disease β (further we will call sometimes these measures shortly **Beta** measures); on the other hand the fast identification and medication of virus spreaders, hospitalization, quarantining and similar, increase the rate γ of removal from the group of virus spreaders (further we will call sometimes these shortly **Gamma** measures). It is important to assess how these expensive and resource intensive measures implemented by the authorities can contribute to the prevention and control of the COVID-19 infection, and how long they should be maintained, Tang et al. [2020], Tang et al. [2019].

In order to meet the challenge of Controlled spread of the COVID-19 (and similar) epidemics, one needs to develop new mathematical models which better describe reality. Based on the widely used conventional epidemiological model SEIR, in the present research we propose a new model

TVBG-SEIR which incorporates a specific spline model for the time-varying transmission β and removal γ rates.

The present chapter is organized as follows: In Section 2.2 we recall the deterministic SEIR model and introduce some notions and notations. In Section 2.3 we introduce the discretization of the SEIR model which is used in the algorithms. In Section 2.4 we introduce and provide all technical details of the TVBG-SEIR spline model. In Section 2.5 we provide an application of the TVBG-SEIR model to Bulgarian data, which are used to illustrate the work of the **SBT-COVID-19 Tool** for prediction scenarios. In Section 2.6 and Section 2.7 we provide more examples of analysis, for the Italian data, and for the German data. In Section 2.8 we describe the technical details of the **SBT-COVID-19 Tool** for prediction scenarios (projections). In Section 2.9 we provide some recent references about models with time-varying transmission rates and their fitting to the data (calibration).

2.2 The classical deterministic SEIR model: Notions and Notations

We will introduce the classical deterministic SEIR model by providing all notions and notations as fully as possible, to enable the replicability of the calculations and experiments in the present research. Thus we also provide a detailed account of the *discretization* of the well-known classical SEIR model.

In section 1.2 we have introduced the Compartmental models with all necessary notions and notations.

In section 1.3 we have provided the list of official (observed) data available in Internet via the majority of the popular links, some of which we have provided in section 1.4.

2.2.1 Definition of the classical continuous SEIR model

The main point of developing the compartmental deterministic SEIR model is to provide some tractable approximations $S(t), E(t), I(t), R(t)$ to the time series of the "reality data" $Sr(t), Er(t), Ir(t), Rr(t)$ explained in section 1.2. The most widely used is the model based on a system of Ordinary Differential Equations with variables $S(t), E(t), I(t), R(t)$ which is given as follows:

$$S'(t) = -\frac{\beta(t) S(t) I(t)}{N} \tag{2.1}$$

$$E'(t) = \frac{\beta(t) S(t) I(t)}{N} - \sigma E(t) \tag{2.2}$$

$$I'(t) = \sigma E(t) - \gamma(t) I(t) \tag{2.3}$$

$$R'(t) = \gamma(t) I(t) \tag{2.4}$$

Let us explain the notations and the correspondence to the reality data of the Compartmental model:

1. Here the term $\beta(t) I(t)/N$ expresses the rate at which new individuals (as a proportion of the total population size) are infected by the already infectious $I(t)$ individuals, (cf. Keeling and Rohani [2008], p.18). Here and further $\beta(t)$ is called Transmission rate of the infection, which we call further simply Beta.

2. As already said, the coefficient $\gamma(t)$ is the Removal rate; it is determined by the reciprocal of the infectious period, after which either the person is *recovered* (and no more infectious) or *dead* (again, no more infectious). Here and further $\gamma(t)$ is called Removal rate, and sometimes we call it simply Gamma.

3. The coefficient σ is the latent rate, or the rate of "becoming symptomatic" (where $\frac{1}{\sigma}$ is the average of the **incubation period**). In the present paper we use the constant value

$$\sigma = \frac{1}{5.9}$$

which represents a reasonable approximation, as the recent research shows, Mcaloon et al. [2020]. (Let us note that previously the rate was $\frac{1}{5.9}$, see Lauer et al. [2020].)

4. The curve $S(t)$ corresponds to the reality data time series $Sr(t)$. The quantity $\sigma E(t)$ is equal to the daily new infectious cases $Idata(t)$. However, the curve $I(t)$ of the SEIR model is equal to the so-called Active Cases which are defined by the equation

$$AC(t) = Total\_Infected(t) - Rr(t)$$

Here $Total\_Infected(t)$ is the **cumulative** sum of $Idata(t)$ until the date t.

The usual applications of the SEIR model are with constant rates $\beta(t)$ and $\gamma(t)$. One assumes that the initial values $S(0), E(0), I(0)$, and $R(0)$ are given and the system is solved for the times $t \geq 0$, where t is an integer. Obviously, we have the equation

$$S'(t) + E'(t) + I'(t) + R'(t) = 0 \qquad \text{for } t \geq 0,$$

which implies that the sum

$$S(t) + E(t) + I(t) + R(t)$$

is a constant for every $t \geq 0$. It is assumed that the following equation holds

$$\begin{aligned} N &= S(0) + E(0) + I(0) + R(0) \\ &= S(t) + E(t) + I(t) + R(t) \end{aligned} \tag{2.5}$$

where N is the total population in the country of interest XX.

Let us remind that one of the most important properties of the SEIR model is that it provides a direct way to express the basic Reproduction ratio:

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \tag{2.6}$$

2.3 Discretization of the SEIR model

In practice one uses a discretization of the continuous SEIR model. In the current work we use the following discretization of the SEIR model which is derived from the Euler method for approximate solution of the initial value problem (2.1)-(2.4):

$$S_{n+1} = S_n - \frac{\beta_n S_n I_n}{N} \tag{2.7}$$

$$E_{n+1} = E_n + \frac{\beta_n S_n I_n}{N} - \sigma E_n \tag{2.8}$$

$$I_{n+1} = I_n + \sigma E_n - \gamma_n I_n \tag{2.9}$$

$$R_{n+1} = R_n + \gamma_n I_n \tag{2.10}$$

Here S_n, E_n, I_n and R_n are respectively the values of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ on the day $t = n$, and the initial values for day $n = 0$ are S_0, E_0, I_0 and R_0. The above system is iteratively solved for all integers $n \geq 0$. We assume that the size N of the population remains unchanged (hence no usual birth and mortality are taken into account). As in the continuous case, the total sum of the above is assumed to satisfy

$$N = S_n + E_n + I_n + R_n \tag{2.11}$$

which makes one of the equations in (2.7)-(2.10) redundant.

It is well known that the above Euler method for approximating the solution of (2.1)-(2.4) is less accurate than the Runge-Kutta which is widely used, see e.g. Stoer and Bulirsch [2002].

Again, it is very important for the modeling process to realize what is the correspondence between the variables of the discrete model and the officially announced data: On the day n the value R_n corresponds to the sum of the cumulative recovered plus fatalities data, i.e. to $Rdata(n) + Deaths(n)$. The announced observed daily data of newly infected $Idata(n)$ correspond to the amount σE_n which is clear from equation (2.9). Below we use this correspondence to define the quadratic function $F(\Theta)$ for fitting our models.

We have to make an important remark about the notations used in the present text. It was clearly explained in [Kounchev et al., 2021b, see Bibliography to chapter 3] that in the literature people use very often misleading notations by mixing four different settings:

1. the continuous setting of SEIR model, (2.1–2.4)
2. the discrete setting of SEIR model, (2.7)-(2.10)
3. the official data provided in section 1.3
4. the real situation best described by the Compartmental framework described in section 1.2, and which is meant to be approximated by the SEIR model.
5. In particular, misleading is the usage of the curve $I(t)$ of the SEIR model which corresponds to the Active Cases in the empirical data. Things will become clearer after we discuss the correspondence of the variables in the SEIR model with the other data in section 3.5.

Remark 2 *Let us remark that the continuous model (2.1–2.4) and the above discrete approximation (2.7)-(2.10) have essential differences in the long-term behavior which has been the subject of much research. It is important to note that the qualitative properties of the solution to the differential equation and of the discrete equation differ essentially - the continuous case is simpler as usual.*

2.4 The TVBG-SEIR model defined

The SEIR models have proved to be very efficient in situations where the main parameters β and γ are constants, in natural conditions, where no special control by the authorities is exercised, i.e. no intervention measures are undertaken to change the transmission and the removal rates in the course of the epidemics. This is very often the case with the seasonal flus where the medical authorities do not undertake actively special measures to restrict the social behavior of the citizens, although nowadays the vaccinations change the natural picture. However due to the specific of the COVID-19 the situation has become more dramatic and it has required the interference of the governments in order to avoid the overloading of the National Health systems. The authorities have introduced very strong restrictive measures which have essentially influenced the dynamics of the parameters β, γ. For the majority of the states these measures have been introduced not only in one step but most often at least in *two steps*.

In view of the above it makes sense to seek for mathematical models which try to model as best as possible the dynamics of the parameters β

and γ. We have decided for spline structure with two important breakpoint nodes, *Node1, Node2* - which reflect the control exercised by the authorities in the form of two consecutive restriction measures. Also, it is natural to assume that between the dates the control measures change the parameters $\beta(t)$ and $\gamma(t)$ in a monotone way, i.e. $\beta(t)$ is decreasing whereas $\gamma(t)$ is increasing.

2.4.1 Technical Description of the TVBG-SEIR model

1. We denote the *StartDate* by T_1; this corresponds to a date when the first cases of COVID-19 are announced, eventually we may choose T_1 to be a date when the steeper growth of the epidemic starts. We denote by T_4 the *EndDate* (usually chosen to be Today).

2. We choose two interior nodes in the interval $[T_1,\ T_4]$ for the interpolation splines modeling the coefficients $\beta(t), \gamma(t)$: *Node1* $= T_2$ and *Node2* $= T_3$. This corresponds to two steps of the introduction of restrictive measures imposed by the authorities of the country XX. Normally, the date T_2 may be the First restrictive measures date, or a date close to it, and T_3 may be the Second restrictive measures date, or a date close to it.

3. The model is supposed to reflect the natural expectation that once there are official restrictions, they will implicate an essential change in the Transmission and Removed rates although not immediately. We assume that the rate $\beta(t)$ is monotone decreasing with the time, which corresponds to the natural expectation that the more restrictive the measures the smaller the Transmission rate. Respectively, the rate $\gamma(t)$ is assumed to be monotone increasing, meeting the expectation that the stronger the measures, the bigger the removal rate.

4. We assume that $\beta(t)$ and $\gamma(t)$ are constant between the start date T_1 and the first node T_2, i.e. $\beta(T_1) = \beta(T_2)$ and $\gamma(T_1) = \gamma(T_2)$. This corresponds to the still life of the society (without containment measures) when the rates $\beta(t)$ and $\gamma(t)$ are nearly a constant.

5. To be more precise, the splines which we consider are not the usual polynomial, but the so-called exponential splines depending on a parameter in the exponent, which makes a fast decay to the next target value of the $\beta(t)$ rate; respectively this makes fast increasing to the target value of the rates $\gamma(t)$. This corresponds to the expectation that the speed by which the society switches from one level of the restrictive measures to another is relatively fast, and it is reflected by the size of the exponent we decide to choose. In fact, we use shape preserving exponential splines which are just C^1 (smooth) and do not

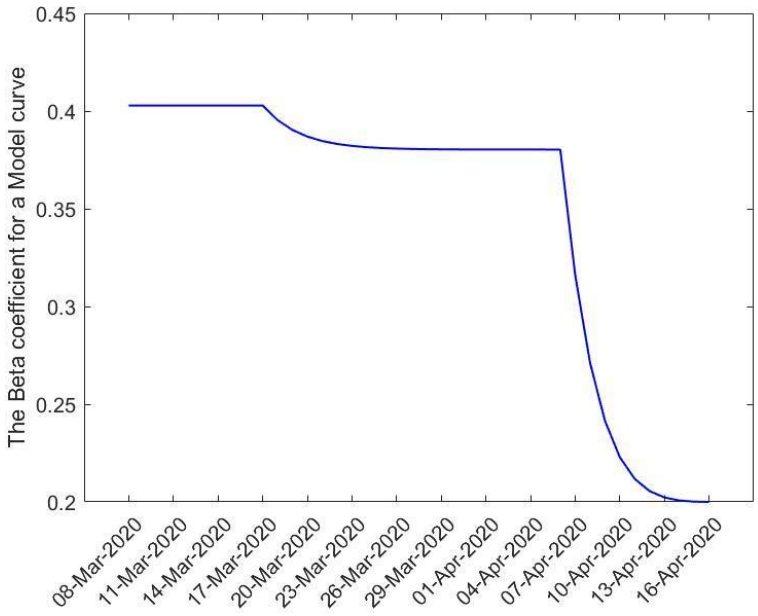


Figure 2.2: The rates $\beta(t)$ for a Model curve

need additional boundary conditions. Alternatively, one may use C^2 (twice differentiable, twice smooth) exponential splines which would be more technical due to the necessity to choose boundary conditions (at the initial and the terminal points). For the practical purposes, there are different spline functions implemented in Matlab/Octave, R, Python.

On the following Figures 2.2, 2.3 we provide examples of the dynamics of $\beta(t)$ and $\gamma(t)$ rates:

6. An important property of the TVBG-SEIR model is that due to the above spline model for the $\beta(t)$ and $\gamma(t)$ parameters, where there is a fast transition to the next target value, a classical SEIR model with constant $\beta(t)$ and $\gamma(t)$ holds during larger sub-intervals. In particular, this permits to provide a reliable estimate of the Basic Reproduction Number.

7. The **Basic Reproduction number** (**ratio**) $\mathcal{R}_0$ is a key epidemiological value for all models of epidemics, see Keeling and Rohani [2008], Lipsitch et al. [2003], Wallinga and Lipsitch [2007], Heffernan et al. [2005]. Following Heffernan et al. [2005] (formula (2.4)), for the case of the SEIR models with constant rates $\beta(t)$ and $\gamma(t)$, the basic re-

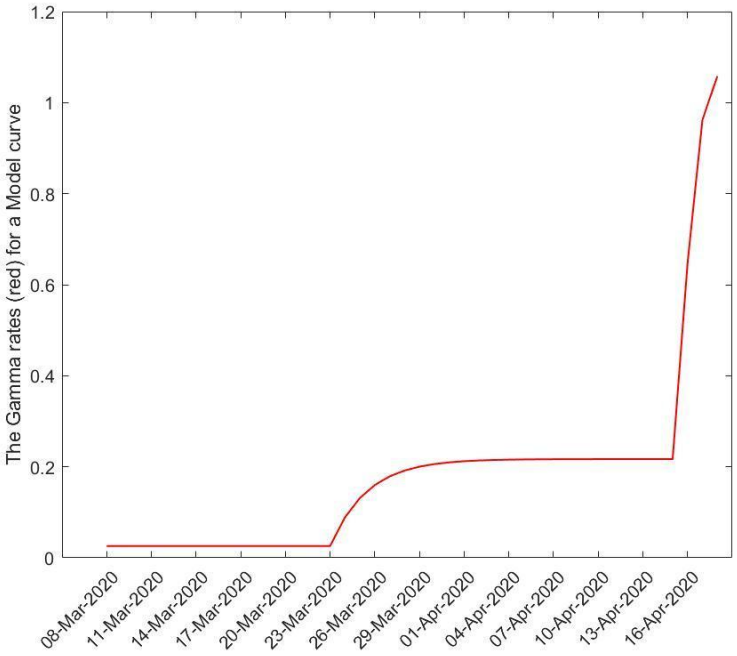


Figure 2.3: The rates $\gamma(t)$ for a Model curve

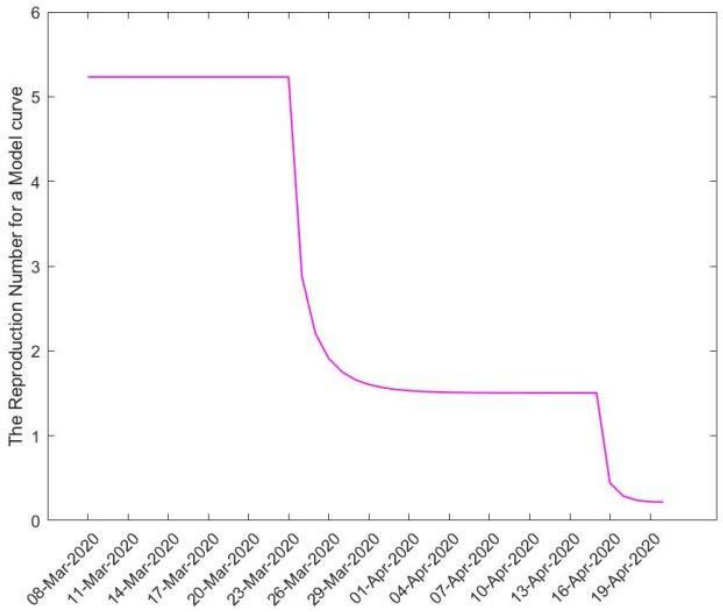


Figure 2.4: An example of a Time-Changing Basic Reproduction Number obtained by formula $\mathcal{R}_0 = \frac{\beta(t)}{\gamma(t)}$.

production number $\mathcal{R}_0$ is given by the formula

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

where we have assumed that the natural birth and mortality rates are small and also equal. Due to the above remark, we may extrapolate the above formula for all time points of interest by putting:

$$\mathcal{R}_0 = \frac{\beta\,(t)}{\gamma\,(t)}$$

In Figure 2.4 we provide the basic reproduction number $\mathcal{R}_0$, obtained by the last formula, for some specific TVBG-SEIR models.

8. In Figure 2.2 we see that the nodes of the spline satisfy $Node1 = T_2 =$ *17-Mar-2020* and $Node2 = T_3 =$*7-Apr-2020* On the other hand in Figure 2.3 we have chosen a configuration with different nodes, $T_2 =$ *23-Mar-2020*, and $T_3 =$ *15-Apr-2020*.

9. In the above examples of the dynamics of $\beta\,(t)$ and $\gamma\,(t)$ one sees the exponential factor $exp\,(0.4 \cdot (t - t_1))$ by which the curve changes from one level at $t = t_1$ to the next target level. The coefficient 0.4

is judiciously chosen and may be varied, as well as the exponential function may be replaced by a different proper function.

10. The rates $\beta(t)$ and $\gamma(t)$, are defined as interpolation splines on the subintervals defined by the start date T_1, the nodes dates T_2, T_3, and the final date T_4. Thus the whole configuration is defined by eight parameters in total, which we gather in a set Θ, given by

$$\Theta = \{T_2,\ T_3,\ \beta(T_2),\ \beta(T_3),\ \beta(T_4),\ \gamma(T_2),\ \gamma(T_3),\ \gamma(T_4)\}$$

11. The data which we use for the fitting of the discrete TVBG-SEIR model are the official data for daily new infected cases $Idata(t)$ (or their cumulative vectors $cum(Idata)(t)$ and the cumulative data for recovered and fatalities.

12. Finally, we fit the Model to the data by optimizing the positions of the two nodes T_2, T_3, and the levels of $\beta(t)$ and $\gamma(t)$, i.e. by applying Least squares minimization. Namely, we minimize the following quadratic functional $F(\Theta)$ by varying the parameter set Θ :

$$F(\Theta) = \sum_{j=1}^{n} \left(cum\left(Idata\right)(t_j) - \sigma \cdot cum(E)(t_j)\right)^2 + \\ + \left(Rdata\left(t_j\right) + Deaths(t) - R(t_j)\right)^2$$

Here we denoted by $cum(E)(t)$ the cumulative vector of the solution E_n of the discrete SEIR system until the date t.

Let us note that there are different possibilities to choose the functional $F(\Theta)$ which is used by other authors, and one of the most important arguments is the reliability of the officially announced data.

13. As we said, the minimization of $F(\Theta)$ is performed by varying by means of sampling the two nodes $T_2 < T_3$ of the splines in the interval range $[T_1, T_4]$; the interpolation values for the splines,

$$\beta(T_2),\ \beta(T_3),\ \beta(T_4),\ \gamma(T_2),\ \gamma(T_3),\ \gamma(T_4)$$

are also varied. More details about the possible choice of proper models are provided below in Section 2.5.

14. The curves S_n, E_n, I_n, R_n of the discrete SEIR model are obtained by solving the system (2.7)-(2.10) with initial conditions given by

$$S_1 = N - E_1 - I_1 - R_1$$

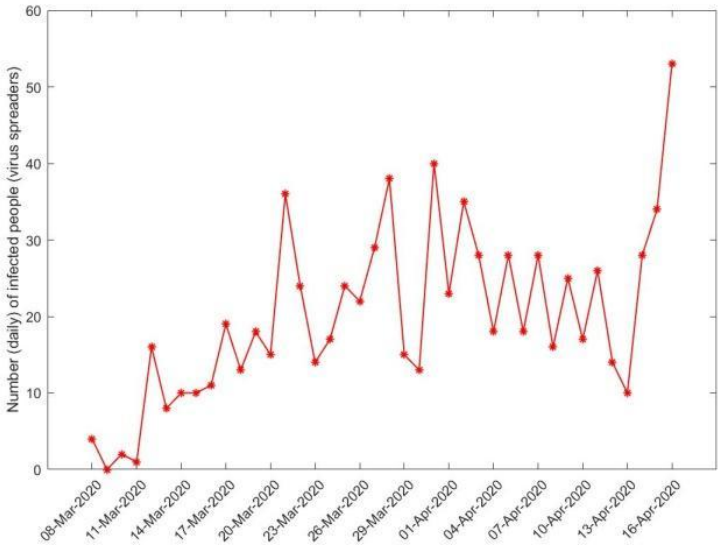


Figure 2.5: An example of daily data of infected cases Idata(t) for Bulgaria, March-April, 2020.

where N is the size of the whole population, and also,

$$\begin{aligned} \sigma E_1 &= Idata\,(1) \\ R_1 &= Rdata\,(1) + Deaths\,(1) \\ I_1 &= Idata\,(1) - R\,(1) \end{aligned}$$

As defined above, the set Θ contains the parameters which determine the (discrete versions of the) splines for $\beta\,(t)$ and $\gamma\,(t)$. Additionally, one may introduce non-negative weights $w_1\,(t)$ and $w_2\,(t)$ which give priority to some of the data.

15. In Figure 2.5 below we provide an example of daily data of infected cases $Idata(t)$ for Bulgaria during March - April, 2020.

16. The cumulative data for recovered and fatalities during March - April, 2020 are provided in Figure 2.6:

17. In Figure 2.7 we provide the fitting of the model curve $\sigma E(t)$ to the data for Bulgaria, $Idata(t)$, during March-April, 2020:

18. Figure 2.8 shows the fitting by the model curve $R\,(t)$ of the Recovered plus Fatalities data for Bulgaria:

19. It is important to remark that we have applied the parsimonious principle for constructing the spline model, by which one has to avoid

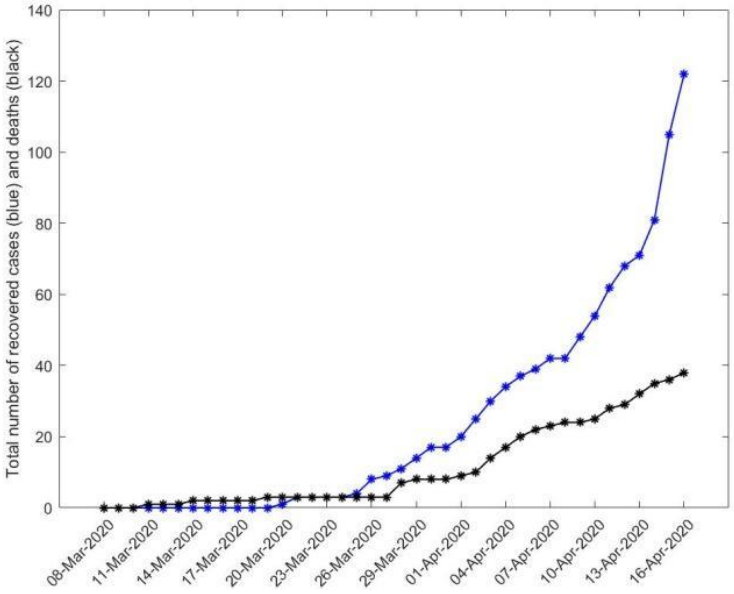


Figure 2.6: An example for cumulative data for recovered and fatalities in Bulgaria, March-April, 2020.

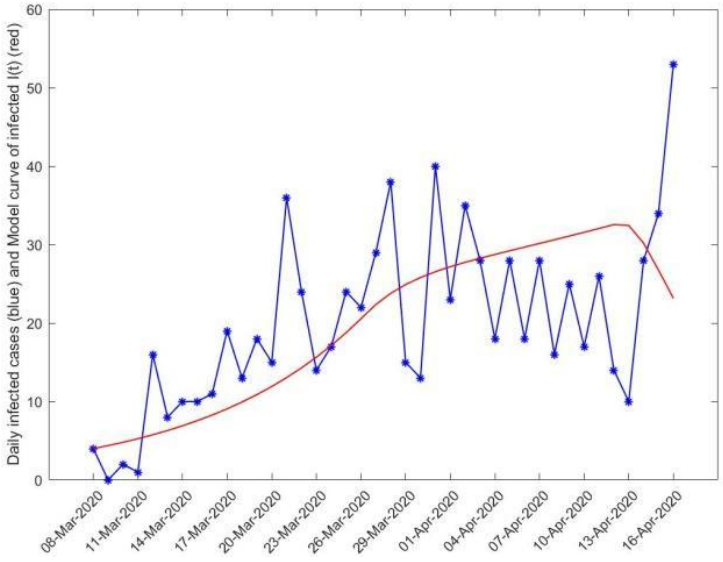


Figure 2.7: Fitting of the model curve $\sigma * E(t)$, to the data curve $Idata(t)$ for the infected cases in Bulgaria, March-April, 2020.

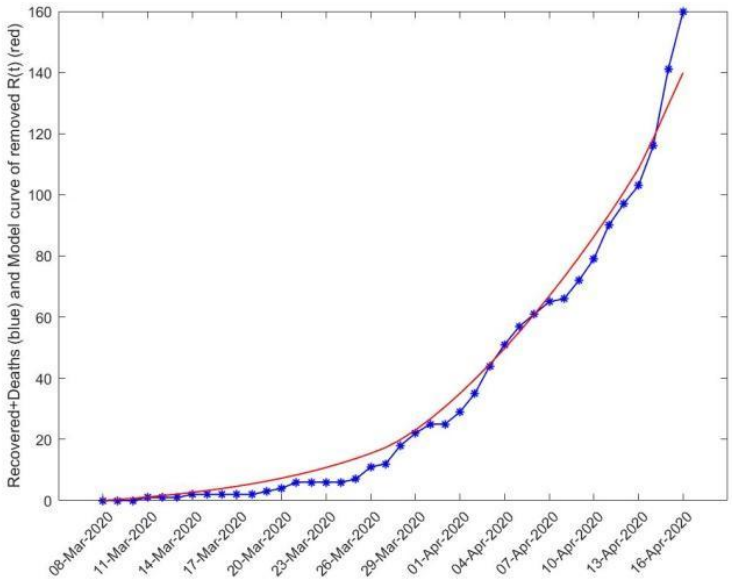


Figure 2.8: Fitting by the model curve R(t) of the Recovered plus Fatalities data for Bulgaria, March-April, 2020.

putting too many nodes in the splines since this will influence the stability of the model, and might cause overfitting, hence would spoil the predictive power of the model.

