

Scenarios for the spread of COVID-19 analyzed by the TVBG-SEIR spline model

Ognyan Kounchev¹, Georgi Simeonov¹, Zhana Kuncheva²

¹Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria

kounchev@math.bas.bg, gsimeonov@math.bas.bg

²C4X Discovery Ltd
London, UK
z.kuncheva@gmail.com

Received: 16 September 2020, accepted: 8 March 2021, published: 15 March 2021

Abstract— We develop a novel TVBG-SEIR spline model for analysis of the coronavirus infection (COVID-19). It aims to analyze the long-term global evolution of the epidemics “controlled” by the introduction of lockdown/open up measures by the authorities. The incorporation of different “lockdown scenarios” varying in time permits to analyze not only the primary epidemic wave but also the arising secondary wave and any further waves.

The model is supplied by a web-based Scenario Building Tool for COVID-19 (called shortly SBT-COVID19) which may be used as a decision support software by (health) policy makers to explore various scenarios. This can be achieved 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-COVID19 Tool permits to assess how long the lockdown measures should be maintained.

Keywords-COVID-19; SEIR model; spline models; time-varying transmission and removed rates; time-varying basic reproduction number

I. INTRODUCTION

A. Context

The coronavirus infection (COVID-19) started in December 2019 in Wuhan, China, and has quickly spread out to almost all other countries in the world. The possible catastrophic impact on the economy and health system has required the active intervention of the authorities to introduce various containment measures. This caused the necessity for creating an appropriate model to plan the type, scale and duration of the measures.

Mathematical models are traditionally used to analyze the long-term global evolution of epidemics, to determine the potential and severity

Copyright: ©2021 Kounchev et al. This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Ognyan Kounchev, Georgi Simeonov, Zhana Kuncheva, 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>

of an outbreak, and to provide critical information for identifying the type of disease interventions and intensity. One of the widely used mathematical models of long-term spreading of epidemics are the so-called deterministic compartmental models (SIR/SEIR type models [4]) which reflect adequately the clinical progression of the virus spreading, epidemiological status of the individuals, and the intervention measures. One of the main purposes of applying such models is to assess how the expensive containment measures imposed by the authorities (home and social isolation/quarantine, travel restrictions, etc.) can effectively reduce the basic reproduction number, transmission rate and risk of the disease. In particular, it is essential to assess how the expensive, 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.

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 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.

B. 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-COVID19 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. 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 *Today*. 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 *Node1*, *Node2*, and *Today* date. It is assumed that $\beta(t)$ and $\gamma(t)$ are constant in the time interval [*StartDate*, *Node1*], and of $\beta(t)$ is monotone decreasing while $\gamma(t)$ is monotone increasing function. The TVBG-SEIR model with time-varying $\beta(t)$ and $\gamma(t)$ will be fitted simultaneously to two sets of data: the daily infected cases (or their cumulative vector), and the removed cases (which are all removed cases until a certain date).

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*) for not very long periods of time. These models are used as a basis for creation of “prediction scenarios” starting from *Today*, with a prediction perspective of about 2 months horizon. These models are updated every day (by the arrival of the official daily data), and the scenarios are renewed accordingly. In a more mathematical language “scenario” 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 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 in the classical sense of the word.

The web-based SBT-COVID19 Tool was designed for visualization of the results of the fitted model (the daily infected cases), and for creating prediction scenarios for the daily infected cases during the next two month horizon, by controlling the future values of Beta and Gamma. It is described in detail in Section VII, and is available at the links

Version 3 (Matlab generated Figures)

http://213.191.194.141:8888/notebooks/TVBG-SEIR-Spline-model_v3.ipynb?token=b5d97bfb7dd062e47ee7ab51837e470a8c226743a4667ee

Version 4 (Python generated Figures)

http://213.191.194.141:8889/notebooks/TVBG-SEIR-Spline-model_v4.ipynb?token=b5d97bfb7dd062e47ee7ab51837e470a8c226743a4667ee

C. Objectives

In the case of the usual seasonal flu the main parameters of the spread of the viruses are the transmission rate β which reflects the power of the transmission of the virus from infected people to susceptibles, the recovery rate γ which is reciprocal to the recovery period (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. Due to the long incubation period and 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.