20. As we said, the result of the minimization of the quadratic functional $F(\Theta)$ is a set of parameters $\underline{\Theta}$ for which the minimum is attained, whereby there may be multiple solutions. Once we have found some model based on the parameters $\underline{\Theta}$ we proceed to constructing prediction scenarios . We choose some date T_6 which we call **Horizon**, say at most 2 (two) months from *Today* $= T_4$, by putting $T_6 = T_4 + 2\ months$

Then a scenario is defined by choosing an additional node T_5 which is a *Third restriction measures date* and the parameters

$$\{\beta(T_5),\ \beta(T_6),\ \gamma(T_5),\ \gamma(T_6)\}$$

We put

$$\begin{aligned}\beta(T_5) &= \text{Coef1} * \beta(T_4)\\ \beta(T_6) &= \text{Coef11} * \beta(T_5)\\ \gamma(T_5) &= \text{Coef2} * \gamma(T_4)\\ \gamma(T_6) &= \text{Coef22} * \gamma(T_5)\end{aligned}$$

The coefficients Coef1, Coef11, Coef2, Coef22 are used further, to control and represent our scenario building in the SBT-COVID-19 Tool. Their meaning and choice is explained in detail in Section 2.8, where we introduce the Tool.

2.5 Analysis of COVID-19 spread in Bulgaria in October-December, 2020

In the present section we provide an application to Bulgarian COVID-19 data and prediction scenarios (projections) generated by the SBT-COVID-19 Tool. We will use "prediction scenario" instead of "projection" terminology.

In the SBT-COVID-19 Tool one may find online the results for analyzing of the COVID-19 data and for generating scenarios in the case of Bulgaria, for the period March-August, 2020, (see Kounchev et al. [2021b]). There we have shown the possibility for a next wave of the infection.

Here we demonstrate how to *generate prediction scenarios* (projections) based on the Bulgarian data for the period 1 October, 2020 – 3 January, 2021. In the results provided below, it is clearly visible that the model reflects properly the wave of the epidemic in October-November and its decline at the end of December due to the lockdown imposed on November 25^{th}, 2020. It also hints the appearance of a next wave for *certain scenarios*, which correspond to special choices of the splines for $\beta(t)$, $\gamma(t)$ after *Today* date, which model relaxation of the containment measures. We have to emphasize that the date *Today* and *Third restriction measures date* are the only nodes of the splines for $\beta(t)$, $\gamma(t)$ in the interval after *Today* date. Hence, the only parameters which determine a prediction scenario are the Third restriction measures date and the values of $\beta(t)$, $\gamma(t)$ at them.

The SBT-COVID-19 Tool will be described in detail in Section 2.8.

We provide the visualizations of the discrete SEIR model fitting, which are available in the SBT-COVID-19 Tool. The ***thick red curve*** on the Figures below shows the fitted model curve until $Today = T_4$ for the daily new infected cases $\sigma E(t)$ and the blue stars show the official data for them, namely $Idata(t)$. The ***thin red curve*** shows the *prediction scenarios*, after *Today*.

Definition 3 *Under* **scenario** *we understand a choice of the coefficients* Coef1, Coef2 *which indicate whether we relax the measures (i.e. we set them to* 0.2, 0.4, 0.6, 0.8*), retain the measures (=* 1.0*) or tighten measures (i.e. we set them equal to* 1.2, 1.4, 1.6, 1.8*) which determine the parameters* $\beta(t)$, $\gamma(t)$ *of the epidemic after* Today *date, as well as of the coefficients* Coef11, Coef22*, which indicate a relaxation (if set equal to* 1.2, 1.6, 1.8*) of*

the two types of measures after the HORIZON date (for which we have three possible choices, namely, 5, 15, 25 *days from* Today*).*

1. For Bulgaria we have considered the data from the StartDate which is $T_1 =$ *1-Oct-2020*, until the end date *Today*, equal to $T_4 =$ *3-Jan-2021. Third restrictions date* $= T_5 =$ *28-Jan-2021,* and the *Horizon* date is $T_6 =$ *22-Feb-2021.*

2. As we explained in Section 2.4 the minimization of the functional $F(\Theta)$ consists of considering many pairs of nodes T_2 , T_3 (about 150 for a three-month period) for the splines $\beta(t)$, $\gamma(t)$. We select the pair T_2 , T_3 and the corresponding parameters $\underline{\Theta}$ (which define ***Model1***) for which the minimum $F(\underline{\Theta})$ of $F(\Theta)$ is attained. However there are also other parameter vectors Θ for which the functional $F(\Theta)$ attains values very close to the optimal value $F(\underline{\Theta})$. We denote these by $\underline{\Theta} = \underline{\Theta}^1$, $\underline{\Theta}^2$, $\underline{\Theta}^3$, etc. These vectors define parameters $\beta^{(j)}$, $\gamma^{(j)}$, or equivalently, models, which we denote by *Model1, Model2, Model3*, etc. The curves of the TVBG-SEIR model which correspond to these parameters $\beta^{(j)}$, $\gamma^{(j)}$ play a very useful role, and serve as an alternative to the Bootstrapping procedure as described in the classical textbooks, see e.g. the monograph of Hastie and Tibshirani (2009), Hastie et al. [2009]. Thus, it will provide us also an alternative to finding the Confidence intervals for the obtained results.

3. Let us note that in the example above the maximum value of the functional $F(\Theta)$ is 275.90 (taken over all admissible parameters Θ), while the minimum is 30.96.

4. For the optimistic, *Model1*, we have found T_2 = *9-Nov-2020,* T_3 = *2-Dec-2020*, with $F(\underline{\Theta}) = 30.96$, hence the ratio $\max(F(\Theta))/F(\underline{\Theta})$ is about *9.* Figure 2.9 shows the simplest prediction scenario starting on *Today* $=T_4 =$ *3-Jan-2021.* In the Legend of the Figure, Coef1 $=$ 1 and Coef2 $=$ 1 mean that no change by the authorities will be undertaken starting *Today* and ending on the *Third restrictions date* $= T_5 =$ *28-Jan-2021.* Further, Coef11 $= 1$ and Coef22 $= 1$ mean that no relaxation of the measures will follow starting on *28-Jan-2020.*

5. However, on *28-Jan-2021* only the measures decreasing the coefficient β may be partially relaxed, without appearance of a next wave, i.e. we may afford Coef11 $= 1.4$, This is seen on Figure 2.10:

6. The second wave is inevitable if more relaxation of the measures is allowed by the health authorities: namely, relaxing both measures, i.e. Coef11 $=$ Coef22 $= 1.8$ after *28-Jan-2021* will generate a strong next wave of infections, as seen from Figure 2.11

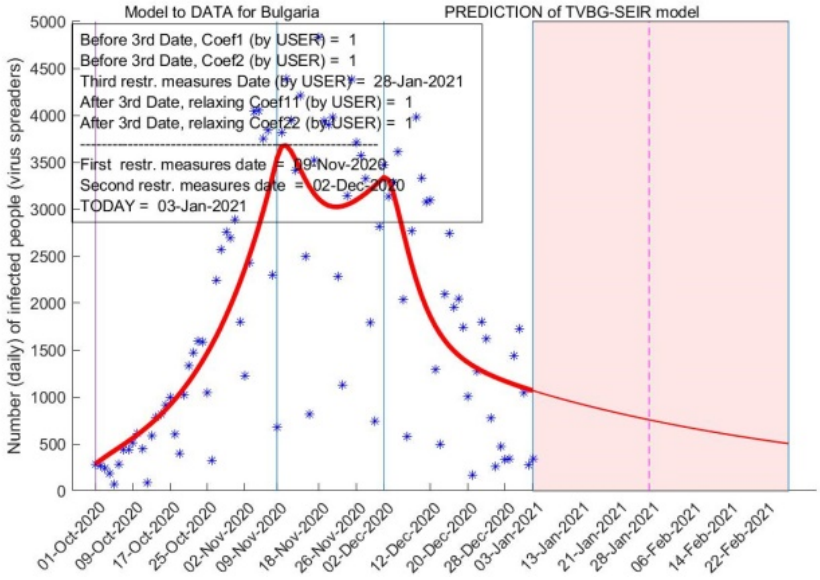


Figure 2.9: Model1: Simplest prediction scenario starting on Today = T4 = 3-Jan-2021.

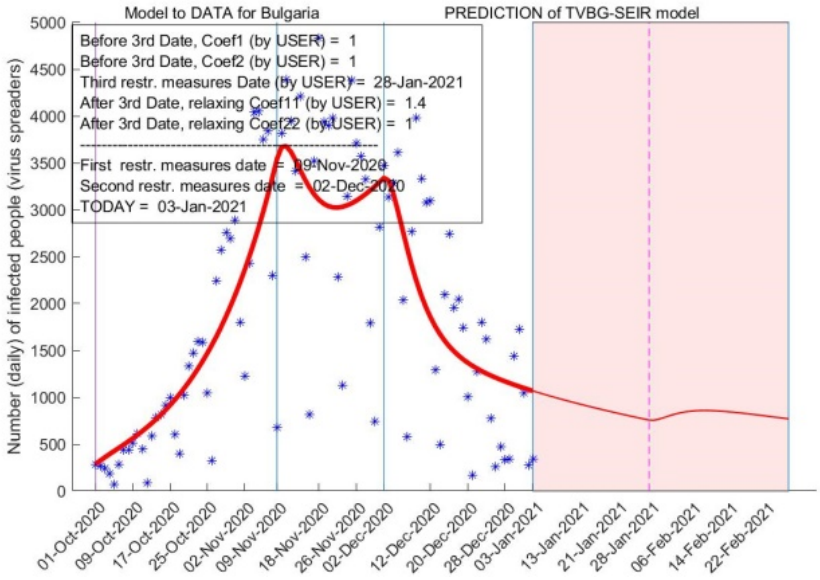


Figure 2.10: Model1: prediction scenario as on Figure 2.8, but with relaxation of β, by Coef11 = 1.4 on 28-Jan-2021

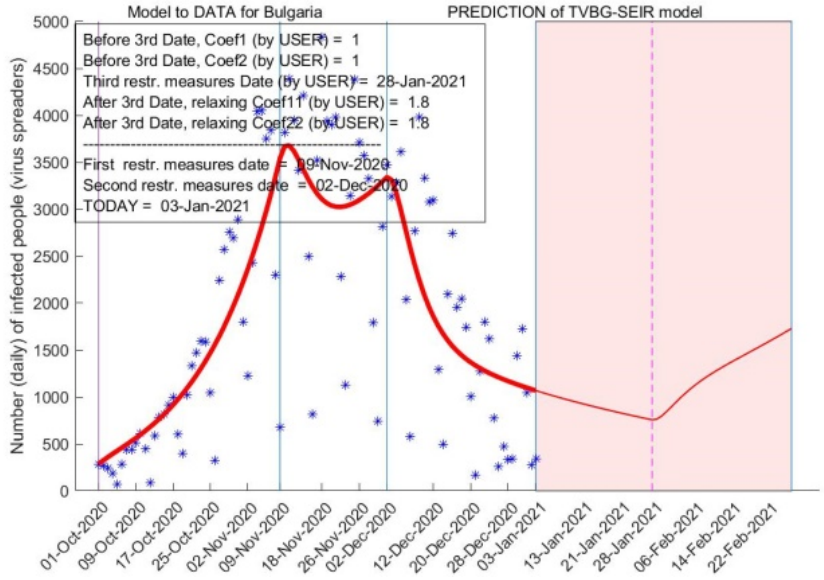


Figure 2.11: The second wave appears after relaxation of both beta and gamma, by putting Coef11 = Coef22 = 1.8 after 28-Jan-2021.

7. Similar are the conclusions with *Model3* for Bulgaria (with $Fval = 31.60$, with next wave appearing as well.

8. For *Model2* (with $Fval = 31.78$) we have the **most optimistic** scenario since we may partially **relax both measures** after *28-Jan-2021* (i.e. Coef11 = Coef22 = 1.4), and no second wave will appear, as seen from Figure 2.12

As we mentioned above, we may use say ten models *Model2, Model2,...*, *Model11* and generate their curves $I(t)$ to obtain estimate of the Confidence intervals at every time t. However we found the above presentation using optimistic and pessimistic scenarios more simple and clear.

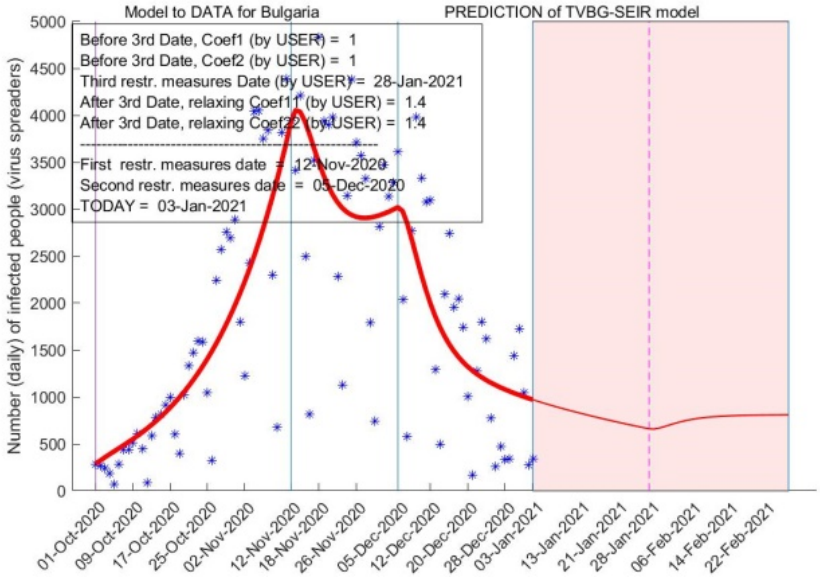


Figure 2.12: No second wave appears for Model3.

2.6 Application to Italian COVID-19 data in October-December, 2020, and scenarios generated by the SBT-COVID-19 Tool

In the present section we provide similar results obtained by our *SBT-COVID-19 Tool* for the Italian data.

1. The results about Italy considered till *Today = 1-Jan-2021* are similar to Bulgarian. For *Model0* we have T_2 = *6-Nov-2020*, T_3 = *20-Nov-2020*, and
$$Fval = F\left(\underline{\Theta}\right) = 153.09.$$
See Figure 2.13.
A strong relaxation after the Horizon date *26-Jan-2021* results in a strong next wave seen in the following Figure 2.14:

2. *Model2* has T_2 = *3-Nov-2020*, T_3 = *17-Nov-2020*, and
$$Fval = F\left(\underline{\Theta}\right) = 159.77$$
For it we obtain the following scenario, see Figure 2.15:

3. Just as in *Model1*, further strong relaxation gives a strong next wave provided in Figure 2.16:

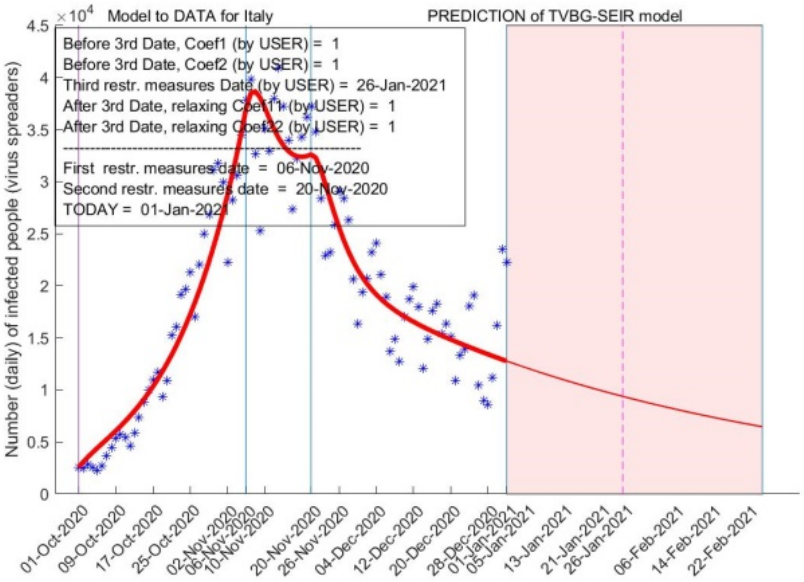


Figure 2.13: Italy, Model0: an optimistic prediction scenario.

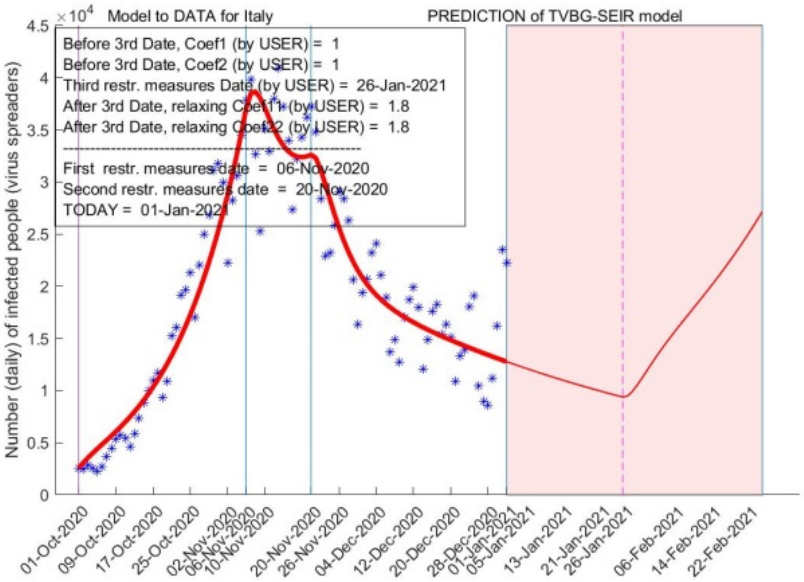


Figure 2.14: Italy, Model0: next wave appears after strong relaxation both restriction measures by putting Coef11 = Coef22 = 1.8

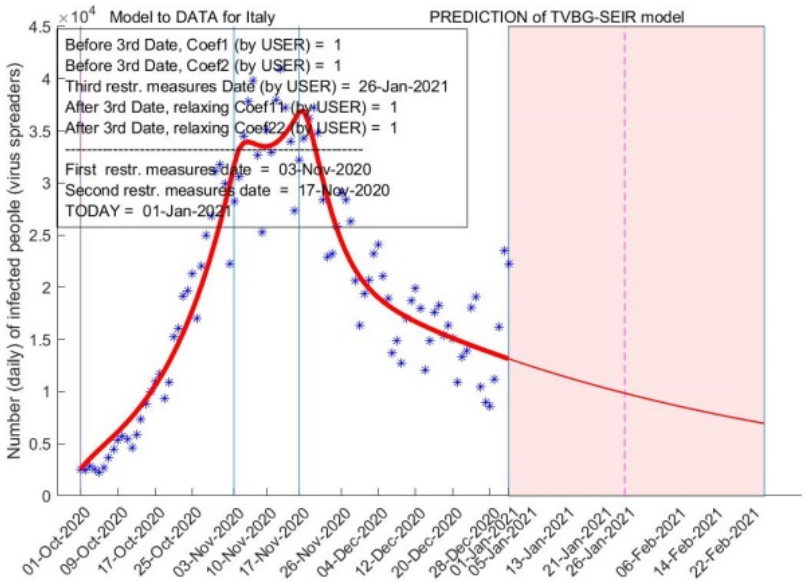


Figure 2.15: Italy, Model2: no relaxation - no next wave.

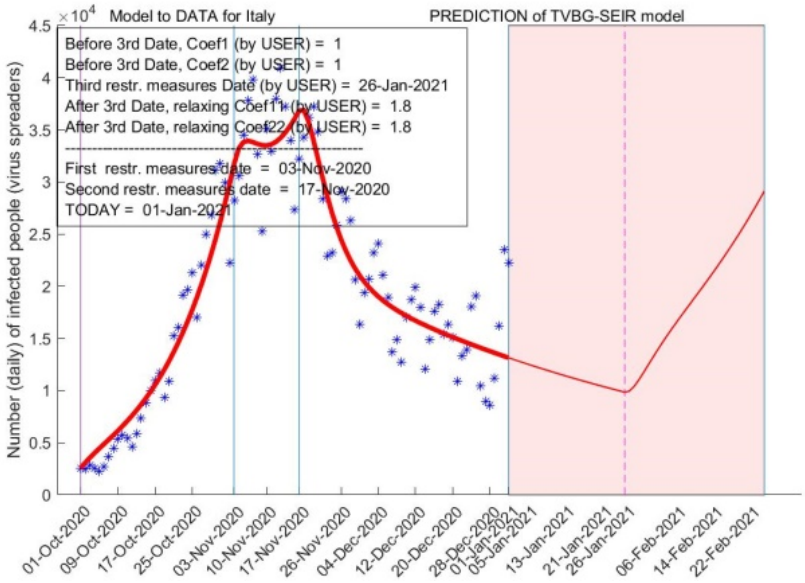


Figure 2.16: Italy, Model2: strong relaxation for both measures causes next wave.

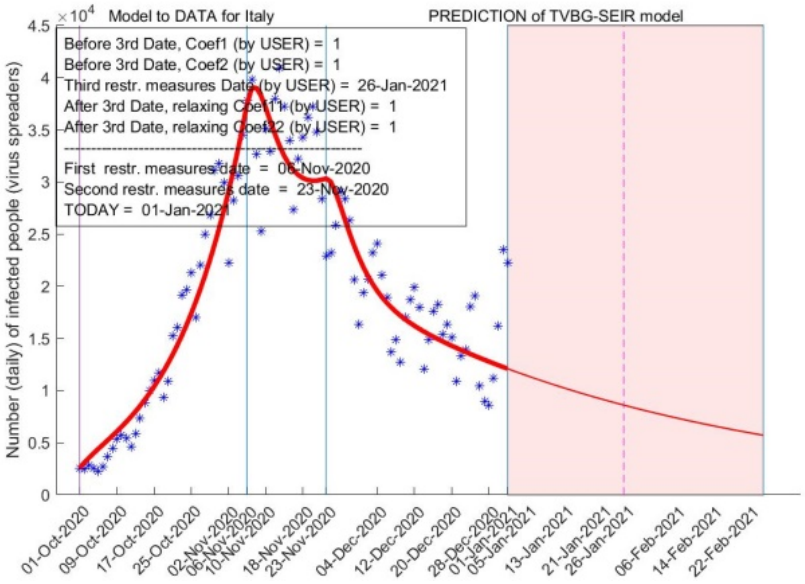


Figure 2.17: Italy, Model3: no relaxation - no next wave.

4. *Model3* has $T_2 = 6\text{-}Nov\text{-}2020$, $T_3 = 23\text{-}Nov\text{-}2020$, and $Fval = F(\underline{\Theta}) = 169.72$.

   It gives the following Figure 2.17

5. Further relaxation after *26-Jan-2021* shows a bigger next wave in the Figure 2.18

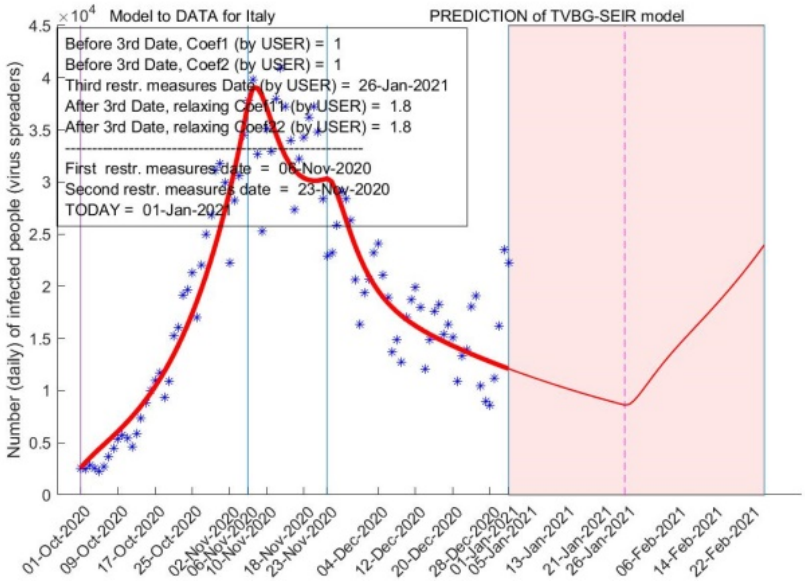


Figure 2.18: Italy, Model3: strong relaxation causes strong next wave.

2.7 Application of the model to German COVID-19 data in March-May, 2020

We provide an analysis of the German data by means of scenarios generated by the SBT-COVID-19 Tool.

The overall observation is, that unlike the data for Bulgaria and Italy, during the period March-May, 2020, they have shown a very strong tendency to explode into a next wave.

1. For the German data, on the *Today* date = *1-Jan-2021*, according to our *Model1* for Germany, we have T_2 = *3-Nov-2020,* T_3 = *11-Nov-2020*, and
$$Fval = F\left(\underline{\Theta}\right) = 160.96.$$
It is seen that if the containment measures remain the same as before *Today*, then Germany is already in the next wave , which is seen from the Figure 2.19

2. However, a moderate tightening of the second measures (Coef2 = 1.2) will result in a calming down, see Figure 2.20.
But a much better result will bring the tightening of the measures influencing the coefficient β, i.e. Coef1 = 1.2 as seen in Figure 2.21:

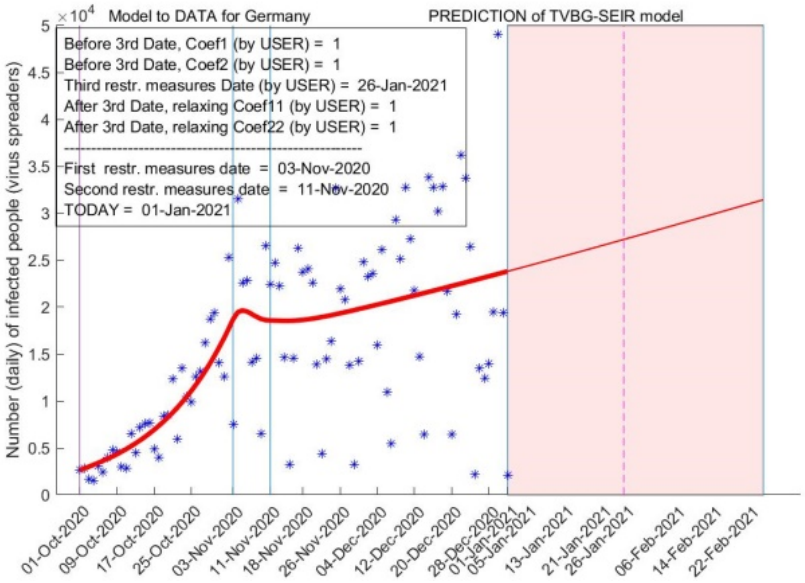


Figure 2.19: Germany, Model1: no more restrictions Today - remains in a wave.

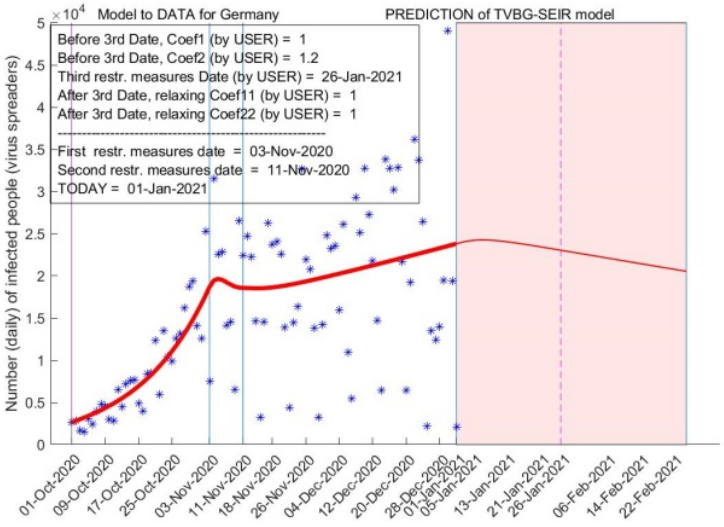


Figure 2.20: Germany, Model1: Restrictions Today cause decay of the wave.

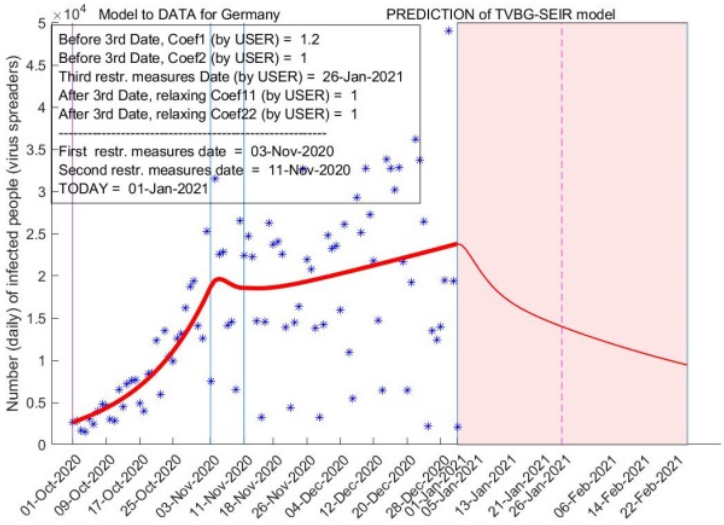


Figure 2.21: Germany, Model1: Stronger Restrictions Today causes strong decay of the wave.

3. For *Model2* we have $T_2 = 3\text{-}Nov\text{-}2020$, $T_3 = 8\text{-}Nov\text{-}2020$, and

$$Fval = F\left(\underline{\Theta}\right) = 169.16,$$

and for *Model3* - $T_2 = 3\text{-}Nov\text{-}2020$, $T_3 = 20\text{-}Nov\text{-}2020$, and

$$Fval = F\left(\underline{\Theta}\right) = 171.61$$

The application of the two models gives a result similar to that of *Model1*.

In Figures 2.22, 2.23, one may choose how strong the tightening of the measures has to be, within *Model2,* in order to obtain a stronger slowdown of the infection progression,:

And for *Model3*: we have a slight difference in Figure 2.24:

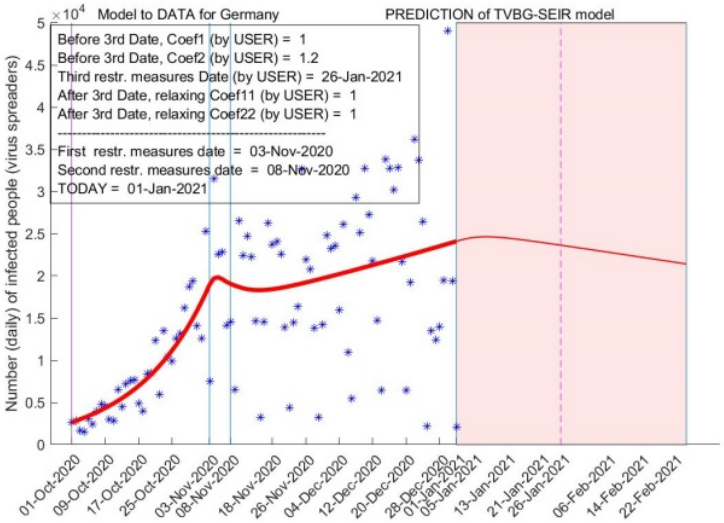


Figure 2.22: Germany, Model2: Restrictions Today causes decay of the wave.

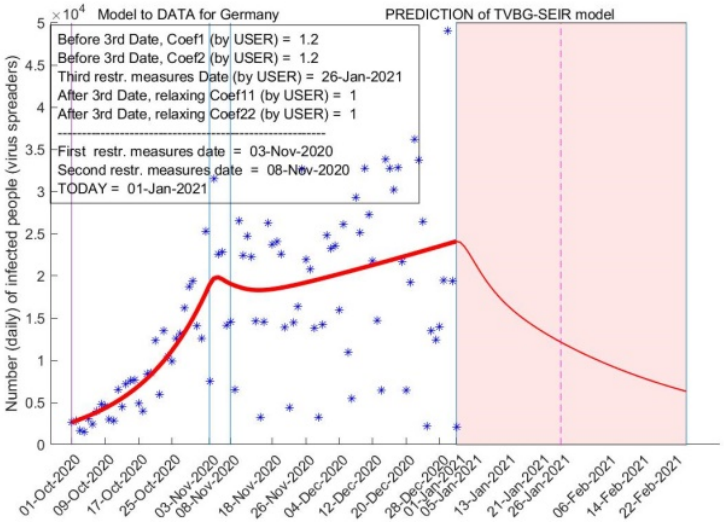


Figure 2.23: Germany, Model2: Strong Restrictions Today cause strong decay of the wave.

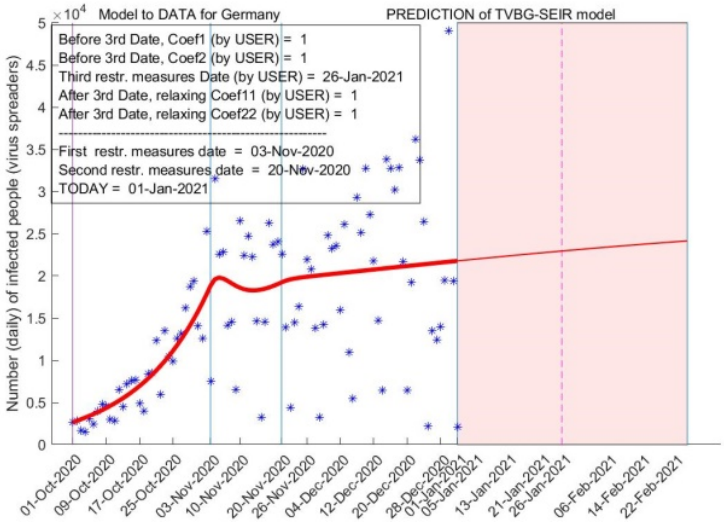


Figure 2.24: Germany, Model: Very similar to previous Model2.

2.8 Description of the SBT-COVID-19 Tool for controlled scenarios

We provide a short informative description of the Web-based tool SBT-COVID-19, which implements the results obtained by the TVBG-SEIR model.

1. We have designed a **SBT-COVID-19 Tool** for the Model scenario predictions (projections) of the Coronavirus (and similar infectious diseases) spread. The **SBT-COVID-19 Tool** is based on the fitting of the TVBG-SEIR model to the official data available on a daily basis as described in Section 1.3).

2. The online tool is available at the site `http://213.191.194.141:8888/notebooks/TVBG-SEIR-Spline-model_v3.ipynb`

3. First of all, we fit the model for the time series in the interval $[T_1,\ T_4]$, where T_4 is Today's date. Then the USER may choose several parameters to make a prediction about the virus spread during the period $[T_4, Horizon]$ where $Horizon$ is chosen to be at most 2 months from $Today\ = T_4$.

4. The first parameter, to be controlled, is the Third restrictive measures date denoted by T_5. The USER may choose several options, say 5, 15,

25 days from *Today* ($= T_4$), i.e. one may select the dates

$$T_5 = \begin{cases} T_4 + 5 \\ T_4 + 15 \\ T_4 + 25 \end{cases}$$

5. Then the USER may decide how to strengthen or relax the Beta measures and the Gamma measures during the period $[T_4,\ T_5]$, by means of the coefficients Coef1 and Coef2 respectively; Coef1 = 1 means that the Beta measures remain the same, while Coef2 = 1 means that the Gamma measures remain the same in the period $[T_4,\ T_5]$. If Coef1 < 1 then this means that the Beta measures are weaker , and also, the smaller Coef1, the weaker are the Beta measures and they will reach a target value at the date T_5, which is defined by the size of Coef1 (Note that Coef1 < 1 means that the rate $\beta(t)$ will be bigger!). In a similar way, if Coef2 < 1, then this means that the Gamma measures will be weaker , and the smaller Coef2, the weaker are the Gamma measures (note that in such case the rate $\gamma(t)$ will be smaller!). A target value (determined by the size of Coef2) will be reached at the date T_5.

6. On the other hand, if Coef1 or Coef2 are bigger than 1, this means strengthening the measures , resp. of Beta measures and Gamma measures in the period $[T_4,\ T_5]$ to some target value defined by Coef1, Coef2.

7. The USER has further the possibility to decide what will happen after date T_5 - to weaken or leave the same the Beta and the Gamma measures. This is decided by the choice of two coefficients – Coef11 for the Beta and Coef22 for the Gamma measures. Coef11 = 1 means that one retains the same level of the Beta measures; Coef22 = 1 means that one retains the same level of the Gamma measures. If Coef11 > 1 then this would relax the Beta measures – the bigger Coef11 the more the relaxation. Coef22 makes the same for the Gamma measures.

8. A similar way to represent the above approach to Scenarios design and visualization is implemented since relatively recently in the popular online tool COVID-19 Projections, `http://www.healthdata.org/covid/`, embedded in the Worldometers.info web-page, which we have mentioned in section 1.4.1. However, initially, until the end of July 2020, they used confidence intervals around the most probable scenario. Instead of using coefficients Coef1, Coef2, they use a more descriptive terminology for worse and better scenarios, as 95% masks usage, short lasting vaccination, etc. Their terminology changes very fast in time.

2.9 Recent research on time-varying transmission rates

As we said, presently it is urgent to consider SIR/SEIR models with time-varying $\beta(t)$ and $\gamma(t)$ rates.

For completeness sake, we mention some research about solving an inverse problem for finding time-varying $\beta(t)$, in a SIR model, for a fixed removal rate γ, Pollicott et al. [2012], from the number of infectious cases. In Boatto et al. [2018], the authors do research and provide further references of research on specific models for the transmission rate $\beta(t)$.

Chapter 3

ATVBG-SEIR: SEIR models for long-term predictions based on splines

The main purpose of the present chapter is to introduce a model ATVBG-SEIR for generation of long term prediction scenarios (projections), which will help us answer the question, how long lockdowns are needed to end the COVID-19 epidemic in a single country, with or without vaccinations. We will provide applications to data from different countries, as Austria, Bulgaria, Germany, Italy, UK and USA. These models are implemented in a Web-based tool, for creating Lockdown Scenarios.

3.1 Short summary

The main purpose of the present Chapter is to present a methodology for the estimation of the duration of the Epidemic of COVID-19 in a single country, accounting for different scenarios. Our methodology is based on a specific SEIR model, called ATVBG-SEIR model, which is explained in detail. We include vaccinations in the model which are carried out according to a vaccination plan provided on a monthly basis. The algorithm takes into account the main constraint of the health system which is the number of Intensive Care Units (ICU) intended for COVID-19 patients (they are e.g. about 1100 in Bulgaria, about 8000 in Germany, etc.). At the end, we present a web-based Lockdown Scenarios Tool, available online at **http://atvbg-seir.eu** based on the algorithm implementing the methodology. Results are demonstrating the efficiency of the tool by applying it to COVID-19 data from Austria, Bulgaria, Germany, Italy, UK and USA.

We have implemented a model for the Seasonal effect (the summer season) in the ATVBG-SEIR model, which is a very challenging task in general.

Another challenge is the modeling of the duration of the vaccine and

natural immunity. As is well-known now thanks to the Chinese experience (see [Nature, 2021, Bibliography to chapter 1], [He et al., 2021, Bibliography to chapter 1]), the **durability** of the natural immunity (after the course of the disease) is about 9 months. On the other hand, there are no definite results in the scientific literature about the durability of the post-vaccination immunity (acquired due to vaccination) – it differs according to different experts from 5 to more than 12 months; in our implementation we have decided for 6 months durability in this case. Hence, we have decided to incorporate in our model the loss of both types of immunity to be 9 months and 6 months respectively. This feature of "immunity loss" makes the model seem to be more complicated but we still succeed to keep it simple and manageable.

3.2 Introduction to the model

There is a list of features of the COVID-19 disease, which distinguish it essentially from the usual seasonal flu, and have to be taken into account by the modellers of the long-term behaviour of the spread. We summarize these features briefly:

1. The main characteristics of the Coronavirus are determined by its insidious properties – namely, majority of people (about 80%, mainly children) have an extremely light asymptomatic course of disease; very few people have a heavy progression of the disease with more than three weeks of symptomatic course, especially those with underlying conditions. As a result, the total amount of infected people who need hospitalization is considerable and quite many need to stay in the ICUs (intensive stations, intensive care units, intensive beds), where they get additional oxygenation. The number of the ICUs reserved for COVID-19 cases is in every particular country limited although somewhat flexible, e.g. in Bulgaria there are about $1,100$ such beds (with a total limit of about$2,000$ ICUs for the whole country), while in Germany they are at least $8,000$ (with a total limit of $40,000$ ICUs), $2,764$ in Italy, $26,900$ in USA, etc. This number of ICUs represents the main restriction which has to be taken care of by the Health authorities who are in charge of imposing intervention measures against the development of the COVID-19 epidemics.

2. Another important characteristic feature of the COVID-19 infection is the very long incubation (latent) period, from 5 to 14 days which causes a very long tail in the curve of the Active cases. Hence, the average time of the course of COVID-19 disease plus the incubation time are several times bigger than that of the seasonal flu; this explains

the magnitude of the basic reproduction number $\mathcal{R}_0$ of COVID-19 exceeding very often six.

3. Those two characteristics enable the very wide proliferation of the virus infection, although its **fatality (mortality)** rate does not differ drastically from that of the seasonal flu (as many specialists claim). Thus, the comparison with the usual "seasonal flu" is superficial if the above remarks are not taken into account.

4. In the months of October through March there are two main factors: the negative effect of winter **drop of immunity**, aggravated by the open **schools/universities** and **kindergartens**. However, starting April-May we have an average increasing of population's immunity.

5. Hence, the issue with the ICUs is the main bottleneck of the health systems. The *empirical* data show that there is a rather **stable correlation** between the number of the ICUs needed currently for COVID-19 patients (called also ICU critical cases) and the number of the so-called Active cases (see a rigorous Definition below).

6. The data about the current occupancy of the ICUs in many countries may be retrieved from the official data sources, with nice visualization, e.g. at the site `https://www.ecdc.europa.eu/en/publications-data/download-data-hospital-and-icu-admission-rates-and-current-occupancy-covid-19`, `https://ourworldindata.org/covid-hospitalizations`, `https://www.worldometers.info/coronavirus/`.

   However many countries have specialized portals which sometimes provide data faster and with a better quality, although sometimes not very regularly.

In our methodology and the related Algorithms and Tool we have focused on the following concepts and technical issues:

First of all, due to the long incubation period of COVID-19, the SEIR model is very suitable for modeling the dynamics of the epidemic, since the people who have contacted infectious individuals stay a considerable amount of time in the Compartment of the Exposed. The main parameters of the SEIR models are transmission rate β (reflecting the spread of the virus from infectious individuals) and the rate of "removal / elimination" γ (reflecting the measures for hospitalization / isolation). Measures to reduce the transmission rate β are: social distancing, wearing masks, fewer contacts, incl. social isolation, less travel, etc. Measures to increase the "removal rate" γ are: rapid detection and quarantine of infected persons through testing and reporting, quickly tracking contacts, etc.

Since the interventions of the health authorities change essentially the dynamics of the epidemic, the classical constant parameter SEIR model does not manage to provide an adequate model. Hence, specific type of SEIR models are necessary which reflect properly the dynamics of the parameters $\beta,\ \gamma$ and the course of the epidemic development.

The idea of the ATVBG-SEIR model is the following:

1. We subdivide the dynamics of the model of the epidemics into two different *altering regimes* (roughly, **0s** and **1s**): The first is a ***StrictRegime*** (mimicking strict measures) during a fixed and short period of time (say 21 days) and has a constant basic reproduction number $\mathcal{R}_0 < 1$. During the Strict regime we use a classical SEIR model with constant parameters $\beta,\ \gamma$ .The values of these parameters have been obtained by *calibration* of the standard SEIR model to certain period of the official data for Bulgaria (or other country) for which it is known that Strict Containment measures have been applied. The other regime is ***RelaxRegime*** (mimicking Relaxed measures) which has basic reproduction number $\mathcal{R}_0 > 1$, and might have a longer (not fixed) duration. During the Relaxed regime we use another classical SEIR model for which we obtain the values of the parameters $\beta,\ \gamma$ by *calibration* of the SEIR model to the official data for Bulgaria (or other country) for a period of time where we know that Relaxed Containment measures have been applied.

As is well-know, the calibrated parameters $\beta,\ \gamma$ of the SEIR model determine the basic reproduction number by means of the formula

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

It would be more correct to write $\mathcal{R}_e$ instead, which is the effective reproduction number in a non-steady dynamical system.

1. Thus, the ATVBG-SEIR model consists of altering Strict regime with Relaxed regime periods, whereby the duration of the Strict regime periods is kept fixed while the duration of the Relaxed regime periods is left flexible to predict the future. The altering regimes of our model may be considered as a hesitation of the society in the ethical dilemma – between the risky expectation for the herd immunity, and the fast somewhat risky vaccine, see Aschwanden (2020), Aschwanden [2020].

2. Our methodology is applicable to every country, but one needs to carefully collect the data from different sources, especially important is to have adequate data for the critical cases needing ICUs. Also, one has to select appropriate "typical time periods" in the development of

the epidemic when the measures introduced by the health authorities were *strict* and efficient, and other time periods when the measures were *relaxed*. The crux of the matter is the very logical assumption that during a period of time when the Containment measures have remained unchanged, the main parameters β, γ of the classical SEIR (or similar compartmental) model will not vary essentially and would provide a good approximation to the data.

3. The focus of our observation is the application of the ATVBG-SEIR model to a forecast of the dynamic of the ICU critical cases (the number of the individuals with COVID-19 who need ICUs), by generating prediction scenarios (projections).

The plan of the present chapter is as follows:

In section 3.3 we show how the curve of the Active cases may be expressed in terms of the curves of the continuous SEIR model which has been introduced in section 2.2.

In section 3.4 we remind the discrete SEIR model from section 2.2 and introduce the notion of Modified Active cases and compare them with the usual notion of Active cases.

In section 3.5 we recall the correspondence between the empirical data and the curves of the discrete SEIR models.

In section 3.6 we incorporate the vaccinations in the SEIR model.

In section 3.7 we introduce the ATVBG-SEIR model.

In section 3.8 we show how the number of ICU critical cases may be expressed within the ATVBG-SEIR model via the curve of the Active cases, and mention the main points of the Algorithm.

In section 3.9 we fill the gap of the usual stopping rule (the 70% rule), related to the equilibrium properties of the SIR/SEIR models. We introduce a second stopping rule.

In section 3.10 we list the main properties of the Algorithm, including initial values for the curves of the ATVBG-SEIR models.

In section 3.11 we provide the Main results of our approach.

In sections 3.12- 3.14 we show the thorny way of model calibration: this is an important step of our Methodology (and Algorithm) – the identification of *Strict* and Relaxed regimes by fitting the ATVBG-SEIR model to Bulgarian, German or other countries data. This step is the very first in the Algorithm.

In section 3.15 we provide a short instruction how to work with the web-based Lockdown Scenarios Tool, which was used to obtain the main results of the present Chapter.

A reader not interested in the methodology and Algorithm but just in the main results of the chapter may go directly to section 3.11.

3.3 Active cases in the continuous SEIR model

The main idea in the development of the SEIR model is to provide some **traceable** approximations $S(t)$, $E(t)$, $I(t)$, $R(t)$ to the actual data. In section 2.2 by the system of equations (2.1)-(2.4) we have introduced the classical differential SEIR model, following the usual references Anderson and May [1991], Hethcote [2000], Keeling and Rohani [2008].

The usual applications of the SEIR model are with constant rates $\beta(t)$ and $\gamma(t)$. One assumes that the initial values $S(0)$, $E(0)$, $I(0)$, and $R(0)$ are given and the system is solved for the times $t \geq 0$. As we have already discussed, see (2.5), it is assumed that the following equation holds,

$$N = S(t) + E(t) + I(t) + R(t)$$

where N is the total population in the country XX.

Let us remind taht one of the most important properties of the SEIR model is that it provides an easy way to express the basic Reproduction ratio:

$$\mathcal{R}_0 = \frac{\beta}{\gamma}$$

which shows how many people are infected by one infectious person during his/her being sick.

We have mentioned the empirical observation that the critical ICU cases are a rather stable percentage of the Active cases. Hence, we need to find an expression for the Active cases within the SEIR model. In the available *empirical data* the Active cases on the day t are given by the formula:

$$\begin{aligned} Active\ Cases\ (t) &= Total\ number\ of\ infected(t) \\ &- Total\ number\ of\ Recovered(t) \\ &- Total\ number\ of\ Fatalities(t) \end{aligned} \tag{3.1}$$

Let us assume for simplicity that the whole interval of interest is the period $0 \leq t \leq T$. It is very natural to define the total number of infectious cases on the day t by the integral

$$TotalInf(t) = \int_0^t \sigma E(s)\, ds$$

which is the total number of people who have come to compartment C_I in the period $[0, t]$ (we use the terminology of the compartment formulation of the SEIR model, see e.g. Keeling and Rohani [2008].

Hence, the number of the Active cases which we denote in the continuous case by $AC(t)$ will be given in the continuous SEIR model by the integral

$$\begin{aligned} AC(t) &= TotalInf(t) - R(t) \\ &= \int_0^t \sigma E(s)\, ds - R(t) \end{aligned} \tag{3.2}$$

It is a classical fact from the theory of Ordinary Differential Equations, that the curves (solutions) of the SEIR model are smooth (differentiable) **at the points t where the parameters** $\beta(t)$, $\gamma(t)$ **are continuous** (see e.g. Katok and Hasselblatt (1999), Katok and Hasselblatt [1997]). The last remark is very important for us since by definition the model ATVBG-SEIR is essentially a SEIR model with piece-wise continuous parameters $\beta(t)$, $\gamma(t)$ with a finite number of break-points.

Hence, at the times t where $\beta(t)$, $\gamma(t)$ are continuous the above formula for $AC\left(t\right)$ implies that the derivative of the Active cases $AC\left(t\right)$ is given by

$$\begin{aligned} AC'\left(t\right) &= \sigma E\left(t\right) - R^{'}\left(t\right) \\ &= \sigma E\left(t\right) - \gamma I\left(t\right) = I'(t) \end{aligned}$$

Hence, in the SEIR model the slope of the Active cases $AC(t)$, is equal to the slope of the infectious cases $I(t)$. Hence, by using standard arguments from the theory of ODEs we may easily prove the following result which provides the interpretation of the variable $I\left(t\right)$ of the SEIR model (cf. Kounchev et al. [2021a], Kounchev et al. [2021c]):

Theorem 4 *Assume that the variable parameters $\beta(t)$, $\gamma(t)$ of the SEIR model have a finite number of points of discontinuity, which we denote by $0 < t_1 < t_2 < \cdots < t_p < T$. Then for every moment of time $t \neq t_j$ for $j = 1, 2, \ldots, p$ with $0 \leq t \leq T$ holds*

$$AC\left(t\right) = I\left(t\right) + C \tag{3.3}$$

where C is a constant.

Next, our main purpose is to find the discrete SEIR model counterpart of the notion of Active cases, which appears to be a non-evident problem.

Remark 5 *Here we have to make an **important** remark about a certain mismatch between the Active cases in the official data and the Active cases in the Compartmental model, see e.g. Keeling and Rohani [2008]. While in the official data under active case one understands a person who is currently sick and may be in the hospital, in the Compartmental model such person is considered to be in the Removed compartment, since in a hospital the person is supposed to be **no more virus spreader** (we assume that the hospitals are working properly).*

3.4 Discretization of SEIR model and Active Cases

In practice one uses a discretization of the continuous SEIR model. We have considered the following discretization of the SEIR model (2.7)-(2.10)

in section 2.3. It is very intuitive, and is in fact derived from the Euler method for approximate solution of the initial value problem (2.1)-(2.4) in section 2.2 (cf. also Keeling and Rohani [2008]),

$$\begin{aligned} S_{n+1} &= S_n - \frac{\beta_n S_n I_n}{N} \\ E_{n+1} &= E_n + \frac{\beta_n S_n I_n}{N} - \sigma E_n \\ I_{n+1} &= I_n + \sigma E_n - \gamma_n I_n \\ R_{n+1} &= R_n + \gamma_n I_n \end{aligned}$$

As in the continuous version of the model, we assumed that the size N of the population remains unchanged (normal birth and mortality are not taken into account), recall equation (2.11):

$$N = S_n + E_n + I_n + R_n$$

An important point in our discretization framework is the way we define the Active cases. It appears that the direct and most obvious definition does not have nice properties.

Now we provide the following definition of what we call **Modified Active Cases**:

Definition 6 We define the **Modified Active cases for the SEIR model**: For $N = 1$ we put

$$MAC_1 = TotalInf\,(1) - R_2 \tag{3.4}$$

and for every integer $P \geq 2$ we put

$$MAC_P = TotalInf\,(1) + \sigma \sum_{n=2}^{P} E_n - R_{P+1} \tag{3.5}$$

Since σE_n represents the newly infected cases, we see that the above sum is in fact expressed by the equation

$$TotalInf\,(P) = TotalInf\,(1) + \sigma \sum_{n=2}^{P} E_n$$

From the above definition and the equations (2.9), (2.10) we obtain the simple formula:

$$\begin{aligned} MAC_P - MAC_{P-1} &= \sigma E_P - R_{P+1} + R_P \\ &= \sigma E_P - \gamma I_P = I_{P+1} - I_P \end{aligned}$$

In the continuous limit it means that the Modified Active Cases are approaching the variable I_t of the continuous SEIR model (perhaps with some constant).

After summing up we obtain the main result:

Theorem 7 *The* **Modified Active Cases** satisfy the equality

$$MAC_P - MAC_1 = I_{P+1} - I_2 \tag{3.6}$$

Or equivalently,

$$MAC_P = I_{P+1} + TotalInf\,(1) - R_2 - I_2.$$

Thus, we see that up to a constant the Modified Active Cases at the moment P (which is an observable *data value) is equal to the variable I_{P+1} of the discrete SEIR model.*

The intriguing **discovery** of the above simple equations is that it is **more natural** to consider the Modified Active Cases instead of the usual and seemingly more natural straightforward definition of Active Cases. Namely, if we put

$$AC_P = TotalInf\,(1) + \sigma \sum_{n=2}^{P} E_n - R_P$$

then we obtain

$$\begin{aligned} AC_P &= MAC_P + R_P - R_{P+1} \\ &= MAC_P - \gamma_P I_P \end{aligned}$$

Further, we obtain the following relation for the (finite) difference:

$$\begin{aligned} AC_P - AC_{P-1} &= \sigma E_P - R_P + R_{P-1} \\ &= \sigma E_P - \gamma_{P-1} I_{P-1} \end{aligned}$$

Hence, after summing up in the variable P we obtain

$$AC_P - AC_1 = I_{P+1} - I_2 + \gamma_P I_P - \gamma_1 I_1$$

Obviously, the last is **not** a closed expression of AC_P by means of I_P since it contains the parameter γ, and lacks the nice properties of the Modified Active Cases.