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 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 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, [12], [13]. 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.

D. Data Acknowledgement

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

- HDX Humanitarian Data Exchange V1.39.3
[time_series_covid19_confirmed_global.csv](https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases)
[time_series_covid19_deaths_global.csv](https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases)
[time_series_covid19_recovered_global.csv](https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases)
<https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>
- Kaggle SRK Sudalairaj Kumar
<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#>
[covid_19_data.csv](https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#)

The paper is organized as follows: In Section I we recall the deterministic SEIR model and introduce some notions and notations. In Section II

we introduce the discretization of the SEIR model which is used in the algorithms. In Section III we introduce and provide all technical details of the TVBG-SEIR spline model. In Section IV we provide an application of the TVBG-SEIR model to Bulgarian data, which are used to illustrate the work of the SBT-COVID19 Tool for prediction scenarios. In Section V and Section VI we provide more examples of analysis, for the Italian data, and for the German data. In Section VII we describe the technical details of the SBT-COVID19 Tool for prediction scenarios. In Section VIII we provide some recent references about models with time-varying transmission rates and their fitting to the data (calibration).

II. THE CLASSICAL DETERMINISTIC SEIR MODEL: NOTIONS AND NOTATIONS

The main purpose of the authors' presentation is to provide all notations and formulas as fully as possible, to enable the replicability of the calculations and experiments in the paper. Thus we also provide a detailed account of the discretization of the well-known classical SEIR model.

A. Compartmental models

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 [2]. One may use a deterministic approach using a system of ODEs or a stochastic approach which is more complicated. The deterministic approach is what we follow, namely our choice is the 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 below.

For a detailed and excellent introduction to the compartmental models we refer to the monograph

[4]. We provide a short description of the deterministic SEIR model which will be the main approach in our research.

The classical SEIR model is based on the consideration of four compartments, C_S , C_E , C_I and C_R which are described as follows: Compartment C_S : its size is $Sr(t)$ – 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.

Compartment C_E : its size is $Er(t)$ – the number of "exposed" people at time t - these are the people who 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 that the average incubation period is 11.5 days [3]. 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. Compartment C_I : its size is $Ir(t)$ – 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 they go to compartment C_R , but they have stayed in C_I only until they have been diagnosed (and these are the official data which we obtain – denoted further by $I_{data}(t)$). Compartment C_R : its size is $Rr(t)$ – the number of recovered or deceased (or immune) individuals, which are all called "removed". Normally they come from compartment C_I after becoming healthy and no more virus spreaders. Officially these data are provided in a cumulative way.

B. Official data available for modeling purposes

We do not have the "reality data" $Sr(t)$, $Er(t)$, $Ir(t)$, $Rr(t)$. The available official data $Idata(t)$ are 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 for seasonal flu (and it is considered to be similar for COVID-19) the size of C_I is much bigger than that indicated by the official data $Idata(t)$, and we have the inequality $Idata(t) \leq Ir(t)$.

The officially announced data $Rdata(t)$ contains the cumulative number of recovered cases and the $Ddata(t)$ is the cumulative number of fatalities.

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 inequalities 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.

C. Definition of the 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 above time series of the "reality data" $Sr(t), Er(t), Ir(t), Rr(t)$. 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:

$$\begin{aligned} S'(t) &= -\beta(t)S(t)I(t)/N \\ E'(t) &= \beta(t)S(t)I(t)/N - \sigma E(t) \\ I'(t) &= \sigma E(t) - \gamma(t)I(t) \\ R'(t) &= \gamma(t)I(t) \end{aligned} \tag{1}$$

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

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-Rohani, (2008), p.18). Here and further $\beta(t)$ is called Transmission rate of the infection, which we call further simply Beta.

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. The coefficient σ is the latent rate, or the rate of "becoming symptomatic" (where $1/\sigma$ is the average of the incubation period). In the present paper we use the constant value $\sigma = 1/5.2$ which represents a reasonable approximation, as the recent research shows, see [3].