3.5 Correspondence between Official data and discrete SEIR model

This is the most important part of our theoretical discussion.

1. The most important remark is that usually the notation for the curve $I(t)$ in the SEIR model is somewhat misleading. Usually, it is mixed with the daily newly infected cases. However, on the day t the daily *New_Infected* cases are equal to

   $$NewInfected_t = \sigma E_t$$

   where E_t is the variable of the discrete SEIR model.

2. We have seen above that in the *continuous* SEIR model the curve $I(t)$ of the SEIR model coincides with the Active Cases on the day t, up to a piece-wise constant function, recall Theorem 7. However, in the *discrete* SEIR model the discrete curve I_t coincides up to a constant with the Modified Active Cases MAC_{t-1}.

Let us remark that it is not difficult to prove that the Active Cases and the Modified Active Cases coincide in the limit with the continuous SEIR model. Indeed, they differ only by the amount $R_P - R_{P+1}$ which is supposed to be in general small.

Conclusion 8 *It is important that we have found a nice correspondence between the discrete curve I_t and the curve of the Modified Active Cases (in both the discrete SEIR model and in the empirical data). We have already said that there is a drawback in the usual definition of the empirical Active Cases, since it does not correspond nicely to the discretized SEIR model. Although it is expected that practically the two formulas are very close (since their difference $R_P - R_{P+1}$ is supposed to be small), at least for a rigorous numerical analysis it would be better to replace the usual notion of Active Cases with the notion of Modified Active Cases, if we want to have a smoothly working simple algorithms.*

Perhaps the above drawback may be resolved by a more sophisticated discretization of the SEIR model, which would though make the algorithms more complicated.

3.6 The modeling of vaccinations

It is possible to introduce without big effort the vaccinations in our continuous and discrete SEIR models.

There is a bunch of papers devoted to the modeling of the vaccination via SIR and SEIR models, as well as to optimal control of the vaccination process. We refer to the following sources: Biswas et al. (2014), Biswas et al. [2014], Neilan and Lenhart (2010), Neilan and Lenhart [2010], Brauer and Castillo-Chavez (2001) Brauer et al. [2012], Brauer et al. (2008) Brauer [2008].

We will consider very simplified setting and parameters of the vaccinations. We will reformulate the SEIR model by introducing a vaccination process. For simplicity, we assume that the vaccine is 100% effective so that all vaccinated susceptible individuals become immune; technically this may be achieved by multiplication of an efficiency coefficient. We will also assume that the immunity lasts forever; technically this may be achieved by revaccinations, which we will not add at the end of the day.

We will denote by $u(t)$ the percentage of susceptible individuals $S(t)$ being vaccinated per unit of time. Another formulation is to use directly the number $V(t)$ of the vaccinated persons, i.e. $V(t) = u(t) S(t)$. In particular, if we vaccinate $30,000$ people per month, then we will have to subtract $30000/30 = 1000$ persons on the day t from the number of the susceptible individuals $S(t)$. Hence, the modification of the SEIR model (2.1)-(2.5) becomes:

$$S'(t) = -\frac{\beta(t) S(t) I(t)}{N} - V(t) \tag{3.7}$$

$$E'(t) = \frac{\beta(t) S(t) I(t)}{N} - \sigma E(t) \tag{3.8}$$

$$I'(t) = \sigma E(t) - \gamma(t) I(t) \tag{3.9}$$

$$R'(t) = \gamma(t) I(t) + V(t) \tag{3.10}$$

The discrete counterpart of the above system of equations is a modification of the corresponding discretized SEIR system (2.7)-(2.10):

$$S_{n+1} = S_n - \beta_n S_n I_n / N - V_n \tag{3.11}$$

$$E_{n+1} = E_n + \beta_n S_n I_n / N - \sigma E_n \tag{3.12}$$

$$I_{n+1} = I_n + \sigma E_n - \gamma_n I_n \tag{3.13}$$

$$R_{n+1} = R_n + \gamma_n I_n + V_n \tag{3.14}$$

3.7 The ATVBG-SEIR model

In order to make an adequate SEIR model, and an appropriate algorithm, we take into account the main characteristics of the COVID-19 epidemic.

1. Our approach to modeling the dynamics of the epidemic is to use a *highly simplified* SEIR model with variable parameters, which we call *Alternating Time Varying Beta-Gamma-SEIR* **(**abbreviated as **ATVBG-SEIR)** (it reminds of the bang-bang controls in the theory of Optimal Control). The model has only two alternating regimes: **Strict regime** *(mimicking strict containment measures)* and **Relax regime** *(mimicking weakened containment measures)* with varying rates β, γ that are alternately piece-wise constants as explained below.

2. The main constraint is the upper limit of **Intensive Care Unit** beds (**ICUs**) reserved for COVID-19 patients, which in the case of Bulgaria does not exceed 1100, while in Germany they are at least $8,000$ (with a total limit of $40,000$ ICUs), $2,764$ in Italy, $26,900$ in USA, etc.[1]

[1]It does not seem that there are very precise data about the upper limit of ICUs reserved for Covid-19 cases since different sources show different numbers.

3. For simplicity, we choose the *ATVBG-SEIR model* to use fixed periods of **about 21** days (or more) for the *Strict regime*s, while the periods for *Relax regimes* are determined iteratively by our algorithm upon hitting the ICU beds upper limit.

4. The most time consuming and not obvious task in constructing the ATVBG-SEIR model is to identify the Strict regime and the Relaxed regime. In sections 3.12-3.14 below we show in more detail our experience in identifying Strict and Relaxed regimes in Bulgaria and Germany. Without loss of generality, a possible *Relax* ***regime*** for Bulgaria was obtained after calibration of the standard SEIR model (constant parameters) for the period *1.10.2020—25.11.2020* during which schools, kindergartens, restaurants, bars were open. It has a basic reproductive number $\mathcal{R}_0 \approx 4$. Another possibility is to take the period *15.09.2020—15.10.2020,* which has basic reproductive number $\mathcal{R}_0 \approx 2$. For the last period it is important that it is immediately after partial containment measures , while the former time period stretches over a large period of relaxed measures.

5. We have fewer opportunities to calibrate the **ATVBG-SEIR** model during a **Strict** regime since strict measures were imposed in Bulgaria during the period March-May, 2020, at the beginning of the epidemic, but a lot of Bulgarians guest workers have arrived from abroad. Hence, no reasonable calibration of the classical SEIR model could be made relying just upon the official data without taking into account the social dynamics.

6. We chose several **Strict** regimes for Bulgaria. One of them has a basic reproduction number $\mathcal{R}_0 = 0.9$ and is obtained on the basis of data for the period *1.4.2020-15.05.2020.*

7. However, we have considered also **Strict** regimes with reproduction numbers that are much less than one (i.e. we have a declining of the epidemic): $\mathcal{R}_0 \approx 0.8,\ 0.5,\ 0.2$

EXAMPLE: Below we provide an example of *ATVBG-SEIR* model parameters for the period '*12-Dec-2020*' till '*11-Feb-2022*'. We have displayed the graphs of the parameters: $\beta(t)$ (**blue**) and $\gamma(t)$ (**red**). For each day the pair of numbers $(\beta(t), \gamma(t))$ **alternately** takes the following pairs of values

$$StrictRegime = (0.4649,\ 0.8415)$$
$$RelaxRegime = (0.1276,\ 0.0279)$$

i.e. $\beta(t)$ **takes alternately** the values 0.4649 and 0.1276, and $\gamma(t)$ **takes alternately the values** 0.8415 and 0.0279. In this example, the periods of the *Strict regime* are **21 days**, as seen in Figure 3.1

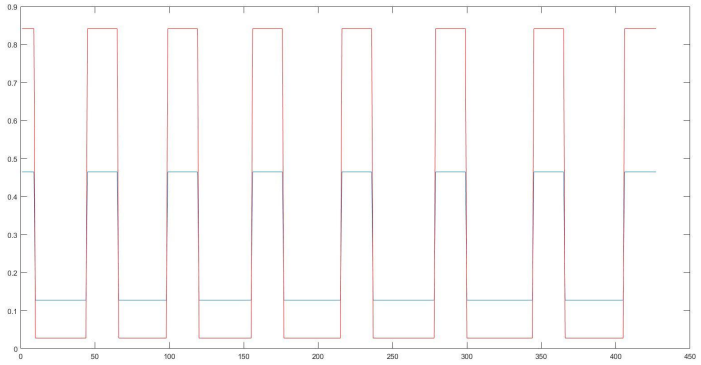


Figure 3.1: An example of parameters $\beta(t)$ (blue) and $\gamma(t)$ (red) for the ATVBG-SEIR model, for the period 12-Dec-2020 till 11-Feb-2022.

3.8 Determining the number of ICUs and the Algorithm

Below we discuss the most important elements of our Algorithm:

1. Empirical data to date show that the number of ICUs occupied by critical COVID-19 cases is very strongly correlated with the so-called **Active cases**. The Active cases are an important notion since the course of the COVID-19 disease has very different duration for different people. Practically, the number of Active cases is calculated on the ground of the official daily data for each day t by the formula

$$\begin{aligned} AC\,(t) \;&=\; Total\ number\ of\ infected \\ &-\; Total\ number\ of\ Recovered \\ &-\; Total\ number\ of\ Fatalities \end{aligned}$$

2. According to the **empirical** observations, the number of patients with COVID-19 accommodated in ICU beds varies; e.g., for Bulgaria it is about 0.55%, and for Germany about 1.35%, of the Active cases. In the implementation of our algorithm we have complied with these percentages.

   Our main goal is to account for the number of COVID-19 ICU cases in the framework of the **ATVBG-SEIR model**. We have seen that the Active cases within the continuous SEIR model are determined by the formula

$$AC\,(t) = TotalInf(t) - R\,(t) = \int_0^t \sigma E\,(s)\,ds - R\,(t)$$

and we have justified an analog for the discrete model ATVBG-SEIR, where we use the Modified Active Cases. Hence we put

$$ICU\,(t) = 0.55\% * AC\,(t)$$

in the ATVBG-SEIR model for Bulgaria. Here the percentage 0.55% has been obtained as a mean value for a sufficiently long period of time.

3. Another important point supporting the realism of the results is, that by the opinion of many medical experts, the infected cases are at least $3-4$ times more than the officially registered. In our model we take this into account. This phenomenon is easily explained, since many people with a positive PCR test and people with symptoms which have not been medically investigated do not enter the official statistics, and are not a subject to mandatory isolation and control by health authorities; this issue has been discussed in various sources. The escalation coefficient $3-4$ or a different one is an important input from the experts in our Algorithm.

4. According to the experts, the epidemic ends if 70% of the population has acquired immunity, by means of vaccination or previous health history. In our algorithm we have set this condition as a **stopping rule** for the iterations of the algorithm, which upon fulfilling automatically determines the duration of the model ATVBG-SEIR. However this stopping rule cannot be always fulfilled due to the existence of equilibrium state. In the section 3.9 below we will discuss another stopping rule, which becomes active if the above rule is not applicable.

3.9 Equilibrium of SEIR models and the end of an epidemic

As we said above, it is a common truth among the specialists that the immunity of 70% of the population N marks the end of an epidemic. This has been discussed in the recent paper Gomes et al. (2020), Gomes and et al. [2020].[2]

However, when applying SIR or SEIR models, it is not guaranteed that such a 70% threshold may be attained without vaccine, due to some fundamental properties of these models, which are related to the existence of equilibrium states (see e.g. Keeling and Rohani [2008]). Indeed, one may

[2] In the same publication it is proved that the worst case scenario from the point of view of epidemic size and duration is the homogeneity of the population network, i.e. every heterogeneity of the (social connections) network reduces the size and the duration of the epidemic.

prove rigorously that for $t \longrightarrow \infty$ there is an asymptotic value of the removed cases $R(t)$, which is denoted by R^*. More generally, for $\mathcal{R}_0 > 1$, the so-called *endemic equilibrium* exists and is defined by the limiting constants S^*, E^*, I^*, R^*, where

$$S^* = \frac{1}{\mathcal{R}_0}, \quad E^* = 0, \quad I^* = 0, \quad R^* = 1 - \frac{1}{\mathcal{R}_0}$$

and a zero mortality rate (= birth rate) is adopted, for details see Keeling and Rohani [2008], p. 42. Hence, if the population under consideration has N^* individuals, then the maximal percentage which may be removed is

$$R^* N^* = \left(1 - \frac{1}{\mathcal{R}_0}\right) N^*$$

Since we work with a model starting at a part of the population, say with initial value $R(0)$, then we obtain $N^* = N - R(0)$ and we see that quantity

$$R\left(0\right) + \left(1 - \frac{1}{\mathcal{R}_0}\right) N^* = R\left(0\right) + \left(1 - \frac{1}{\mathcal{R}_0}\right) \left(N - R\left(0\right)\right)$$

has to be bigger than $0.7 * N$, if we want to satisfy the stopping criterion. Hence, the condition

$$1 - \frac{1}{\mathcal{R}_0} \geq \frac{0.7N - R\left(0\right)}{N - R(0)}$$

has to be satisfied if we want to remove 70% of the population N. We see that this condition is not satisfied below in Figure 3.6 (for Germany). Indeed, as seen in Figure 3.6, we have the last Relax regime starting on the date $T =$ *July 1, 2021*, and the final value of the removed is

$$R\left(T\right) = 2.955195 \cdot 10^7.$$

This value serves as $R(0)$ for the next interval [1 $July$, 2021, 4 $Jan.$ 2022] where we have the Relax regime parameters for a classical SEIR model. Since $R = 83 \cdot 10^7$ we see that

$$\frac{0.7N - R\left(0\right)}{N - R(0)} = 0.53412 > 0.5 = 1 - \frac{1}{\mathcal{R}_0}$$

This shows that we cannot satisfy the above condition

$$1 - \frac{1}{\mathcal{R}_0} \geq \frac{0.7N - R\left(0\right)}{N - R(0)}$$

hence, the algorithm will not stop its execution. This is the major motivation to introduce **another stopping rule** for our algorithm.

We introduce another criteria for decaying epidemic which may be considered as a practical end of the epidemic. In particular, if the number of newly infected cases forecasted by the model, namely $\sigma E\left(t\right)$, or the number of the Active cases (given by $I\left(t\right)$) fall under some threshold, this is a clear sign about a strongly suppressed epidemic, and might be considered practically as ended epidemic.

3.10 The algorithm

1. First of all, let us note that it is practically useless to take into account the whole information from the very beginning of the pandemic (since January or February, 2020) - we just need to take no more than three months of the data (a more detailed discussion would need a rigorous proof based on the theory of dynamical systems, see Katok-Hasselblatt (1999), Katok and Hasselblatt [1997]). We cut the data for the last three months and we denote the corresponding interval of time by $[T_1, T_2]$.

2. Thus we have the non-trivial problem to find the initial data for the SEIR model using the the officially announced data $TotalInf\,(t)$ (infectious cumulative data), $Idata\,(t)$ (today's new infections), $Rdata\,(t)$ (the recovered cumulative data), and $Deaths\,(t)$, (the fatalities cumulative data). We use the notion of Modified Active Cases, for defining the initial data for the discrete SEIR model for the interval $[T_1, T_2]$ where we consider T_1 as the initial point of the model.

We use the conclusions of Section 3.5 where we have discussed the correspondence between the empirical data and the discrete SEIR data. For defining $I0$ we use the Modified Active Cases:

$$\begin{aligned} E0 &= \left(\frac{1}{\sigma}\right) * Idata(1) \\ R0 &= Rdata\,(1) + Deaths(1) \\ I0 &= TotalInf\,(1) - Rdata\,(2) - Deaths(2) \\ S0 &= N - E0 - I0 - R0 \end{aligned}$$

Then we execute the algorithm as we have already explained until one of the stopping criteria is met.

We have already discussed the **stopping criteria** for the algorithm. As we have discussed above in Section 3.9, there are two conditions upon which the algorithm signals the end of the epidemic:

1. If at a certain moment t the removed cases calculated by the model $R(t)$ reach (or eventually go beyond) the threshold $0.7{*}N$ where N denotes the population.

2. If at a certain moment t the Active cases $I(t)$ calculated by the model reach (and fall under) a threshold value. The definition of this threshold value needs further study from empirical data.

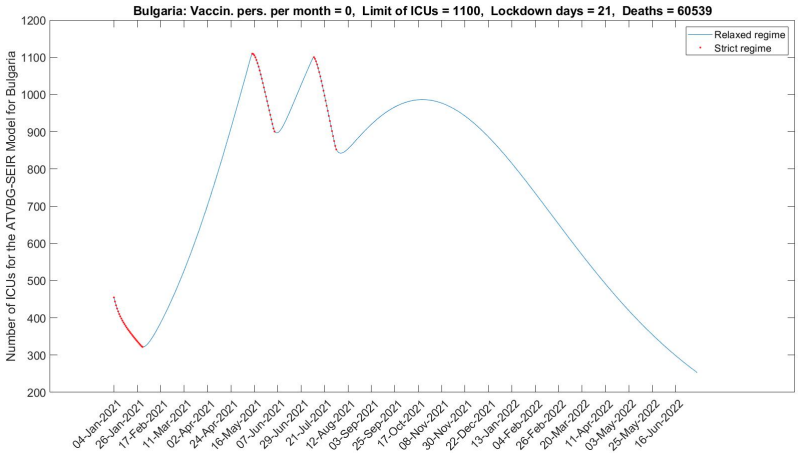


Figure 3.2: Optimistic scenario for Bulgaria - the curve of ICUs; no vaccinations.

3.11 Main results obtained with the Lockdown Scenarios Tool without Seasonal effect

3.11.1 Results for Bulgaria:

Following the *optimistic scenario* for Bulgaria (defined in Section 3.13 below), namely

$$\begin{aligned} RelaxRegime &= (\beta = 0.0436,\ \gamma = 0.0215)\,, \qquad \mathcal{R}_0 = 2.0, \\ StrictRegime &= (\beta = 0.0199,\ \gamma = 0.0288)\,, \qquad \mathcal{R}_0 = 0.6915 \end{aligned}$$

without vaccinations, the epidemic ends on *06-Jul-2022*, with two major lockdowns in the spring/summer, while 70% of the population got immune, and an estimated total number of fatalities $60,539$, see Figure 3.2.

For the same *optimistic scenario*, as seen from Figure 3.3, with vaccination plan $120,000$ persons ($240,000$ vaccinations) per month, the epidemic ends on *26-Dec-2021*, with one major lockdown, while70% of the population got immune, and an estimated total number of fatalities $45,591$:

On the other hand if we choose the *pessimistic scenario* (in Section 3.13 below), namely,

$$\begin{aligned} RelaxRegime &= (\beta = 0.0902,\ \gamma = 0.0159)\,, \qquad \mathcal{R}_0 = 5.7 \\ StrictRegime &= (\beta = 0.0218,\ \gamma = 0.0270)\,, \qquad \mathcal{R}_0 = 0.8056 \end{aligned}$$

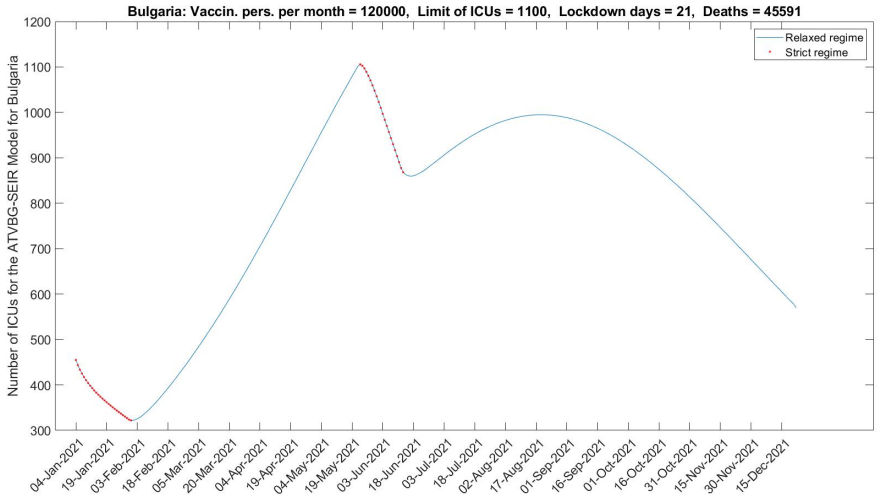


Figure 3.3: Optimistic scenario for Bulgaria - the curve of ICUs; with vaccinations.

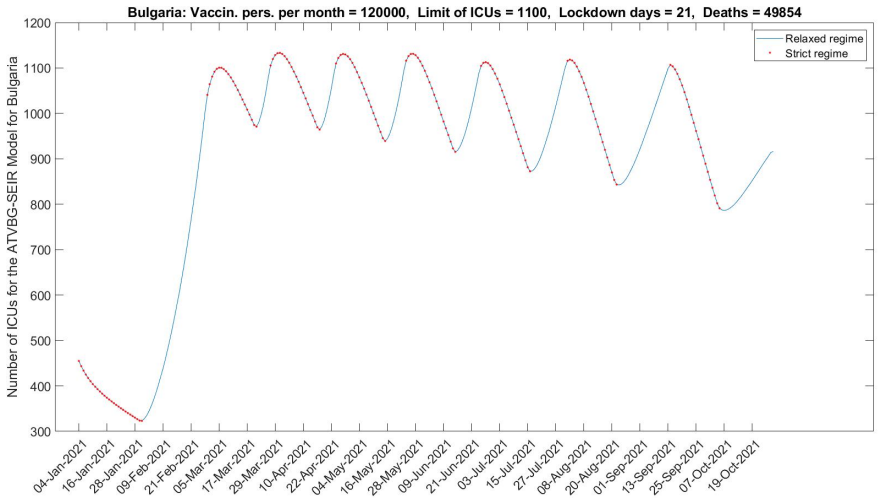


Figure 3.4: Pessimistic scenario for Bulgaria: the curve ICUs; with vaccinations

with vaccination plan120, 000 persons (240, 000 vaccinations) per month, the epidemic ends in November, with a series of Strict lockdowns, as seen from the Figure 3.4, and with an estimated total number of fatalities 49, 854:

3.11.2 Estimating the number of fatalities for optimistic/pessimistic scenario

The perspective provided by the two extreme scenarios is meant to compensate the lack of seasonal effect component in our models. Typically, the "pessimistic scenario" results in more lockdowns, though in a shorter duration of the epidemic. On the other hand, some main characteristics seem to remain rather stable for either scenario, in particular, the number of fatalities.

In order to obtain a rough estimate of the number of fatalities, i.e. the curve $Deaths\,(t)$ for the model ATVBG-SEIR, we recall that it does not participate directly in the model but is included in the removed cases, in the curve $R\,(t)$. We may avoid the consideration of a more complicated model of the type of SEIRD (where the variable D corresponds to the curve $Deaths\,(t)$) by the introduction of an empirical ratio. We have tested different approaches, and we have found that the usual mortality ratio

$$\mathcal{D}_0 \approx \frac{Deaths\,(t)}{TotalInf\,(t)} \tag{3.15}$$

for the empirical data is rather stable. The basic statistical analysis shows that in all cases of the countries which we have considered, the above mortality ratio has the least standard deviation compared to other ratios as

$$Deaths\,(t)\,/CumulativeICU\,(t) \quad \text{and} \quad Deaths\,(t)\,/Recovered(t).$$

For us it is important that for dates t beyond Today, we will be able to use the ATVBG-SEIR model, namely, we will calculate $TotalInf\,(t)$ by using the curve $E(t)$ of the discrete ATVBG-SEIR model, since it is the sum of $\sigma E(t)$. Hence, we may scenario-forecast the curve $Deaths\,(t)$; we obtain an estimate of the fatalities curve in the future by putting

$$Deaths\,(t) = \mathcal{D}_0 * TotalInf\,(t)$$

which we use practically only on the very last date t of the forecasted epidemic.

The ratio $\mathcal{D}_0$ depends on the particular country, and we take into account the escalation coefficient showing the real number of totally infected . Hence, we obtain that the ration $\mathcal{D}_0$ for Bulgaria has an average about 0.012 until January 2021. For the same period, for Germany, it is 0.011; for Austria – 0.008; for USA – 0.012, for Italy - 0.033, for UK - 0.032.

We have to say that what we obtain is an estimate of the curve $Deaths\,(t)$, which does not take into account that an adequate Vaccination plan targets on the first line those who are vulnerable to COVID-19 (with heavy accompanying diseases), and the expected fatalities will be definitely less.

We may compare our scenario-predictions with other resources in the Web, e.g. with the platform of the *IHME* institute, http://www.healthdata.org/.

The forecast of their product "Projections tool", available on the website shows about 11,000 cases on the date $t = \textit{June 1, 2021}$, see `https://covid19.healthdata.org/bulgaria?view=total-deaths&tab=trend`. However it is not clear how do they account for the vaccinations, see Figure 3.5.

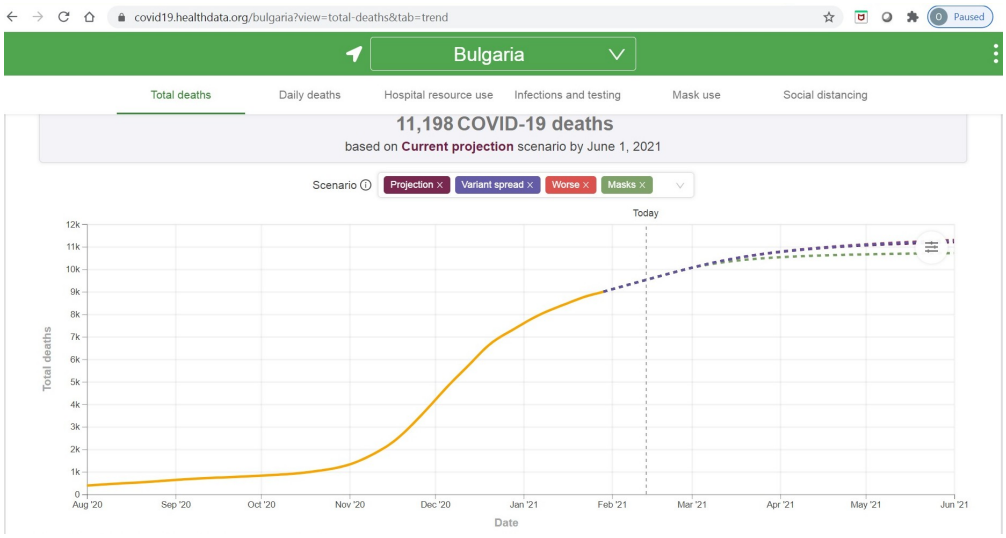


Figure 3.5: The prediction (projection) for the Total Fatalities, of the "IHME projection tool" for Bulgaria, with Today date March 1, 2021, and prediction for June 1, 2021.

On the other hand, our model shows, that if we assume the limit 1100 of the ICU beds in Bulgaria, and 120,000 vaccinated persons per month, then on date $t =$ June 1, 2021, the curve $Deaths\,(t)$ is equal to 30,261 cases (in the pessimistic scenario), or 24,245 cases in the optimistic scenario. No accounting for the priority of Vaccination plan for the vulnerable cases has been applied.

3.11.3 Results for Germany

To obtain results for Germany, the algorithm has applied the second stopping criterion for "End of epidemic" (discussed in sections 3.9, 3.10). Below is the Figure **without vaccinations**, where the algorithm has stopped since the number of ICUs of the model

$$ICU_{model}\,(T) = 198$$

on the final date $T = 4\ Jan.$ 2023 has become less than

$$0.01 * BEDS = 200$$

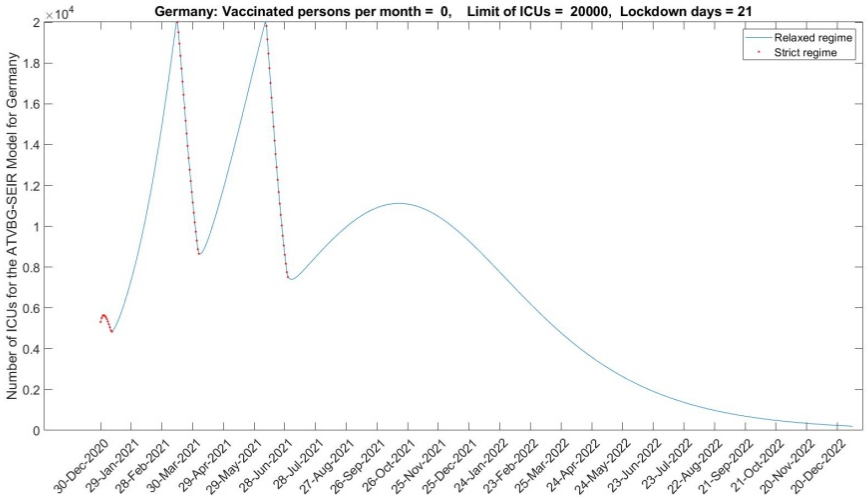


Figure 3.6: Germany, optimistic scenario, without vaccinations.

number of ICU beds. However on 4 Jan. 2023 less than 70% of the population have got immunity, see Figure 3.6.

With $1,400,000$ vaccinated persons per month we obtain the following Figure 3.7 showing that the epidemic ends on Nov. 7, 2021, while the end of epidemic is due to the first stop criterion - reaching the immunity of 70% of the population.

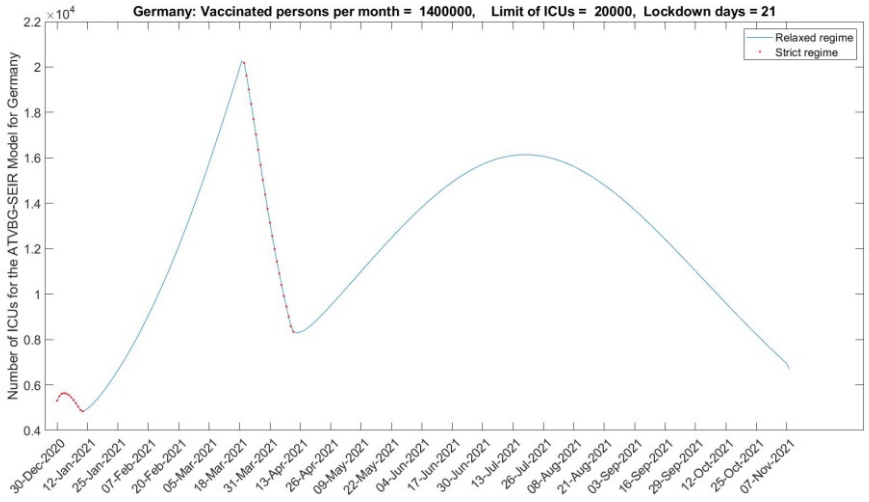


Figure 3.7: Germany, optimistic scenario, with vaccinations.

3.12 Identification of Strict and Relaxed regimes by fitting SEIR model

In order to find proper Strict regime and Relaxed regime parameters, we have fitted the classical SEIR model for different intervals of time in the history of the epidemic in Bulgaria, up to today. This resulted in different pairs of parameters $(\beta,\ \gamma)$ which reflect the real state of the society. One of the most important facts is that one may find the basic reproduction ratio $\mathcal{R}_0$ for some specific periods of time; this is called effective Reproduction ratio, denoted by $\mathcal{R}_e$. Below we provide the fitting of the SEIR model to different periods of time and the resulting parameters $(\beta,\ \gamma)$.

For finding the parameters β, γ we fit the model for a period of (strongly) Relaxed measures, with open schools, restaurants, social venues, etc. This was the period of time *30 Sep., 2020 - 25 Nov., 2020*. We obtained the following values

$$\beta = 0.1276,\ \gamma = 0.0279$$

hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \approx 4$$

The following Figures show the quality of the fitting which we consider as satisfactory:

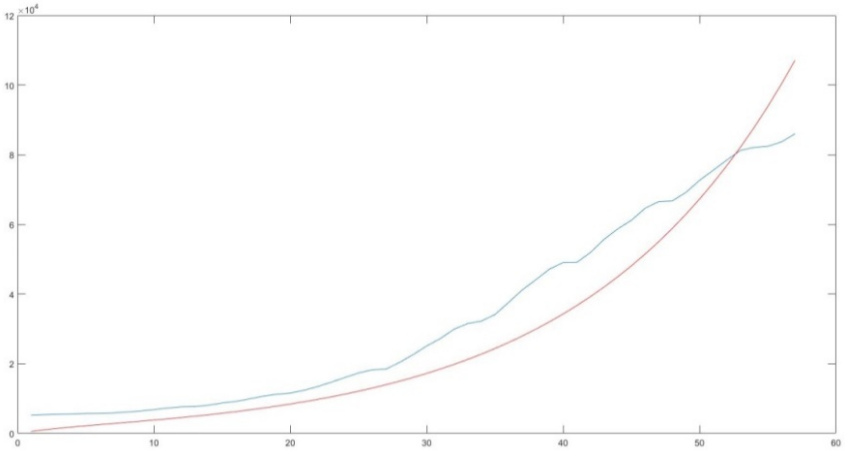


Figure 3.8: Comparison between the Active cases vs. the infectious curve $I(t)$ of the SEIR model.

3.12.1 Active cases data vs. $I(t)$ of SEIR model

In Figure 3.8 we provide a comparison between the Active cases vs. the infectious curve $I(t)$ of the SEIR model.

3.12.2 Recovered+Deaths data vs. Removed ($R(t)$) of SEIR model

In Figure 3.9 we provide a comparison between the Recovered and Deaths cases vs. the Removed curve $R(t)$ of the SEIR model.

Results for Austria, Italy, UK and USA are provided in the Tool, where one may choose the Relax and Strict regimes.

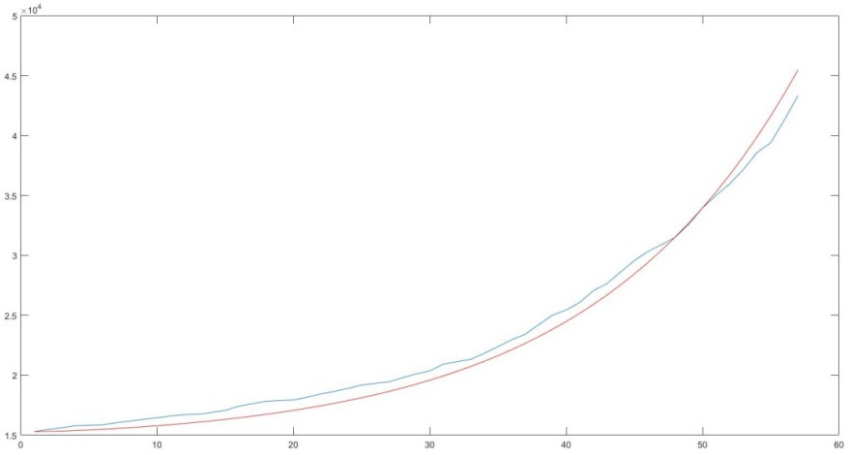


Figure 3.9: Comparison between the Recovered + Deaths cases vs. the Removed curve $R(t)$ of the SEIR model.

3.13 Fitting of the classical SEIR model to Bulgarian data

As we explained, one needs a serious research to find proper Strict regime and Relaxed regime which will constitute the pieces of the splines for the variable parameters $\beta(t)$, $\gamma(t)$ of the ATVBG-SEIR model.

We tried to find the parameters β, γ of classical SEIR in times of Strict measures, during the months April-August. However the very fact that the resulting basic reproduction ratio $\mathcal{R}_0$ is always bigger than one shows that something goes wrong. The only explanation is that there existed a very big incoming *flow of work emigration and students* (mainly Bulgarian citizens) from abroad.

1. During the period *6 April, 2020 – 15 May, 2020*, we have had very strict measures in Bulgaria. We have fitted the SEIR model and obtained the following parameters

   $$\beta = 0.0517,\ \gamma = 0.0156$$

   which give the basic reproduction ratio

   $$\mathcal{R}_0 = \frac{\beta}{\gamma} = 3.3118$$

2. During the period *15 May, 2020 – 31 July, 2020*, we had some relaxation – restaurants, etc. are open, but NO schools, and we have the

summer seasonal effect, we got the following parameters for the fitted SEIR model:

$$\beta = 0.0434,\ \gamma = 0.0254$$

which results in

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.7092 > 1$$

Again, the explanation is that many seasonal workers and Bulgarian students came from abroad and brought the infection.

3. Also for the period *15 May, 2020 – 31 August, 2020*, we obtained:

$$\beta = 0.0495,\ \gamma = 0.0327$$

which gives

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.5142 > 1.$$

4. For *1 June, 2020-15 Sept., 2020*, we have

$$\beta = 0.0685,\ \gamma = 0.0469$$

Hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 1.4601 > 1$$

5. However only close to September we got $\mathcal{R}_0 < 1$, which may be explained with a very strong seasonal effect. For the period *1 Aug., 2020-20 Sept, 2020*, we got the parameters

$$\beta = 0.0252,\ \gamma = 0.0363$$

Hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 0.6958 < 1$$

6. On the other hand, we have chosen as a **Strict** regime the one which we got in the winter, for *1 Dec., 2020-4 Jan., 2020*, namely

$$\beta = 0.0208,\ \gamma = 0.0291$$

Hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 0.7134 < 1$$

7. We have chosen as a **Relax** regime the one in the period *15 Sept., 2020 – 15 Oct., 2020*, namely

$$\beta = 0.0436,\ \gamma = 0.0215$$

Hence,

$$\mathcal{R}_0 = \frac{\beta}{\gamma} = 2.0$$

Based on the above we have defined one possible **Relaxed** regime by putting:

$$RelaxRegime = (\beta = 0.0902,\ \gamma = 0.0159)$$

For the **Strict regime** we have inspected different periods. For the period of officially introduced Strict measures (no schools, no restaurants, etc.), *1 Dec. – 29 Dec., 2020*, we have fitted the SEIR model and obtained the parameters

$$(\beta = 0.0199,\ \gamma = 0.0288)$$

with basic reproduction ratio

$$\mathcal{R}_0 = \frac{\beta}{\gamma} \approx 0.691.$$

Hence, we have put

$$StrictRegime = (\beta = 0.0199,\ \gamma = 0.0288)$$

In our research to find appropriate parameters for the Strict and Relaxed regimes, we have inspected different sub-intervals of the period *1 April – 15 Sep., 2020*.

For the period with partial measures (no schools, but open restaurants, etc.), *1 Aug – 15 Sep., 2020,* we obtained

$$(\beta = 0.0312,\ \gamma = 0.0367)\,, \qquad \mathcal{R}_0 = 0.8524$$

Further, for the period with relaxed measures (open schools, open restaurants, etc.), *1 Oct.-30 Oct., 2020*, we obtained the following parameters of the fitted classical SEIR model

$$(\beta = 0.0871,\ \gamma = 0.0162)\,, \qquad \mathcal{R}_0 = 5.4$$

and for the period with the same relaxed measures, *15 Sep.-15 Oct., 2020*, we obtained

$$(\beta = 0.0436,\ \gamma = 0.0215)\,, \qquad \mathcal{R}_0 = 2.0$$

As a result of the above, one may define several models which may be considered in the range of most optimistic to most pessimistic:

For **Bulgaria**:

The **optimistic** scenario is with the following parameters:

$$\begin{aligned} RelaxRegime &= (\beta = 0.0436,\ \gamma = 0.0215)\,, \qquad \mathcal{R}_0 = 2.0 \\ StrictRegime &= (\beta = 0.0199,\ \gamma = 0.0288)\,, \qquad \mathcal{R}_0 = 0.6915 \end{aligned}$$

The **mild** scenario:

$$\begin{aligned} RelaxRegime &= (\beta = 0.0599,\ \gamma = 0.0141)\,, \qquad \mathcal{R}_0 = 4.239 \\ StrictRegime &= (\beta = 0.0218,\ \gamma = 0.0270)\,, \qquad \mathcal{R}_0 = 0.8056 \end{aligned}$$

The **pessimistic** scenario:

$$RelaxRegime = (\beta = 0.0902,\ \gamma = 0.0159)\,, \qquad \mathcal{R}_0 = 5.7$$
$$StrictRegime = (\beta = 0.0218,\ \gamma = 0.0270)\,, \qquad \mathcal{R}_0 = 0.8056$$

For **Germany**:
The optimistic scenario:

$$RelaxRegime = (\beta = 0.0835,\ \gamma = 0.0419)\,, \qquad \mathcal{R}_0 = 2.0$$
$$StrictRegime = (\beta = 0.0337,\ \gamma = 0.0816)\,, \qquad \mathcal{R}_0 = 0.4125$$

The pessimistic scenario:

$$RelaxRegime = (\beta = 0.0946,\ \gamma = 0.0368)\,, \qquad \mathcal{R}_0 = 2.5713$$
$$StrictRegime = (\beta = 0.0337,\ \gamma = 0.0816)\,, \qquad \mathcal{R}_0 = 0.4125$$

For **UK**:
The optimistic scenario:

$$RelaxRegime = (\beta = 0.0435,\ \gamma = 0.0209)\,, \qquad \mathcal{R}_0 = 2.0818$$
$$StrictRegime = (\beta = 0.0134,\ \gamma = 0.0150)\,, \qquad \mathcal{R}_0 = 0.8968$$

The pessimistic scenario:

$$RelaxRegime = (\beta = 0.0499,\ \gamma = 0.0047)\,, \qquad \mathcal{R}_0 = 10.6923$$
$$StrictRegime = (\beta = 0.0134,\ \gamma = 0.0150)\,, \qquad \mathcal{R}_0 = 0.8968$$

For **Austria**:
The optimistic scenario:

$$RelaxRegime = (\beta = 0.1333,\ \gamma = 0.0790)\,, \qquad \mathcal{R}_0 = 1.6870$$
$$StrictRegime = (\beta = 0.0750,\ \gamma = 0.1130)\,, \qquad \mathcal{R}_0 = 0.6632$$

The pessimistic scenario:

$$RelaxRegime = (\beta = 0.1233,\ \gamma = 0.0574)\,, \qquad \mathcal{R}_0 = 2.1490$$
$$StrictRegime = (\beta = 0.0750,\ \gamma = 0.1130)\,, \qquad \mathcal{R}_0 = 0.6632$$

For **USA**:
The optimistic scenario:

$$RelaxRegime = (\beta = 0.0108,\ \gamma = 0.0058)\,, \qquad \mathcal{R}_0 = 1.8793$$
$$StrictRegime = (\beta = 0.0159,\ \gamma = 0.0243)\,, \qquad \mathcal{R}_0 = 0.6535$$

The pessimistic scenario:

$$\begin{aligned} RelaxRegime &= (\beta = 0.0141,\ \gamma = 0.0053)\,, \qquad \mathcal{R}_0 = 2.6426 \\ StrictRegime &= (\beta = 0.0159,\ \gamma = 0.0243)\,, \qquad \mathcal{R}_0 = 0.6535 \end{aligned}$$

For **Italy**:

The optimistic scenario:

$$\begin{aligned} RelaxRegime &= (\beta = 0.0317,\ \gamma = 0.0111)\,, \qquad \mathcal{R}_0 = 2.8627 \\ StrictRegime &= (\beta = 0.0258,\ \gamma = 0.0328)\,, \qquad \mathcal{R}_0 = 0.7885 \end{aligned}$$

The pessimistic scenario:

$$\begin{aligned} RelaxRegime &= (\beta = 0.0554,\ \gamma = 0.0100)\,, \qquad \mathcal{R}_0 = 5.5381 \\ StrictRegime &= (\beta = 0.0258,\ \gamma = 0.0328)\,, \qquad \mathcal{R}_0 = 0.7885 \end{aligned}$$

It should be noted that the final duration of the epidemic forecasted by the optimistic and the pessimistic scenarios is not that big; the main difference is that one needs more lockdowns (periods with Strict regime) in the case of the pessimistic scenarios.

3.14 Fitting of SEIR model to German data

We have carried out similar research of the German data.

As in the case of Bulgarian data, we have found appropriate periods with relaxed (*1 Sep. – 31 Oct., 2020*) and strict measures (*10 Apr – 10 June, 2020*), to which we have fitted the classical SEIR model. Thus we have chosen the following *StrictRegime* and *RelaxRegime* for the ATVBG-SEIR model for Germany:

$$\begin{aligned} StrictRegime &= (\beta = 0.0337,\ \gamma = 0.0816)\,, \qquad \mathcal{R}_0 = \frac{\beta}{\gamma} \approx 0.4125 \\ RelaxRegime &= (\beta = 0.0835,\ \gamma = 0.0419)\,, \qquad \mathcal{R}_0 = \frac{\beta}{\gamma} \approx 2.0 \end{aligned}$$

3.15 Lockdown Scenarios Tool – a short description

The above results have been obtained by the specially designed Lockdown Tool, at `http://atvbg-seir.eu`.

We have implemented the methodology of the ATVBG-SEIR model in the Lockdown Tool, where the end-user may play with several parameters which is the best way to get a feeling of the strengths of the model to make forecasts.

1. **P1: (C3)** (default value is **3**, input by a Slider) – this is the escalation parameter, which is an escalation factor showing the ratio between the unreported cases and reported cases, i.e.

$$\textbf{All cases} \ = \ C3 * \textbf{Reported cases}$$

2. **P2: (C21)** (default is **21**; input by a Slider; min=15, max=45). This is a fixed number of days showing the lengths of the **lockdown** periods (of Strict measures).

3. **P3: (C06)** (default for Bulgaria is 0.6% = 0.006, for Germany is 1.4% = 0.014, for Italy is 0.5% = 0.005; input by a Slider; min = 0.1% and max = 3%). This number shows the ICUs as a percentage of the Active Cases. We find it by a regression on the data for every country, during the last three months; data available at `https://www.worldometers.info/coronavirus/#countries`

4. **P4: (BEDS)** (default for Bulgaria is 1,100, for Germany is 10,000; input by a Slider). This number shows the upper **limit** of ICU beds reserved for the COVID-19 cases. It is announced in the database `https://www.worldometers.info/coronavirus/#countries`, but seems to be variable for every country since the authorities always keep higher reserves for emergencies. It is difficult to download true data about the daily occupancy of the ICU beds.

5. **P5: (C70)** (default value is 70% = 0.7; input by a Slider; min = 40%; max = 90%). This number is the percentage of the population that has to be immune, for the epidemic to be assumed ended.

6. **P6: (C01)** (default value is 1% = 0.01; input by a Slider; min = 0.2%, max = 5%). This is the percentage of the upper Limit of ICUs. If the model requires number of ICUs less than this, we assume end of the epidemic.

7. **P7: (VACCINATIONS)** (default value is **0**; Input in a box). This is the total number of vaccinated **persons** per month (as usually planned); let us remind that every vaccinated person gets two vaccinations in a short time.

8. **P8: (VACCINATIONS PLAN)** - A table to enter the number of **vaccines** (NOT vaccinated people) delivered per month. It is though assumed as in **P7** that a vaccinated person gets two vaccines in a short interval of time.

9. **P9**: **(User Defined Scenario**) – The user may enter the values of *StrictRegime* and *RelaxRegime* for every particular country. We

have found the values of some optimistic and pessimistic scenarios for Bulgaria, Germany and other countries for which we have found sufficiently good data in section 3.13 above.

The main parameters which may be manipulated are P4 and P7–P8, and they are strongly influencing the dynamics of the epidemic in a single country. The reason is that in critical situations health authorities are apparently forced to take additional ICU beds from the reserve (for other diseases) and give them to COVID-19 patients. Varying the Vaccination plan by the parameters P7 or P8 may be very helpful for optimizing the future vaccination strategy.

On the other hand, we have added in the Tool a dropdown list of optimistic/pessimistic scenarios for every country, constructed by using the list of several pairs of *RelaxRegime* with different basic reproduction numbers $\mathcal{R}_0$.

3.16 Final Remarks

There are two main factors which may reduce the duration of the epidemic provided by our models.

1. In principle, in our study we do not account seasonal effects. An excuse is that while in the winter the immunity of the people is very low, it is very high in the summer, which causes a compensating effect. However we compensate this by providing optimistic and pessimistic scenarios.

2. We also ignore the fact that the immunization targets on the first line many super-spreaders, which may very essentially influence the dynamics and reduce the duration of the epidemic.

3.17 Modeling of the Seasonal effect for COVID-19 spread

Due to the lack of sufficient information at present, the best approach to modeling the seasonal behaviour of the Covid-19 spread appears to be a direct analogy with the seasonal behaviour of the standard Flu, see the review article Audi et al. [2020].

A large number of scientific papers have investigated the rich nonlinear effects caused by periodically varying contact rates $\beta(t)$ in *epidemic models*, and some excellent reviews exist, see e.g. Altizer et al. [2006], Grassly and Fraser [2006].

The modeling of the seasonal effect for the standard *Flu* (Flu peak and Flu low seasons) appears to be a very non-trivial task and there is a

huge amount of research in the area, see David et al. [2000], Buonomo et al. [2018]. However it seems that there is no final satisfactory solution, and the seasonality of these infection spreads is still under active research. The main problem is to compare the amplitude of the infectious curve during the peak season and the level of the infectious curve during the low season. However not less trivial problem is to determine which is the best shape of the curve during the low season, but also to determine the borders betweent the peak and the low seasons.

The first acceptable from epidemiological point of view model, was designed by Dietz Dietz [1976], Buonomo et al. [2018]. He was the first to investigate the effects of one-year periodic contact rate $\beta(t)$ in the classical SIR and SEIR models. He considered a periodically varying contact rate $\beta(t)$ given by the model:

$$\beta(t) = \beta_m (1 + A\cos(\omega t)).$$

Here, the parameter A measures the degree of seasonality of the contact rate. Further, more complicated models were considered by others, e.g.

$$\beta(t) = \beta_0 (1 + \beta_1 \phi(t))$$

where the degree of the seasonality was measured by the function

$$\phi(t) = \frac{0.68 + \cos(\omega t)}{1 + (2/3)\cos(\omega t)}.$$

In the Science paper David et al. [2000], another version of the above was used:

$$\beta(t) = \beta_0 (1 + \beta_1)^{\phi(t)}.$$

From this simple model very interesting conclusions were made: It is not necessary to complicate the analysis of the measles among children by explicitly modeling age structure in the host population, namely, term-time forced SEIR model behaves almost identically to favored about year 2000 age-structured models (the year of the publication of David et al. [2000] is 2000), indicating that the critical ingredient in measles models is a realistic seasonal forcing function rather than explicit modeling of heterogeneous transmission.

Others use various simple models, as in Irena et al. [2019], which we provide visualized in Figure 3.10:

The difficulty to determine the amplitude of the infectious curve for a particular year is clear from the following Figure 3.11, (available at the CDC web-site, of Health and Services [2020], "Missouri Weekly Influenza Surveillance Report 2018-2019 Influenza Season"), which provides the infectious curve for the years $2015 - 2019$ in Missouri:

In case we do not need a precise estimate but just a crude one, we may use a simple shift of the curves during the summer season with about $4.5 - 5$

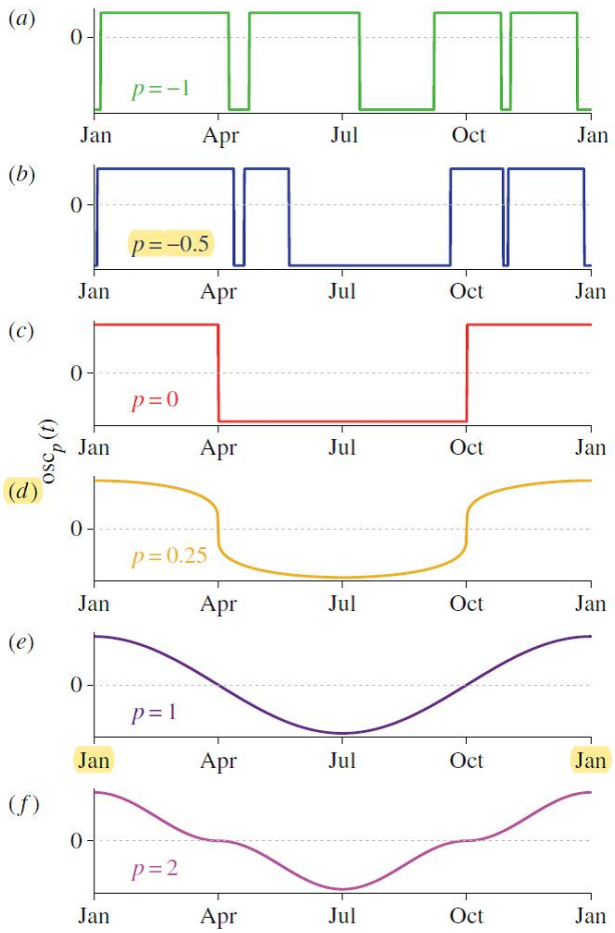
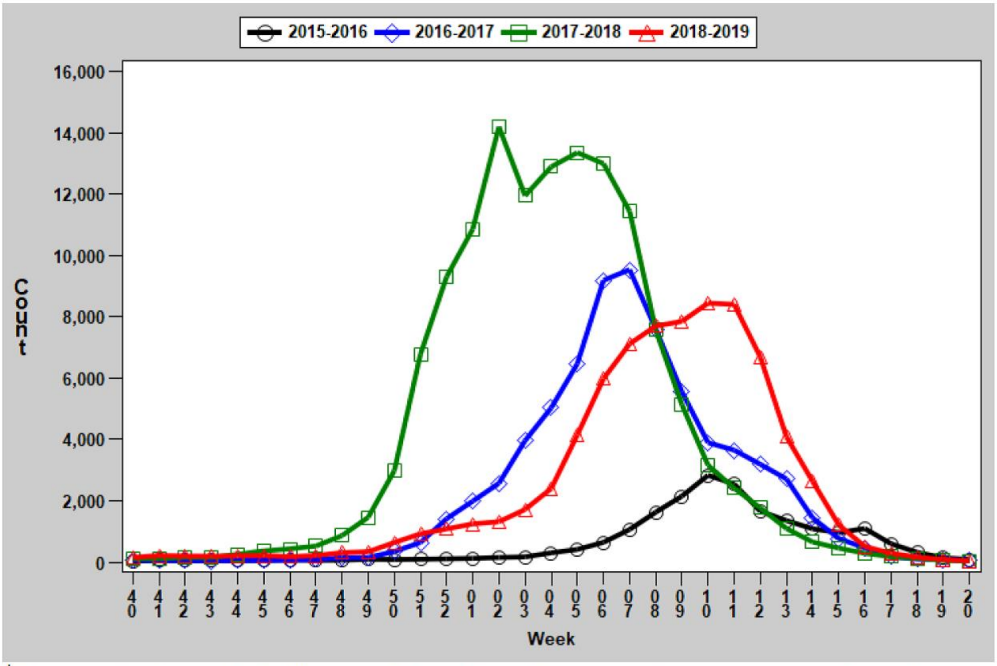


Figure 3.10: Various seasonal models.

Figure 4. Number of Laboratory-positive† Influenza Cases by CDC Week, Missouri, 2015-2019*



†Laboratory-positive influenza includes the following test methods: rapid influenza diagnostic tests (antigen), reverse transcriptase polymerase chain reaction (RT-PCR) and other molecular assays, immunofluorescence antibody staining (Direct (DFA) or Indirect (IFA)), or viral culture.
*2018-2019 season-to-date through the week ending May 18, 2019 (Week 20).Data Source: Missouri Health Information Surveillance System (WebSurv).

Figure 3.11: Influenza cases in Missouri for the years 2015-2019.

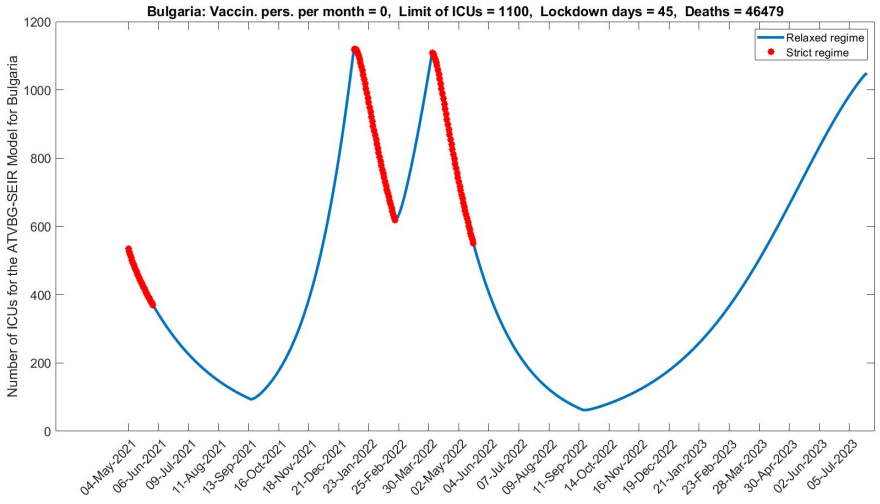


Figure 3.12: Lockdown Scenario Tool, for Bulgaria, starting on May 4, 2021,

months. A related rough estimate of the number of infectious cases and number of fatalities is also not difficult to calculate.

However, another more sophisticated but still simple to implement approach to the seasonality gives a reasonable approximation to the real picture, and extends the ATVBG-SEIR model. Namely, we assume that during the low season *15-May-2021 – 15-sep-2021,* the "summer effect" is **verysimilar** to the effect caused by the Strict Containment measures, i.e. we use the parameters $\beta(t)$ and $\gamma(t)$ in the ATVBG-SEIR model to be the same as in the $StrictRegime = (\beta = 0.0199,\ \gamma = 0.0288)$, provided above in section 3.13.

By applying the Lockdown Scenario Tool, for Bulgaria, starting on May 4, 2021, we obtain the resulting curve of ICUs on Figure 3.12.

We see that there appears a "slowing period" of about $4-5$ months for the summer of 2021 and 2022.

3.18 Modeling of the Loss of Immunity (after 9 months)

An important problem for the long-term control of the further development of the COVID-19 pandemic is to proper model the effects of the Limited immunity due to either vaccination or to the natural course of the disease.

The data from Wuhan ([Nature, 2021, see Bibliography of Chapter 1], [He et al., 2021, see Bibliography of Chapter 1]) show that the natural

immunity lasts for about 9 months for the majority cases. On the other hand the immunity due to the vaccinations is quite disputable – some sources say it lasts for 5 months to 12 or more.

We have to reconsider the system of equations in the SEIR model: First of all we have the data $Rdata(t)$ and $Deaths\,(t)$ which are usually considered as representative until the StartDate for the SEIR model. On the other hand, after the $StartDate$ (of our SEIR model), we consider the variables S, E, I, R. We may extend the whole series of S_{n+1} by defining

$$S_n = Sdata_n, \qquad \text{for} \quad n \leq StartDate;$$

Here, for $n < StartDate$, we have put

$$\begin{aligned} Sdata_n &= N - Rdata_n - Deaths_n - AC_n \\ Rrec_n &= Rdata_n, \end{aligned}$$

Let us recall that in our algorithm we have assumed that the number fatalities will be determined by the formula

$$Deaths\,(t) = \mathcal{D}_0 * TotalInf\,(t)$$

As we have decided for our model SEIR to have a coefficient $\mathcal{D}_0$ (see (3.15)) we put

$$Rrec_n = R_n - Deaths_n, \qquad \text{for} \quad n \geq StartDate$$

Thus we have defined the time series $Rrec_n$ for all times n, i.e. for $n < StartDate$ and for $n \geq StartDate$.

We assume that the loss of **natural immunity** (after the course of disease) is after *nine months* which are roughly 270 days but the loss of the **vaccination immunity** is after *six months*. Now we reconsider our SEIR system with Limited Immunity condition:

$$\begin{aligned} S_{n+1} &= S_n - \frac{\beta_n S_n I_n}{N} - V_n + Rrec_{n-270} + V_{n-180} \\ E_{n+1} &= E_n + \frac{\beta_n S_n I_n}{N} - \sigma E_n \\ I_{n+1} &= I_n + \sigma E_n - \gamma_n I_n \\ R_{n+1} &= R_n + \gamma_n I_n + V_n - Rrec_{n-270} - V_{n-180} \end{aligned}$$

The simplest approach to render the above in a Matlab code is to take an absolute start as January 1, 2020 or so for all countries except for China, or even earlier, and put zeros before the real start of the data series.

For the Matlab code, we have to modify the basic function SEIR_spline where we have to use the Historical data in order to create the above new variables $Rrec_n$ and V_n.

Chapter 4

Software Platforms Used for the Web-based Tools

In the present chapter we provide more details of the implementation of the Web-based instruments.

Brief overview of the Jupyter Project

The name Jupyter is used for a project and a community for developing open source software and services for interactive computing. Project Jupyter started as a spin-off project from IPython developed by Fernando Perez in 2001 as an enhanced Python interpreter. Interactive Python or namely IPython is originally developed for the Python language as a command shell for interactive computing. The Jupyter system supports over 100 programming languages (called "kernels" in the Jupyter ecosystem) including Python, Java, R, Julia, Matlab, Octave, Scheme, Processing, Scala, and many more. Out of the box, Jupyter will only run the IPython kernel, but additional kernels may be installed. Software applications under Jupyter project are intended to support interactive data science and scientific computing. We provide the list of the main components which are contained in the Jupyter project:

- Jupyter Notebook - A web based interface to programming environments of Python, Julia, R and many others
- JupyterLab - Modern web based integrated interface for all notebooks, editors, consoles, etc.
- Jupyter Client - This is a service which contains the reference implementation of the Jupyter protocol. It is also a client library for starting, managing and communicating with Jupyter kernels.
- Jupyter kernels - These are execution environments for every programming language of Jupyter.

- IPykernel - Package that provides IPython kernel to Jupyter.
- QtConsole - Qt based terminal for Jupyter kernels similar to IPython.
- nbviewer - HTML viewer for Jupyter notebooks
- nbconvert - Converts Jupyter notebook files in other formats

For more details, we refer to Chapter 5 of the online book "Jupyter Notebook ecosystem", see `https://jupyter4edu.github.io/jupyter-edu-book/jupyter.html`.

On the other hand, IPython offers more features compared to the standard Python:

- Acts as a main kernel for Jupyter notebook and other front end tools of Project Jupyter.
- Possesses object introspection ability. Introspection is the ability to check properties of an object during runtime.
- Syntax highlighting.
- Stores the history of interactions.
- Tab completion of keywords, variables and function names.
- Magic command system useful for controlling Python environment and performing OS tasks.
- Ability to be embedded in other Python programs.
- Provides access to Python debugger.

Brief overview of the Jupyter Notebook

The Jupyter Notebook was developed as a concept of a computable document, see `https://ipython-books.github.io/chapter-3-mastering-the-jupyter-notebook/`. Using "notebook" or "notebook documents" as a web application one can create and share computable documents that contain live executable programming code, visualizations and interactive web elements. The Jupyter Notebook is one of the ideal tools to help you gain the web environment for performing data science analysis in real time.

The Notebook ecosystem: Jupyter notebooks are represented as JavaScript Object Notation (JSON) documents. JSON is a language-independent, text-based file format for representing structured documents. As such, notebooks can be processed by any programming language, and they can be converted to other formats such as Markdown, HTML, LaTeX/PDF, and others.

There is an ecosystem of tools around the Notebook. Notebooks are being used to create slides, teaching materials, blog posts, research papers, and even books.

Architecture of the Jupyter Notebook: Jupyter implements a two-process model, with a kernel and a client. The client is the interface offering the user the ability to send code to the kernel. The kernel executes the code and returns the result to the client for display. In the Read-Evaluate-Print Loop (REPL) terminology, the kernel implements the Evaluate, whereas the client implements the Read and the Print of the process.

The client can be a Qt widget if we run the Qt console, or a browser if we run the Jupyter Notebook. In the Jupyter Notebook, the kernel receives entire cells at once, so it has no notion of a notebook. There is a strong decoupling between the linear document containing the notebook, and the underlying kernel.

All communication procedures between the different processes are implemented on top of the ZeroMQ (or ZMQ) messaging protocol (`http://zeromq.org`). The Notebook communicates with the underlying kernel using WebSocket, a TCP-based protocol implemented in modern web browsers.

Connecting multiple clients to one kernel: In a notebook, typing connect-info in a cell gives the information we need to connect a new client (such as a Qt console) to the underlying kernel.

JupyterHub: JupyterHub, available at `https://jupyterhub.readthedocs.io/en/latest/`, is a Python library that can be used to serve notebooks to a set of end-users, for example students of a particular class, or lab members in a research group. It handles user authentication and other low-level details.

Security in notebooks: It is possible for an attacker to put malicious code in a Jupyter notebook. Since notebooks may contain hidden JavaScript code in a cell output, it is theoretically possible for malicious code to execute surreptitiously when the user opens a notebook. For this reason, Jupyter has a security model where HTML and JavaScript code in a notebook can be either trusted or untrusted. Outputs generated by the user are always trusted. However, outputs that were already there when the user first opened an existing notebook are untrusted.

The security model is based on a cryptographic signature present in every notebook. This signature is generated using a secret key owned by every user.

History and Development of IPython and Jupyter System

IPython was originally developed by Fernando Perez in 2001. Its current version is IPython7.0.1 which requires Python 3.4 version or higher.

IPython 6.0 was the first version to support Python 3. Users having Python 2.7 should work with IPython's version 2.0 to 5.7 The concept of computational notebooks started in 80s decade when MATLAB and Mathematica were released. These GUI frontends to the interactive shell had features like text formatting, adding graphics, table and adding mathematical symbols. SAGE notebook is also a web based notebook. Creaters of IPython started working on notebook interface for IPython shell in 2005. IPython notebook soon added support of other languages like R and Julia. It was in 2014, that Perez started Jupyter project as a spin-off project from IPython, since IPython project was becoming big with products like notebook server and Qt console added to it. Since IPython 4.0, all additional components were shifted to Project Jupyter and adding support of other languages to IPython notebook. IPython continues to focus on improvement of its enhanced interpreter feature. It also provides primary kernel to Jupyter notebook frontend.

Brief overview of the Bokeh Library and Environment

Bokeh library and Environment are competitive to Jupyter NOtebook, see `https://docs.bokeh.org/en/latest/docs/reference/server.html`. However, Bokeh is not only competitive but may be also combined with Jupyter Notebook.

Thus, we used Bokeh for the frontend part of the implementation – it includes libraries with widgets, controls, interactive UI elements: "bokeh.server" is the analog to the Jupyter Notebook. It provides a customizable Bokeh Server Tornadocore (webserver) application, popular among the web-programming people. The architecture of Bokeh is such that high-level "model objects" (representing things like plots, ranges, axes, glyphs, etc.) are created in Python, and then converted to a JSON format that is consumed by the client library, BokehJS. By itself, this flexible and decoupled design offers advantages, for instance it is easy to have other languages (R, Scala, Lua, etc.) drive the exact same Bokeh plots and visualizations in the browser. The capability to synchronize between Python and the browser is the main purpose of the Bokeh Server. By far the most flexible way to create interactive data visualizations using the Bokeh server is to create Bokeh Applications, and serve them with a Bokeh server. In this scenario, a Bokeh server uses the application code to create sessions and documents for all clients (typically browsers) that connect. On the Figure below: A Bokeh server (left) uses Application code to create Bokeh Documents. Every new connection from a browser (right) results in the Bokeh server creating a new document, just for that session. The diagram of the Bokeh server is provided in Figure 4.1.

In our implementation we created a Bokeh application which is run NOT under Bokeh server but within a Jupyter Notebook. This required the creation of integration buffer and customized configuration. The application

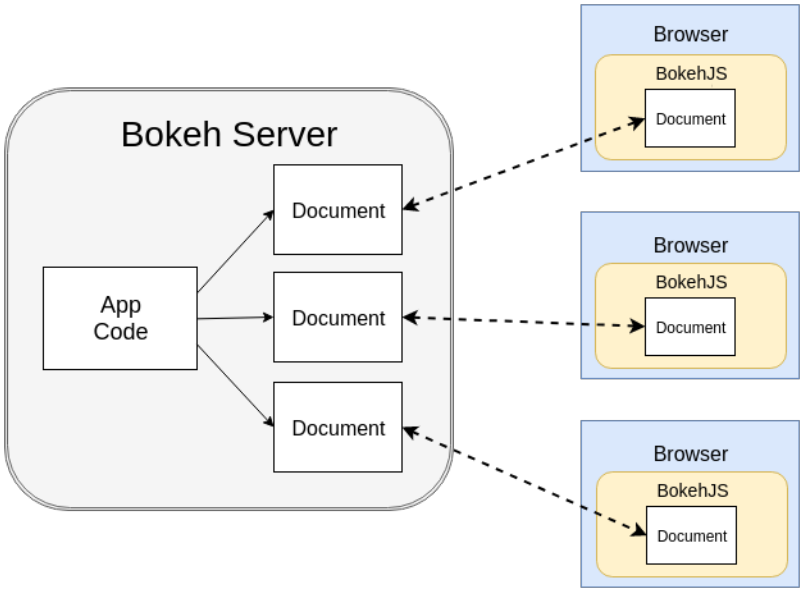


Figure 4.1: Bokeh server

code is executed in the Bokeh server every time a new connection is made, to create the new Bokeh Document (aka Notebook) that will be synced to the browser. The application code also sets up any callbacks that should be run whenever properties such as widget values are changes.

Running a Bokeh server (application): see the link `https://docs.bokeh.org/en/latest/docs/user_guide/server.html` The primary purpose of the Bokeh server is to synchronize data between the underlying Python environment and the BokehJS library running in the browser.

Bokeh server makes it easy to create interactive web applications that connect front-end UI events to running Python code. Bokeh creates high-level Python models, such as plots, ranges, axes, and glyphs, and then converts these objects to JSON to pass them to its client library, BokehJS. For more information on the latter, see the description of BokehJS. This flexible and decoupled design offers some advantages. For instance, it is easy to have other languages, such as R or Scala, drive Bokeh plots and visualizations in the browser.

However, keeping these models in sync between the Python environment and the browser would provide further powerful capabilities:

- respond to UI and tool events in the browser with computations or queries using the full power of Python
- automatically push server-side updates to the UI elements such as widgets or plots in the browser

- use periodic, timeout, and asynchronous callbacks to drive streaming updates

This is where the Bokeh server comes into play. The documentation of Bokeh is available at the link `https://github.com/bokeh/bokeh`.

Bokeh is not only an environment, but may be used as an interactive visualization library for modern web browsers. It provides elegant, concise construction of versatile beautiful graphics, and affords high-performance interactivity over large or streaming datasets. Bokeh can help anyone who would like to quickly and easily make interactive plots, dashboards, and data applications. With Bokeh, one can create JavaScript-powered visualizations without writing any JavaScript yourself. It is available in Anaconda Python distribution. Bokeh is a Sponsored Project of NumFOCUS, see `https://docs.bokeh.org/en/latest/index.html`.

The following references are very useful for initial encounter with Jupyter and Bokeh:

1. Karlijn Willems, Jupyter Notebook Tutorial: The Definitive Guide, at `https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook`

2. Tutorialspoint – Jupyter Tutorial, at `https://www.tutorialspoint.com/jupyter/ipython_introduction.htm`

Chapter 5

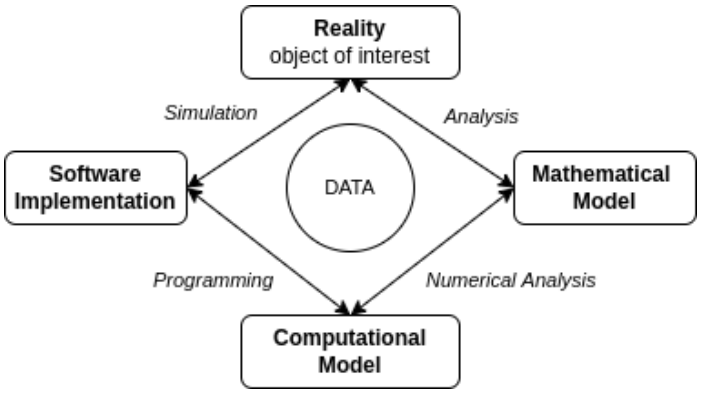
Software Implementation of the Tools

In the present chapter we provide more details about software development of implemented web based tools and some key concepts are given.

The Dissertation represents a contribution to Computational Epidemiology by modeling of contagious diseases in particular of COVID-19. Study of new models for analyzing of epidemic outbreaks including their software implementations as Web-based interactive instruments and tools for analysis, predictions and decision support.

5.1 Overview of some key concepts

When creating models of processes happening in real world the models are based on an understanding of the process leading to the observed data. Mathematical model is a simplified mathematical construct related to a part of reality. In other cases we can make difference between Data modeling and Mathematical modeling. One can discover and match an interesting pattern of data, but not able or with little possibility to explain the process by building a mathematical model and this modeling is referred to data modeling. Calibrating (finding values for parameters) for the models is required by running simulations / models to generate data and make fitting with the real and observed data. Software models are related to the development process and there are many different methodologies and software approaches.



5.2 Software approach

Test-Driven Development (TDD) is used for implementation of the tools described by following stages:

1. Write a test, watch it fail. If it passes, the code already covers the required functionality.
2. Write a code to pass the test. All previous tests have to pass. The new code adds to the existing functionality.
3. Refactor: revise code if necessary improve the code without changing behavior or any functionality at this stage.
4. Repeat 1-2-3

5.3 Software Application

Implemented tools can be used by users as a regular web applications, beyond that the tools have special mode with more complex structure as computable documents known as IPython (Jupyter) Notebooks and they are interactive runtime environments made of cells and group of cells. Each cell may contain both data and executable code. The Model View Controller (MVC) design pattern or something that looks like it (for example widgets in Bokeh) is also applicable for implemented tools and almost all UI-based applications.

5.4 Software Environment

The Web based tools are stored and computed on a server side. The access and running is provided by web browsers. The environment allows in parallel

multiple users to run independently their own experiments. If needed additional security restrictions could be enabled for user authorizations, access levels and etc. The implemented web based tools can be used as Software as Service within cloud computing platforms and provide all benefits like scalability and load balancing.

Bokeh library is chosen and used because its wide array of widgets, plot tools, and UI events that can trigger real Python callbacks, it allows to develop interactive web applications that connect front-end UI events to running Python code.

5.5 Data sources

The Data sources are provided as datasets / collections stored in various formats: CSV, XLSX, JSON. For some datasets is provided API interface / REST. Serial data is needed to be processed different order of columns and different labels are used. There are data sources storing data in rows instead of columns. Also used different formats of dates. All of this requires additional data processing.

5.6 Data pipeline

- Retrieve data – using program interface to connect and extract datasets for each country
- Process data – transform it into single input data format for TVBG / ATVBG tools.
- Verify data – check records for gaps / missing inconsistent data
- Data cleansing – / Remove or Correct gaps if it possible to use average / smooth
- Data generation – in order to find the best prediction scenarios it's required to run thousands times models which is heavy computational task and produce significant amount of data.

5.7 Input Data

The CSV format is chosen as major data format for the implemented web based tools. This is common widely used format for serial data / time series. If necessary format can be easily changed to any other suitable format for serial data.

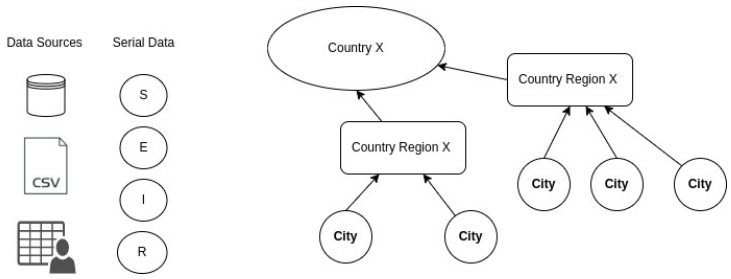
The main Input data is daily updated time series of the following:

- S – Susceptible

- E – Exposed
- I – Infectious
- R – Removed
- ICU – used in ATVBG-SEIR

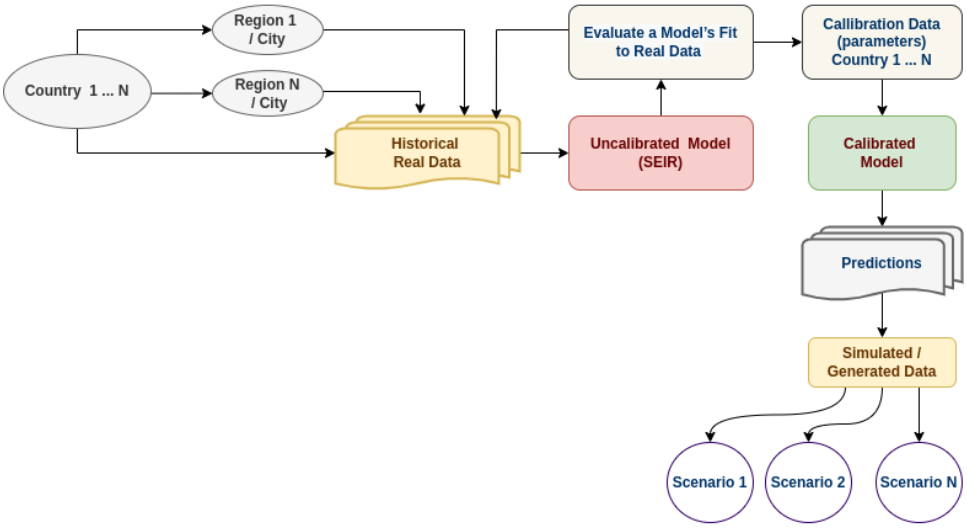
Model parameters are determined by the user or / and by fitting SEIR model with real data. For more detailed description see section for Compartmental SEIR models NumPy and Pandas are specialized Python libraries for storing and manipulating data series, multi-dimensional data. Pandas provides many different IO connectors to SQL, BigQuery, CSV, HDF5, JSON etc.

If needed the data could be scaled at different levels. It depends of data sources and size of regional units. For countries with large territory and population it is reasonable to do regional data modeling and apply the developed tools for predictions scenarios on separate regions.



5.8 Software Tools basic processing steps

The following figure shows back-end part of implemented web tools. It's possible to run tools in parallel for each country by different regions independently.



5.9 TVBG-SEIR Tool Web Interface

TVBG-SEIR tool with short-term SEIR model based on splines for COVID-19 prediction scenarios (see Chapter 2). The following is a part of web interface of the tool.

We assume that in the Country XX the COVID-epidemy started on date T1. We identify two Key dates T2 and T3 when the Govt of XX has introduced Restrictive measures to reduce the Infection Transmission rate (given by a coefficient Beta in the SIR/SEIR models) and to increase the Removal rate (given by a coefficient Gamma in the SIR/SEIR models)). In China these are 30-Jan-2020 and 12-Feb-2020. In Bulgaria they are 14-Mar-2020 and 20-Mar-2020. The model sets two nodes of the splines for Beta and Gamma which are chosen by optimization for fitting to the data.

Beta is reduced by the following measures: washing hands, wearing masks, keeping social distance and soc. isolation (no meetings, staying at home), visiting less stores. Gamma is increased by: fast identification of infected people and subsequent fast hospitalization/isolation until they recover, applying proper medication, etc. The USER may create different scenarios of the Future by using the following Control parameters:

5.9.1 Program input control parameters

1. **Third Date:** Fix a Third date **T4** to **strengthen / weaken** the Restrictive measures during the next days.

2. **Coef1:** Decide whether to relax or to strengthen the Transmission rate of today (Beta1) by choosing a number Coef1 in the interval [0.2,

1.4]; here 1 means the same "amount of measures" as until today; 0.2 means about 5 times less measures; in fact, the formula for the new value B2 of the rate is roughly Beta2 = (1.6-Coef1)*Beta1.

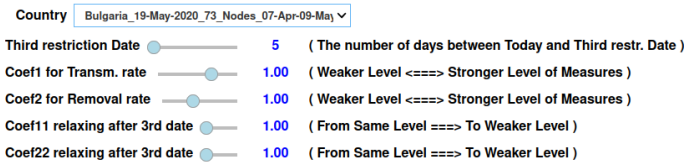
3. **Coef11:** This coefficient says how much the USER wants to relax the Level of the Restrictive measures (which determine the Transmission rate) after the Third Date **T4**. For example, if Coef11 is 1.8, this means that we impose "5 times less measures immediately after Third Date". If Coef11 is 1, this means that we preserve the same level of the measures immediately after Third Date. Roughly, we have the new Transmission rate **Beta3 = Coef11*Beta2**.

4. **Coef2:** The same about the new value **Gamma2** of the Removal rate; we put roughly **Gamma2 = Coef2*Gamma1**, where Gamma1 is the rate until today.

5. **Coef22:** This is analogous to coefficient **Coef11** but relaxes the measures which determine the Removal rate (Gamma).

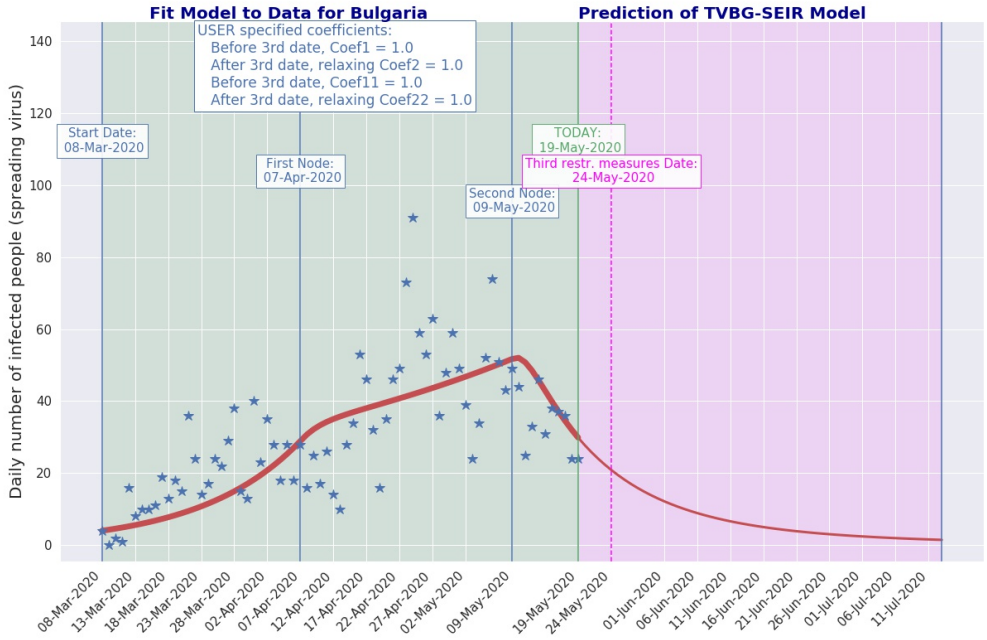
5.9.2 Program input control Widgets

Here is an EXAMPLE of input data (Interactive Widgets, Sliders, Drop list)



5.9.3 Program output result

The output result from TVBG-SEIR Tool for a given input parameters.



5.9.4 Program code

```
from ipywidgets import interact, interactive, fixed, interact_manual
from ipywidgets import Button, HBox, VBox, widgets, Label
from IPython.display import Image, display, HTML, Javascript
from pathlib import Path
import os, glob, math

display(HTML("""<style>.container { width:98% !important;}
.widget-readout, .widget-label {font-size: 16px; font-weight: bold} .widget-readout
{color: blue;}
.widget-hslider {width: 360px;} .widget-dropdown {width: 420px; font-weight: bold;}
</style>"""))

datapath = "FIGURES/"
Country = [dI for dI in os.listdir(datapath) if os.path.isdir(os.path.join(datapath,dI))
Country.sort()
if '.ipynb_checkpoints' in Country:
Country.remove('.ipynb_checkpoints')

def getlist(Country):
datapath = 'FIGURES/' + Country + '/'
```

```
listOfImageNames = [datapath + dI for dI in os.listdir(datapath)]

N = len(listOfImageNames)-1
def f(n):
if N != -1:
imgfile = Path(datapath + str(n) + '.jpg')

if imgfile.exists():
display(Image(imgfile))
else:
display(Image("datasets/DIVERGENT_MODEL.jpg"))
style = {'description_width' : 'initial', 'handle_color' : 'lightblue'}
Td = widgets.IntSlider(min=5, max=25, step=10, value=5, description='Third resti
Coef1 = widgets.FloatSlider(min=0.2, max=1.5, step=0.2, value=1.0, description=
Coef2 = widgets.FloatSlider(min=0.6, max=1.8, step=0.2, value=1.0, description=

Coef11 = widgets.FloatSlider(min=1.0, max=1.8, step=0.4, value=1.0, description
Coef22 = widgets.FloatSlider(min=1.0, max=1.8, step=0.4, value=1.0, description
L_Td = HBox([Td, Label(value="( The number of days between Today and Third resti
L_Coef1 = HBox([Coef1, Label(value="( Weaker Level <===> Stronger Level of Meas
L_Coef2 = HBox([Coef2, Label(value="( Weaker Level <===> Stronger Level of Meas
L_Coef11 = HBox([Coef11, Label(value="( From Same Level ===> To Weaker Level )"
L_Coef22 = HBox([Coef22, Label(value="( From Same Level ===> To Weaker Level )"
ui = widgets.VBox([L_Td, L_Coef1, L_Coef2, L_Coef11, L_Coef22], style=style )

def fsliders(Td, Coef1, Coef2, Coef11, Coef22):
Coef11, Coef22 = Coef11-0.80, Coef22-0.80
Td = math.ceil(Td/10)
C1, C2 = Coef1/0.2, Coef2/0.2
C1, C2 = round(C1), round(C2)
Coef11, Coef22 = Coef11/0.4, Coef22/0.4
Coef11, Coef22 = math.ceil(Coef11), math.ceil(Coef22)
Num = 49*3*3*(Td-1) + 7*3*3*(C1-1) + 3*3*(C2-3) + 3*(Coef11-1) + Coef22
f(Num)
output = widgets.interactive_output(fsliders, {'Td': Td, 'Coef1': Coef1, 'Coef2
output.layout.height = '900px'
output.layout.width = '90%'
output.layout.align_items = 'stretch'
controls = widgets.VBox([ui])
display(ui,output)

interact(getlist,Country=Country);
```

```
from IPython.display import HTML
HTML('''<script>
code_show=true;
function code_toggle() {
if (code_show){
$('div.input').hide();
} else {
$('div.input').show();
}
code_show = !code_show
}
$( document ).ready(code_toggle);
</script>
<form action="javascript:code_toggle()">
<input type="submit" value="Toggle on/off the raw code."></form>''')
```

5.10 ATVBG-SEIR Tool Web Interface

Lockdown Scenarios Tool based on ATVBG-SEIR Model for long-term predictions based on splines (see Chapter 3) ATVBG-SEIR is a tool that can be used to predict the length of the COVID19 pandemic and the number of lockdowns needed to extinguish it in Germany and in Bulgaria. The users can use this tool either with the default provided parameters or by changing them using the provided Slider buttons and arrows to fine-tune. Default parameter values incl. limits are described in the legend below. If you are interested in the results for other countries, please contact the authors of the tool which are happy to collaborate. The ATVBG-SEIR tool is provided to advance the knowledge about COVID19 pandemic and is intended to be used only for research, education and public decision making. Program input control parameters

5.10.1 Program input control parameters

1. **P1:** Escalation parameter, which is an escalation factor showing the ratio between the unreported cases and reported cases, i.e. All cases = P1 * Reported cases. (default value is 3, input by a Slider; min=1, max=10)

2. **P2:** Fixed number of days showing the lengths of the lockdown periods (of Strict measures). (default is 21; input by a Slider; min=7, max=60)

3. **P3:** This number shows the ICUs as a percentage of the Active Cases. (default for Bulgaria is 0.6

4. **P4:** This number shows the upper limit of ICU beds reserved for the Covid-19 cases. (default for Bulgaria is 1100, for Germany is 10000; input by a Slider). It is announced in the database `https://www.worldometers.info/coronavirus/#countries`, but seems to be variable for every country since the authorities always keep higher reserves for emergencies. It is difficult to download true data about the daily occupancy of the ICU beds.

5. **P5:** This number is the percentage of the population that has to be immune, for the epidemic to be assumed ended. (default value is 70

6. **P6:** This is the percentage of the upper Limit of ICUs. If the model requires number of ICUs less than this, we assume end of the epidemic. (default value is 1

7. **P7:** VACCINATIONS number is the total number of vaccinated persons per month (as usually planned); (default value is 0; Input in a box). Let us remind that every vaccinated person gets two vaccinations in a short time. Actual values (per month) for Bulgaria are about 120,000; for Germany - about 1,400,000.

8. **P8:** A table to enter the number of vaccines (NOT vaccinated people) delivered per month.

9. **P9:** The user enters parameters (Beta, Gamma) for Strict regime and Relax regime.

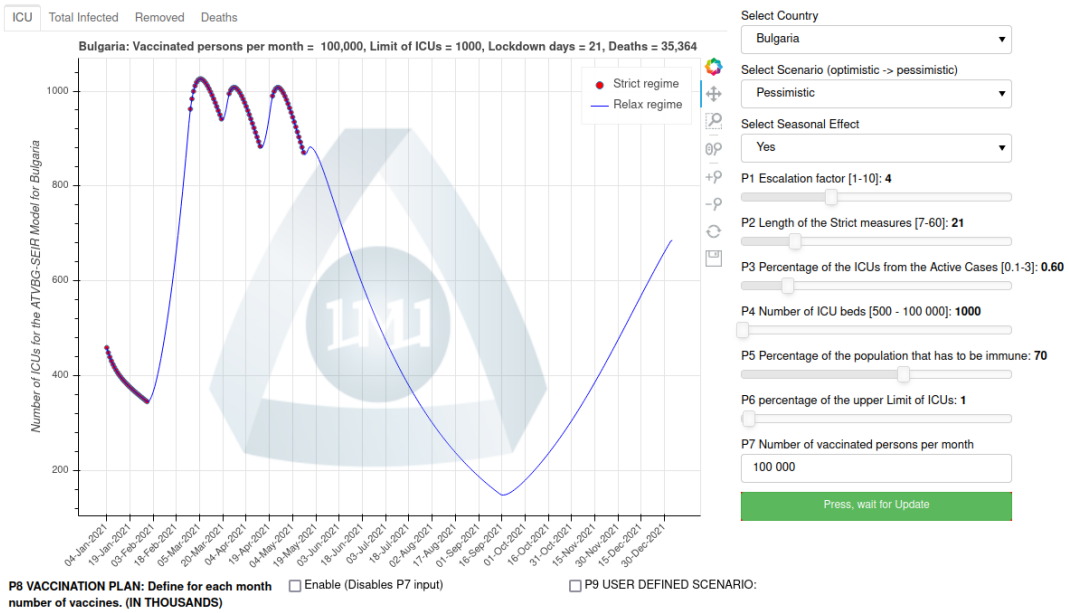
NOTE:

- If not Figure displayed, please refresh the page (from the Browser) or click on Cell (on the top bar), then Run All, and wait a bit, since the calculations of the Model take a minute or even more.

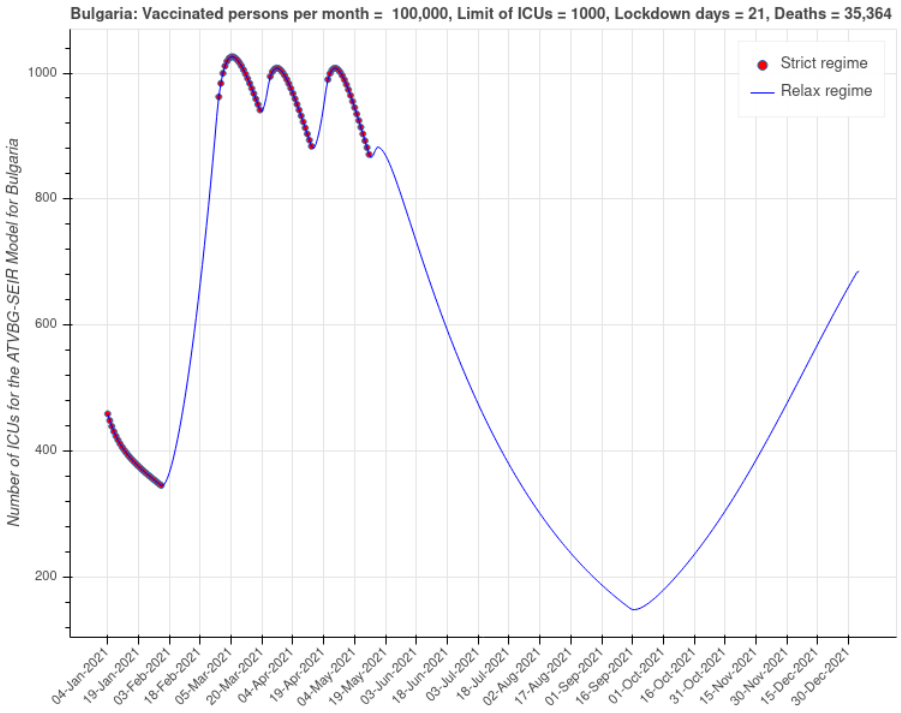
- If necessary to fine-tune the sliders of the parameters below (especially for P4), click on the Slider button, then use the arrows to fine-tune.

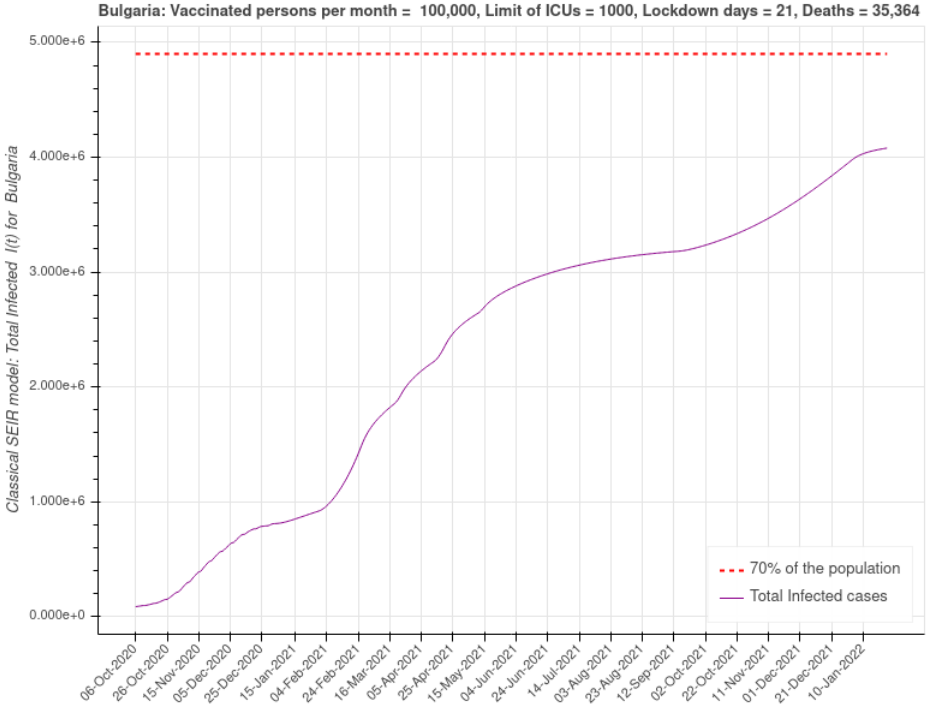
- The number of fatalities (Deaths) is calculated without taking into account that the vaccination is targeting the most vulnerable to Covid-19, hence practically will be much less.

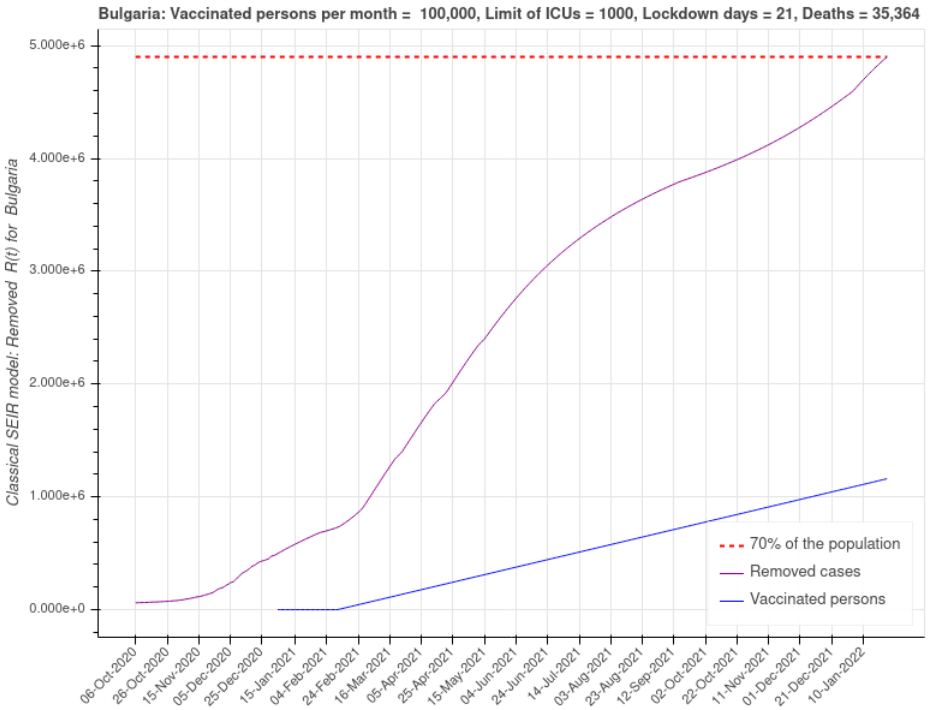
5.10.2 Program input control Widgets

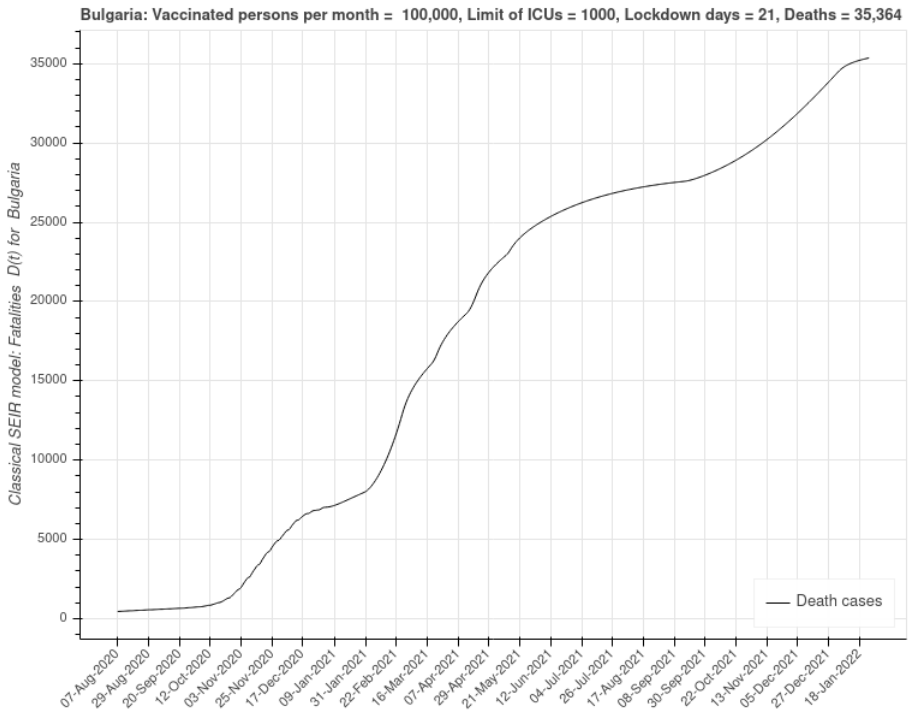


5.10.3 Program output result




Bulgaria: Vaccinated persons per month = 100,000, Limit of ICUs = 1000, Lockdown days = 21, Deaths = 35,364
Classical SEIR model: Total Infected I(t) for Bulgaria
5.000e+6
4.000e+6
3.000e+6
2.000e+6
1.000e+6
0.000e+0
70% of the population
Total Infected cases
06-Oct-2020
26-Oct-2020
15-Nov-2020
05-Dec-2020
25-Dec-2020
15-Jan-2021
04-Feb-2021
24-Feb-2021
16-Mar-2021
05-Apr-2021
25-Apr-2021
15-May-2021
04-Jun-2021
24-Jun-2021
14-Jul-2021
03-Aug-2021
23-Aug-2021
12-Sep-2021
02-Oct-2021
22-Oct-2021
11-Nov-2021
01-Dec-2021
21-Dec-2021
10-Jan-2022


Bulgaria: Vaccinated persons per month = 100,000, Limit of ICUs = 1000, Lockdown days = 21, Deaths = 35,364
Classical SEIR model: Removed R(t) for Bulgaria
5.000e+6
4.000e+6
3.000e+6
2.000e+6
1.000e+6
0.000e+0
70% of the population
Removed cases
Vaccinated persons
06-Oct-2020
26-Oct-2020
15-Nov-2020
05-Dec-2020
25-Dec-2020
15-Jan-2021
04-Feb-2021
24-Feb-2021
16-Mar-2021
05-Apr-2021
25-Apr-2021
15-May-2021
04-Jun-2021
24-Jun-2021
14-Jul-2021
03-Aug-2021
23-Aug-2021
12-Sep-2021
02-Oct-2021
22-Oct-2021
11-Nov-2021
01-Dec-2021
21-Dec-2021
10-Jan-2022



5.10.4 Program code

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
# LOCKDOWN TOOL based on ATVBG-SEIR Model
import sys, os, warnings
sys.path.insert(1, '../src')

import csv
import math
import pandas as pd
import numpy as np
import scipy as sc

from scipy.io import loadmat
from scipy.optimize import minimize
import scipy.interpolate as sci

import time
from datetime import date, datetime
from lockdowns_utils import *
```

```
from ipywidgets import interact
from bokeh.plotting import figure, show
from bokeh.plotting import output_file, output_notebook, reset_output
from bokeh.io import push_notebook

from bokeh.layouts import row, column, widgetbox
from bokeh.models.widgets import Slider, TextInput, Dropdown, Button, Select, P
from bokeh.models import ColumnDataSource, Div, CheckboxGroup
from bokeh.application import Application
from bokeh.application.handlers import FunctionHandler

from IPython.display import clear_output, Image
import bokeh
from bokeh.io import curdoc
from bokeh.io import reset_output

# from tornado.log import enable_pretty_logging
# enable_pretty_logging()

from IPython.display import HTML, Javascript
# display(Javascript('IPython.notebook.execute_cells_below()'))
display(HTML("<style>.container { width:98% !important;} </style>"))

# output_notebook()
warnings.filterwarnings('ignore')

#Setting up initial parameters
Ime = 'bg' # 'ger', 'it', 'bg'
Sigma = 1/5.2
Drate = 0.03
t_step = 1
KK = 0.4 # steepness of the splines
RK = 1 # Runga-Kutte option
C3 = 3 # multiplier

C21 = 21 # THIS IS THE LENGTH OF THE Strict LOCKDOWNS - WE TAKE IT CONSTANT
C06 = 0.6 # Set how many ICU are occupied as a Percentage of the Active cases
BEDS = 1100 # Set Limit for max number of ICU beds
VACCINE = 0
# Define global data: with Python this all has to go in the same cell - alterna

DATAcell = { 'bg' : {'Country':'Bulgaria', 'datafile':'bulgaria_coronavirus__.c
'it' : {'Country':'Italy', 'datafile':'italy_coronavirus__.csv', 'N':60*pow(10,
'ger': {'Country':'Germany','datafile':'germany_coronavirus__.csv', 'N':83*pow(
```

```
'sp': {'Country':'Spain', 'datafile':'spain_coronavirus.csv', 'N':47*pow(10,6), 'St-End_
'uk': {'Country':'UK', 'datafile':'uk_coronavirus__.csv', 'N':66*pow(10,6), 'St-End_Date
'rus': {'Country':'Russia', 'datafile':'russia_coronavirus.csv', 'N':150*pow(10,6), 'St-
'us': {'Country':'USA','datafile':'usa_coronavirus__.csv', 'N':333*pow(10,6), 'St-End_Da
'fr': {'Country':'France', 'datafile':'france_coronavirus.csv', 'N':67*pow(10,6), 'St-En
'au': {'Country':'Austria', 'datafile':'austria_coronavirus__.csv', 'N':9*pow(10,6), 'St
}

# setting up the workspace
basedir = os.getcwd()
datadir = os.path.abspath("../datasets/")
os.makedirs(datadir, exist_ok=True)

# download for bg
# Data_Download(Ime,datadir)

#########################################################
# Setup widgets
C3 = Slider(title="P1 Escalation factor [1-10]",
            value=4, start=1, end=10, step=1)
C21 = Slider(title="P2 Length of the Strict measures [7-60]",
            value=21, start=15, end=45, step=1)
C06 = Slider(title="P3 Percentage of the ICUs from the Active Cases [0.1-3]", value=0.6,
BEDS = Slider(title="P4 Number of ICU beds [500 - 100 000]", value=1000, start=500, end=
C70 = Slider(title="P5 Percentage of the population that has to be immune", value=70, st
C01 = Slider(title="P6 percentage of the upper Limit of ICUs", value=1, start=0.2, end=3
VACCINATION = TextInput(title="P7 Number of vaccinated persons per month", value="100 00

select = Select(title="Select Country", value='Bulgaria', options = ['Austria','Bulgaria
select_scenario = Select(title="Select Scenario (optimistic -> pessimistic)", value='-',
select_seasonal = Select(title="Select Seasonal Effect", value='-', options = ['-','Yes'

button = Button(label="Press, wait for Update", button_type="success")
status_div = Div(text="", style={'font-size': '200%', 'color': 'red'})
label_vac_div = Div(text="P8 VACCINATION PLAN: Define for each month number of vaccines.
checkbox_dates = CheckboxGroup(labels=['Enable (Disables P7 input)'], active=[])
checkbox_beta_gamma = CheckboxGroup(labels=['P9 USER DEFINED SCENARIO:'], active=[])

M1 = TextInput(title="2021 Jan", value="0")
M2 = TextInput(title="2021 Feb", value="0")
M3 = TextInput(title="2021 Mar", value="0")
M4 = TextInput(title="2021 Apr", value="0")
M5 = TextInput(title="2021 May", value="0")
M6 = TextInput(title="2021 June", value="0")
```

```
M7 = TextInput(title="2021 July", value="0")
M8 = TextInput(title="2021 Aug", value="0")
M9 = TextInput(title="2021 Sep", value="0")
M10 = TextInput(title="2021 Oct", value="0")
M11 = TextInput(title="2021 Nov", value="0")
M12 = TextInput(title="2021 Dec", value="0")

Q1 = row(M1, M2, M3, M4)
Q2 = row(M5, M6, M7, M8)
Q3 = row(M9, M10, M11, M12)

RelaxBeta = TextInput(title="Relax Regime Beta:", value="")
RelaxGamma = TextInput(title="Relax Regime Gamma:", value="")

StrictBeta = TextInput(title="Strict Regime Beta:", value="")
StrictGamma = TextInput(title="Strict Regime Gamma:", value="")

BeGa = row(StrictBeta, StrictGamma, RelaxBeta, RelaxGamma)
StrictBeGa = ["",""]
RelaxBeGa = ["",""]

BeGa.visible = False

VACmonthly = []

# Setup Callbacks

def updatedata():
# Get the current slider values
C3_ = C3.value
C21_ = C21.value
C06_ = C06.value
BEDS_ = BEDS.value
C70_ = C70.value
C01_ = C01.value
VACCINATION_ = VACCINATION.value
Ime = select.value
VACCINATION_ = VACCINATION_.replace(' ','')
VACCINATION_ = VACCINATION_.replace(',','')
if not VACCINATION_.isdigit():
VACCINATION_ = "0"
VACCINATION.value = "0"
if checkbox_dates.active:
VACmonthly = []
```

```
for mvac in [M1, M2, M3, M4, M5, M6, M7, M8, M9, M10, M11, M12]:
VACmonthly.append(mvac.value)
else: VACmonthly = []

if checkbox_beta_gamma.active:
StrictBeGa = [BeGa.children[0].value, BeGa.children[1].value]
RelaxBeGa = [BeGa.children[2].value, BeGa.children[3].value]
else:
StrictBeGa, RelaxBeGa = ["",""], ["",""]
fig = MAIN(DATAcell, C01_, C3_, C21_, C06_, C70_,
           BEDS_, VACCINATION_, VACmonthly, Ime,
           checkbox_dates.active, StrictBeGa, RelaxBeGa,
           select_scenario.value, select_seasonal.value)
tab1 = Panel(child=fig[0], title="ICU")
tab2 = Panel(child=fig[1], title="Total Infected")
tab3 = Panel(child=fig[2], title="Removed")
tab4 = Panel(child=fig[3], title="Deaths")
tabs = Tabs(tabs=[ tab1, tab2, tab3, tab4 ])

layout.children[0] = row(tabs, d1, inputs)
button.disabled=False
status_div.text = ""

def set_country_defaults(attrname, old, new):
Ime = select.value
select_scenario.value = '-'
select_seasonal.value = '-'
if Ime == "Bulgaria":
C3.value = 4
C21.value = 45 # 21
C06.value = 0.6
BEDS.value = 1100
C70.value = 70
C01.value = 1
VACCINATION.value = "120 000"
select_scenario.options = ['-','Optimistic','Realistic', 'Pessimistic']
select_seasonal.options = ['-','Yes','No']
M1.value = "76"
M2.value = "107"
M3.value = "124"
M4.value = "520"
M5.value = "77"
M6.value = "0"
M7.value = "595"
```

```
if Ime == "Germany":
C3.value = 3
C21.value = 21
C06.value = 1.4
BEDS.value = 10000
C70.value = 70
C01.value = 1
VACCINATION.value = "1 400 000"
select_scenario.options = ['-','Optimistic', 'Pessimistic']
select_seasonal.options = ['-']

if Ime == "UK":
C3.value = 3
C21.value = 21
C06.value = 0.16
BEDS.value = 2000
C70.value = 70
C01.value = 1
VACCINATION.value = "1 000 000"
select_scenario.options = ['-','Optimistic', 'Pessimistic']
select_seasonal.options = ['-']
if Ime == "USA":
C3.value = 3
C21.value = 21
C06.value = 0.1835
BEDS.value = 0.3*2.7*33200
C70.value = 70
C01.value = 1
VACCINATION.value = "5 600 000"
select_scenario.options = ['-','Optimistic', 'Pessimistic']
select_seasonal.options = ['-']
if Ime == "Austria":
C3.value = 3
C21.value = 21
C06.value = 1.8
BEDS.value = 1000
C70.value = 70
C01.value = 1
VACCINATION.value = "150 000"
select_scenario.options = ['-','Optimistic', 'Pessimistic']
select_seasonal.options = ['-']

if Ime == "Italy":
```

```
C3.value = 3
C21.value = 21
C06.value = 0.45
BEDS.value = 2764
C70.value = 70
C01.value = 1
VACCINATION.value = "2 000 000"
select_scenario.options = ['-','Optimistic', 'Pessimistic']
select_seasonal.options = ['-']

# Set up layouts and add to document
inputs = column(select, select_scenario, select_seasonal, C3, C21,
                            C06, BEDS, C70, C01, VACCINATION,
                            button, status_div)

d1 = Div(text = '<div style="position: absolute; left:-580px; top:130px">
<img src="imi-bas-logo.png" style="width:380px; height:380px; opacity: 0.2">
</div>')

if checkbox_dates.active:
Q1.visible = True
Q2.visible = True
Q3.visible = True
else:
Q1.visible = False
Q2.visible = False
Q3.visible = False
if checkbox_beta_gamma.active:
BeGa.visible = True
else:
BeGa.visible = False

layout = column(row(Div(text='<div style="width:800px; height:600px">
<img src="bg_default_plot.png"></div>'), d1, inputs),
row(label_vac_div, checkbox_dates, checkbox_beta_gamma), BeGa, Q1, Q2, Q3)

def click_bnt():
status_div.text = "Please wait ... Running ..."
button.disabled=True
button.background = "#FF0000"
curdoc().add_next_tick_callback(updatedata)

def next_tick():
time.sleep(200)
```

```
button.disabled=False
def show_months(attr, old, new):
if checkbox_dates.active:
Q1.visible = True
Q2.visible = True
Q3.visible = True
VACCINATION.disabled=True
else:
Q1.visible = False
Q2.visible = False
Q3.visible = False
VACCINATION.disabled=False

def show_beta_gamma(attr, old, new):
if checkbox_beta_gamma.active:
Q1.visible = False
Q2.visible = False
Q3.visible = False
BeGa.visible = True
if checkbox_dates.active:
Q1.visible = True
Q2.visible = True
Q3.visible = True
VACCINATION.disabled=True
BeGa.children[0].value = BeGa.children[0].value.replace(',','.')
BeGa.children[1].value = BeGa.children[1].value.replace(',','.')
BeGa.children[2].value = BeGa.children[2].value.replace(',','.')
BeGa.children[3].value = BeGa.children[3].value.replace(',','.')
else: BeGa.visible = False
def modify_doc(doc):
doc.add_root(row(layout, width=800, height=900))
doc.title = "ATVBG-SEIR"

select.on_change('value', set_country_defaults)
button.on_click(click_bnt)
checkbox_dates.on_change('active', show_months)
checkbox_beta_gamma.on_change('active', show_beta_gamma)

handler = FunctionHandler(modify_doc)
app = Application(handler)
reset_output()
output_notebook(hide_banner=True)
show(app, notebook_url="http://213.191.194.141:8890")
```

Chapter 6

Contributions, Conclusions, Further directions

6.1 Contributions

6.1.1 Scientific Contributions

The main scientific contributions of the present work are as follows:

1. The building of new models of SEIR type which are with time-varying parameters based on splines. The discrete models are intuitive and are easy to be understood for people inexperienced in continuous Dynamical systems.

2. The first model TVBG-SEIR is based on spline models for the SEIR parameters (of transmission and removal rates), generates scenarios for a short term forecasting, with a two month time horizon; there is a set of parameters (for relaxation and tightening of the containment measures for Covid-19—) which may be varied to obtain different scenarios.

3. The second model ATVBG-SEIR differs essentially from the first, generates scenarios for a long term forecasting and thus permits the generation of projections in the horizon of about several years. It has at disposal a set of parameters which are specific for every country: escalation, length of lockdown periods, percentage of ICUs, limit of ICUs, two types of vaccination plans, and a couple of technical parameters.

6.1.2 Applied Scientific Contributions

The main applied scientific contributions of the present work are as follows:

1. I have developed two web-based instruments (software as a web-service).

2. The first instrument is based on the TVBG-SEIR model and implements an interactive mean (visualization of the scenarios and the curves of the SEIR models) for generation of short-term predictions / projections, via web-controls for the interactive variation of the parameters. It is implemented by means of Jupyter Notebook and Bokeh software package (Version 2.3) for a general client server.

3. The second developed web-based instrument is using the ATVBG-SEIR model, and deals with generation of long-term prediction scenarios/projections. These scenarios include customizable vaccination plans, and also a proper model of the seasonal effect. There is an option allowing for the user to change a large variety of models (and model parameters) in an interactive way.

4. I have provided a short description of the software solution based on Jupyter Notebook and Bokeh software package.

5. In collaboration with prof. Kounchev we applied our methods and models mainly to data of Bulgaria, however we provide also some examples of applications of models to data for different countries, as Germany, Austria, Italy, UK, USA.

6.1.3 Publications

The main publications:

1. Kounchev O., Simeonov G., Kuncheva Zh., 2021. *Estimation of the Duration of Covid-19 Epidemic in a Single Country, with or without Vaccinations. The Case of Bulgaria and Germany*, Comptes rendus de l'Acade'mie bulgare des Sciences, Vol. 74, No. 5, pp. 677-686, DOI: 10.7546 / CRABS.2021.05.05

2. Kounchev O., Simeonov G., Kuncheva Zh., 2021. *Scenarios for the spread of COVID-19 analyzed by the TVBG-SEIR spline model*, Biomath 10 (2021), 2103087, http://dx.doi.org/10.11145/j.biomath.2021.03.087

3. Kounchev O., Simeonov G., Kuncheva Zh., *How Long 'Lockdowns' Are Needed to End the COVID-19 Epidemic in a Single Country, with or without Vaccinations*, (January 14, 2021). Available at SSRN: `https://ssrn.com/abstract=3766521orhttp://dx.doi.org/10.2139/ssrn.3766521`

4. Kounchev O., Simeonov G., Kuncheva Zh., 2020, *The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios*, arXiv:2004.11338, https://arxiv.org/pdf/2004.11338

5. Stanchev, P., Ancheva, H., Pavlov, R., Simeonov, G., *The eleventh national information day: Open science, Open Data, Open Access, Bulgarian Open Science Cloud*, DiPP2020, 2020-September, pp. 275–281

6.1.4 Lectures delivered

1. O. Kounchev, G. Boyadzhiev, G. Simeonov *Short and medium term forecasts for Omicron variant - Examples on Bulgarian data*, (2021), Spring Scientific Session of Sofia University "St. Kliment Ohridski" at Faculty of Mathematics and Informatics (FMI) – 26.03.2022, Section: Covid-19 mathematical models and forecasts, `https://www.fmi.uni-sofia.bg/bg/proletna-nauchna-sesiya-na-fmi-2022`

2. Georgi Simeonov, *ATVBG-SEIR Scenarios Tool for Estimation of Covid-19 duration with and without vaccinations*, (2021), Spring Scientific Session of Sofia University "St. Kliment Ohridski" at Faculty of Mathematics and Informatics (FMI) – 27.03.2021, Section: Covid-19 mathematical models and forecasts, `https://www.fmi.uni-sofia.bg/bg/proletna-nauchna-sesiya-na-fmi-2021`

3. O. Kounchev and G. Simeonov, *How Long 'Lockdowns' Are Needed to End the COVID-19 Epidemic in a Single Country, with or without Vaccinations*, (2021), Interdisciplinary Seminar on Biomathematics and Scientific Computing – 18.02.2021 `https://math.bas.bg/?p=10124`,`https://researchseminars.org/seminar/BMNI`

4. Georgi Simeonov, *NI4OS-Europe – National Open Science initiatives in Europe to the European Open Science Cloud (EOSC)*, Annual reporting session for 2020, department SofTIS, IMI-BAS.

6.2 Conclusions and further directions

The main **conclusions** of the present work are:

1. The compartmental models of the SIR/SEIR family are very flexible approach which is easy to adapt to analyze the data related to the pandemic of Covid-19. In particular, one may model successfully the introduction of containment measures with a strength changing in time, by means of spline coefficients in the usual system of differential or difference equations.

2. The SEIR type models are flexible and allow for a direct incorporation of information about different vaccination plans, hence they provide an adequate modeling of the whole picture, by modeling the combinatin of containment measures and vaccinations.

3. The model TVBG-SEIR represents a useful mean for short term prediction and may be used by the health authorities for short term planning of containment measures. The web-based tool built atop of the model is providing an immediate help to the policy makers for planning in the dynamic pandemic situation.

4. The model ATVBG-SEIR represents a useful mean for long term prediction and may be used for decision making in the long run of the global assessment of the casualties caused by the pandemic as well as for estimating the vaccination policies.

As **further directions**, we would point out to the building of new models for describing the dynamics of disease spread by the new variants of Covid-19 as Delta. It is also a challenge to model the "mixed dynamics" where two variants are dominant - the Alpha, Delta and Omicron variants.

6.3 Statement of Originality

I declare that the dissertation on the topic: "Modeling and Analysis of Big Data for Covid-19 Epidemic" presented in connection with the procedure for obtaining the educational and scientific degree "Doctor" at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences is my work.

Citation of all sources of information, text, illustrations, tables, images and others are marked according to the standards. The main dissertation results and contributions of the dissertation research are original and are not borrowed from other research and publications.

It is original work except where due reference is made. It has not been and shall not be submitted for the award of any degree or diploma to any other institution of higher learning.

Bibliography

D Adam. Special report: The simulations driving the world's response to covid-19. *Nature*, 4 2020.

S Altizer, A Dobson, P Hosseini, P Hudson, M Pascual, and P Rohani. Seasonality and the dynamics of infectious diseases. *Ecol. Lett*, 9:467–484, 2006.

R Anderson and R May. Infectious diseases of humans, 1991.

R Anguelov, J Banasiak, C Bright, Jms Lubuma, and R Ouifki. The big unknown: The asymptomatic spread of covid-19. *Biomath*, 9:2005103, 2020.

C Aschwanden. The false promise of herd immunity for covid-19. *Nature*, 587(7832):26–28, 2020. doi: 10.1038/d41586-020-02948-4.

A Audi, M Alibrahim, M Kaddoura, G Hijazi, Yassine Hm, and H Zaraket. Seasonality of respiratory viral infections: Will covid-19 follow suit? front. *Public Health*, 8:567184, 2020. doi: 10.3389/fpubh.2020.567184.

H Biswas, L Paiva, and Mdr De Pinho. A seir model for control of infectious diseases with constraints. *Mathematical Biosciences and Engineering*, 11 (4), 8 2014. doi: 10.3934/mbe.2014.11.761.

S Boatto, C Bonnet, B Cazelles, and F Mazenc. Sir model with time dependent infectivity parameter: approximating the epidemic attractor and the importance of the initial phase. *HAL*, 2018.

F Brauer. Mathematical epidemiology, 2008.

F Brauer, C Castillo-Chavez, and C Castillo-Chavez. *Mathematical models in population biology and Epidemiology*, volume 2. Springer, New York, 2012. ISBN 978-1-4614-1686-9. doi: 10.1007/978-1-4614-1686-9.

B Buonomo, N Chitnis, and A Onofrio. Seasonality in epidemic models: a literature review. *Ricerche mat*, 67:7–25, 2018. doi: 10.1007/s11587-017-0348-6.

M Chadha and V Potdar. Dynamics of influenza seasonality at sub-regional levels in india and implications for vaccination timing. *PLoS ONE*, 10(5): e0124122, 2015. doi: 10.1371/journal.pone.0124122.

Andrew Czyzewski. Modelling an unprecedented pandemic. the vital role of team-based, collaborative epidemiology and disease modelling in managing pandemics, Jul 2020. URL `https://www.imperial.ac.uk/stories/coronavirus-modelling/`.

C Daley, M Fydenkevez, and S Ackerman-Morris. A systematic review of the incubation period of sars-cov-2: The effects of age, biological sex, and location on incubation period. *MedRxiv*, 2020. doi: 10.1101/2020.12.23.20248790.

J David, Pejman Earn, Rohani, M Benjamin, Bryan Bolker, and Grenfell. A simple model for complex dynamical transitions in epidemics source. *New Series*, 287(5453):667–670, 1 2000.

et al. Dehning, J. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. *Science*, 369, 7 2020. doi: 10.1126/science.abb9789.

K Dietz. The incidence of infectious diseases under the influence of seasonal fluctuations. *Mathematical Models in Medicine*, pages 1–15, 1976.

Editorial. Theorists and experimentalists must join forces. *Nature Computational Science*, 1(5):299–299, 2021. ISSN 2662-8457. doi: 10.1038/s43588-021-00082-3. URL `https://doi.org/10.1038/s43588-021-00082-3`.

S Feng, Z Feng, C Ling, C Chang, and Z Feng. Prediction of the covid-19 epidemic trends based on seir and ai models. *PLoS ONE*, 16(1):e0245101, 1 2021. doi: 10.1371/journal.pone.0245101.

M. Gabriela M. Gomes and Aguas et al. Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold. *medRxiv*, 2020. doi: 10.1101/2020.04.27.20081893.

N Grassly and C Fraser. Seasonal infectious disease epidemiology. *Proc. R. Soc. Lond. B Biol. Sci*, 273:2541–2550, 2006.

Trisha Greenhalgh, Jose L. Jimenez, Kimberly A. Prather, Zeynep Tufekci, David Fisman, and Robert Schooley. Ten scientific reasons in support of airborne transmission of sars-cov-2. *The Lancet*, 397(10285):1603–1605, 2021. ISSN 0140-6736. doi: 10.1016/S0140-6736(21)00869-2. URL `https://doi.org/10.1016/S0140-6736(21)00869-2`.

T Hastie, R Tibshirani, and J Friedman. Springer, 2009.

Z He, L Ren, J Yang, L Guo, L Feng, C Ma, and et al. Wang, C. Seroprevalence and humoral immune durability of anti-sars-cov-2 antibodies in wuhan, china: a longitudinal, population-level, cross-sectional study. the lancet. *The Lancet*, 397(10279):1075–1084, 3 2021. ISSN 0140-6736. doi: 10.1016/S0140-6736(21)00238-5. URL `https://doi.org/10.1016/S0140-6736(21)00238-5`.

J Heffernan, R Smith, and L Wahl. Perspectives on the basic reproductive ratio. *J. R. Soc. Interface*, 2(7):281–293, 6 2005. doi: 10.1098/rsif.2005.0042.

H Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4): 599–653, 12 2000. doi: 10.1137/S0036144500371907.

S Hirve, L Newman, J Paget, E Azziz-Baumgartner, J Fitzner, N Bhat, K Vandemaele, and W Zhang. Influenza seasonality in the tropics and subtropics - when to vaccinate? *PLoS One*, 11(4):e0153003, 4 2016. doi: 10.1371/journal.pone.0153003.

Papst Irena, Earn David, and J. Invariant predictions of epidemic patterns from radically different forms of seasonal forcing. *J. R. Soc. Interface*, page 162019020220190202, 2019.

A Katok and B Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, 1997. ISBN 0521575575.

Matt J Keeling and Pejman Rohani. *Modeling Infectious Diseases in humans and animals*. Princeton University Press, 2008. ISBN 9780691116174. doi: 10.2307/j.ctvcm4gk0.

W Kermack and A Mckendrick. Containing papers of a mathematical and physical character. *Proceedings of the Royal Society of London. Series A*, 115(772):700–721, 8 1927. doi: 10.1098/rspa.1927.0118.

P Koul, S Broor, S Saha, J Barnes, C Smith, M Shaw, M Chadha, and R Lal. Differences in influenza seasonality by latitude northern india. *Emerg. Infect. Dis*, 20(10):1723–1726, 10 2014. doi: 10.3201/eid2010.140431.

O Kounchev, G Simeonov, and Zh Kuncheva. How long 'lockdowns' are needed to end the covid-19 epidemic in a single country, with or without vaccinations. *SSRN*, 1 2021a. doi: 10.2139/ssrn.3766521.

O Kounchev, G Simeonov, and Kuncheva Zh. The tvbg-seir spline model for analysis of covid-19 spread, and a tool for prediction scenarios. *Biomath*, 10(1), 2021b. doi: 10.11145/j.biomath.2021.03.087;previousver-sionsath ttps://arxiv.org/abs/2004.11338.

Ognyan Kounchev, Georgi Simeonov, and Zhana Kuncheva. Estimation of the duration of covid-19 epidemic in a single country, with or without vaccinations. the case of bulgaria and germany. *Comptes Rendus de L'Academie Bulgare des Sciences*, 74(5):677–686, 2021c. doi: 10.7546/CRABS.2021.05.05.

Ellen Kuhl. *Computational Epidemiology - Data-Driven Modeling of COVID-19*. Springer International Publishing, 2021. ISBN 978-3-030-82890-5. doi: 10.1007/978-3-030-82890-5. URL `https://www.springer.com/gp/book/9783030828899`.

J Last. A dictionary of epidemiology, 2001.

S Lauer, K Grantz, Qifang Bi, F Jones, Qulu Zheng, H Meredith, A Azman, N Reich, and J Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9):577–582, 5 2020. doi: 10.7326/M20-0504.

M Lipsitch, T Cohen, B Cooper, J Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Kai Suok, Chorh Chuan Chew, Matthew Tan, David Samore, Megan Fisman, and Murray. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300:1966–1970, 6 2003. doi: 10.1126/science.1086616.

J Liu, Sh, and Xia. *Computational Epidemiology: From Disease Transmission Modeling to Vaccination Decision Making*. Springer, 2020. ISBN 978-3-030-52109-7. doi: 10.1007/978-3-030-52109-7. URL `https://www.springer.com/gp/book/9783030521073`.

C Mcaloon, Á Collins, and K Hunt. Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10:e039652, 2020. doi: 10.1136/bmjopen-2020-039652.

Nature. Covid research: a year of scientific milestones, 25 march - coronavirus antibodies last for months - if you have them, May 2021. URL `https://www.nature.com/articles/d41586-020-00502-w`.

R Neilan and S Lenhart. An introduction to optimal control with an application in disease modeling, 2010.

US Dept. of Health and Human Services. Flu activity & surveillance, Jul 2020. URL `http://www.cdc.gov/flu/weekly/fluactivitysurv.htm`.

L Peeples. What the science says about lifting mask mandates. *Nature*, 5 2021.

L Piroth. Comparison of the characteristics, morbidity, and mortality of covid-19 and seasonal influenza: a nationwide, population-based retrospective cohort study. *The Lancet*, 2020. doi: 10.1016/S2213-2600(20)30527-0.

M Pollicott, Hao Wang, and Howard Weiss. Extracting the time-dependent transmission rate from infection data via solution of an inverse ode problem, 2012.

S Pyne, A Vullikanti, and M Marathe. Big data applications in health sciences and epidemiology. handbook of statistics, 2015.

A Radulescu, C Williams, and K Cavanagh. Management strategies in a seir-type model of covid 19 community spread. *Sci Rep*, 10:21256, 2020. doi: 10.1038/s41598-020-77628-4.

C Siettos and L Russo. Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4):295–306, 2013. doi: 10.4161/viru.24041.

B Srebrov, O Kounchev, and G Simeonov. *Book Chapter 19 - Big Data for the Magnetic Field Variations in Solar-Terrestrial Physics and Their Wavelet Analysis*, pages 347–370. 1 2020. doi: 10.1016/B978-0-12-819154-5.00031-X.

P Stanchev, H Ancheva, R Pavlov, and G Simeonov. The eleventh national information day: Open science, open data, open access, bulgarian open science cloud, digital presentation and preservation of cultural and scientific heritage, 9 2020.

J Stoer and R Bulirsch. Springer-Verlag, Berlin, New York, 2002.

S Straif-Bourgeois, R Ratard, and M Kretzschmar. Infectious disease epidemiology. *Handbook of Epidemiology*, pages 2041–2119, 2014. doi: 10.1007/978-0-387-09834-034.

Biao Tang, Nicola Bragazzi, Qian Li, Sanyi Tang, Yanni Xiao, and Jianhong Wu. *An updated estimation of the risk of transmission of the novel coronavirus*, volume 5, pages 248–255. 2019. doi: 10.1016/j.idm.2020.02.001.

Biao Tang, Xia Wang, Qian Li, Nicola Bragazzi, Sanyi Tang, Yanni Xiao, and Jianhong Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *J. Clinical Medicine*, 9:462, 2020. doi: 10.3390/jcm9020462.

P Van Den Driessche and J Watmough. Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1-2):29–48, 2002.

J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings: Biological Sciences*, 274(1609):599–560, 2 2007.

Chapter 7

APPENDICES

7.1 APPENDIX. Use of Mathematical Models in Epidemiological Studies

Below we provide some quotes from Handbook of Epidemiology, [Straif-Bourgeois et al., 2014, see Bibliography to chapter 1], which represents a justified praise of the mathematical modeling in Epidemiology:

Although mathematical modeling has been around for a long time, until recently, it was not much used as a tool for public health, but was considered a specialized research area for applied mathematicians and theoretical biologists. This started to change with the advent of the HIV pandemic, when mathematical models were first used to predict future epidemic spread, and to analyze the impact of behavior change on HIV incidence (Kaplan and Brandeau 1994). However, the breakthrough for mathematical modeling as a public health tool came with the concerns that smallpox virus could be used in a deliberate release and lead to devastating outbreaks in the only partially immune populations of present societies. How can public health policy be developed against threats with pathogens that are not circulating at present? There is no way to conduct epidemiological investigations, and the only available data in the case of smallpox were from before the eradication era. Therefore, to design policy, knowledge from historical smallpox outbreaks had to be combined with data about present-day society, and possible interventions had to be tested on the basis of this available information. Mathematical modeling provided a flexible tool to do that and was used to analyze possible vaccination strategies and other interventions.

"Later, the experience with the global spread of SARS caused by a novel strain of corona virus – and the threat of a future pandemic with a new strain of influenza A initiated national and international efforts to better prepare for large outbreaks of emerging infections. Mathematical modeling was widely used for investigating optimal strategies for dealing with a new influenza pandemic. These response plans came into action during the pan-

demic with new influenza A H1N1 emanating from Mexico in the spring of 2009. Even as the pandemic was still unfolding, first mathematical modeling studies started to deliver valuable data analyses almost in real time.

Besides supporting public health policy in designing prevention and intervention strategies, mathematical modeling of infectious diseases has contributed greatly to increasing the understanding of the intricate relationships between clinical and biological determinants of infection and human contact and risk behavior patterns that lead to transmission. The importance of core groups of high sexual activity in the transmission dynamics of sexually transmitted infections (Hethcote and Yorke 1984), the impact of concurrent partnerships on the spread of HIV (Morris and Kretzschmar 1997), the importance of hosts being infectious before the appearance of symptoms for disease control (Fraser et al. 2004), and the connectedness of modern societies in a small world network (Watts and Strogatz 1998) are just some examples for how mathematical modeling has shaped the present paradigms of infectious disease epidemiology.

The following story about the team of mathematical modelers in Epidemiology at Imperial college has played an essential role in forming the proper health policy during the outbreak of Covid-19 in UK, see the link `https://www.imperial.ac.uk/stories/coronavirus-modelling/` for the full story:

Media reports have suggested that an update to the *Imperial college team's model in early March was a critical factor* in jolting the UK government into changing its policy on the pandemic. The researchers initially estimated that 15% of hospital cases would need to be treated in an intensive-care unit (ICU), but then updated that to 30%, a figure used in the first public release of their work on *16 March*. That model showed the UK health service, with just over $4,000$ ICU beds, would be overwhelmed.

Government officials had previously talked up a theory of allowing the disease to spread while protecting the oldest in society, because large numbers of infected people would recover and provide herd immunity for the rest. But they changed their course on seeing the new figures, ordering social-distancing measures. Critics then asked why social distancing hadn't been discussed earlier, why widespread testing hadn't happened, and why **modellers** had even chosen the 15% figure, given that a January paper showed that more than 30% of a small group of people with COVID-19 in China needed treatment in ICUs.

"As for the Chinese data on ICUs, clinicians had looked at them, but noted that *only half* the cases seemed to need *invasive mechanical ventilators*; the others were given *pressurized oxygen*, so might not need an *ICU bed*. On the basis of this and their experience with viral pneumonia, clinicians had advised modellers that 15% was a better assumption."

"The key update came the week before **Ferguson** briefed government officials at Downing Street. Clinicians who had been talking to horrified

colleagues in Italy said that *pressurized oxygen wasn't working well* and that all 30% of the severe hospitalized cases would need invasive ventilation in an ICU. Ferguson says the updated models' mortality projections didn't change hugely, because many predicted deaths are likely to occur in the community rather than in hospitals. But the understanding of how health services would be overwhelmed, and the experience of Italy, *led to a sudden focusing of minds*, he says: government officials swiftly pivoted to *social-distancing measures* (see 'Lockdowns keep infections at bay')."