The variable $S(t)$ corresponds to the reality data $Sr(t)$. The quantity $\sigma E(t)$ is equal to the daily new infectious cases $Idata(t)$. However, the quantity $I(t)$ 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. It is assumed that the following equation holds

$$N = S(0) + E(0) + I(0) + R(0), \tag{2}$$

where N is the total population in the country XX. Obviously, since the sum

$$S'(t) + E'(t) + I'(t) + R'(t) = 0$$

it follows that equation (2) holds for every time $t \geq 0$.

III. 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 (1):

$$\begin{aligned} S(n+1) &= S_n - (\beta_n S_n I_n)/N \\ E(n+1) &= E_n + (\beta_n S_n I_n)/N - \sigma E_n \\ I(n+1) &= I_n + \sigma E_n - \gamma_n I_n \\ R(n+1) &= R_n + \gamma_n I_n \end{aligned} \quad (3)$$

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 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

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

which makes one of the equations in (3) redundant. It is well known that the above Euler method for approximating the solution of (1) is less accurate than the Runge-Kutta which is widely used, see e.g. [11]. 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 daily data of newly infected $Idata(n)$ correspond to the amount σE_n which is clear from the third equation in (3). Below we use this correspondence to define the quadratic function $F(\theta)$ for fitting our models.

Remark: One has to note that the continuous model (1) and the above discrete approximation (3) 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.

IV. THE TVBG-SEIR MODEL

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 flues, 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 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.

A. 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 need additional boundary conditions. Alternatively, one may use C^2 (twice differentiable) 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 we see examples of the dynamics of $\beta(t)$ and $\gamma(t)$ rates:

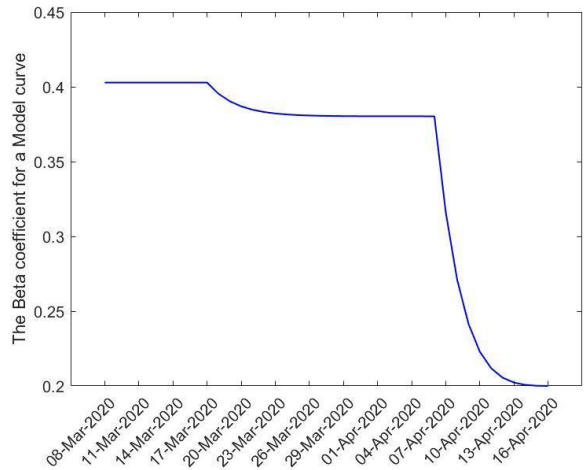


Figure 1. The rates $\beta(t)$ for a Model curve

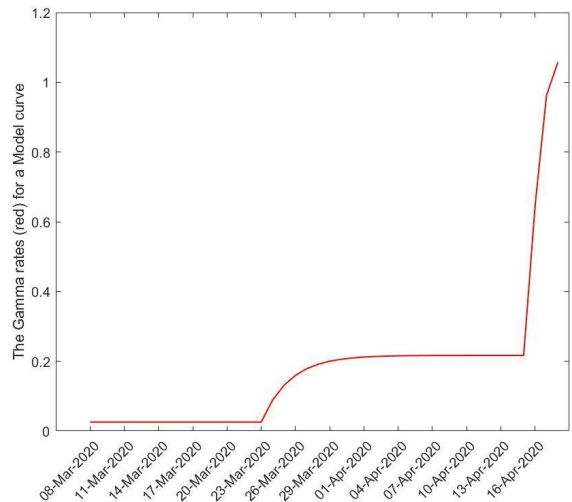


Figure 2. The rates $\gamma(t)$ for a Model curve

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 Reproduction number (ratio) is a key variable for all models of epidemics, see [4], [9], [14], [15]. Following [15] (formula (2.4)), for the case of the SEIR models with constant rates $\beta(t)$ and $\gamma(t)$, the Reproduction number is given by the formula

$$R_0 = \beta/\gamma,$$

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:

$$R_0 = \beta(t)/\gamma(t)$$

On Figure 3 below we provide the Reproduction Number R_0 , obtained by the last formula, for some specific TVBG-SEIR model:

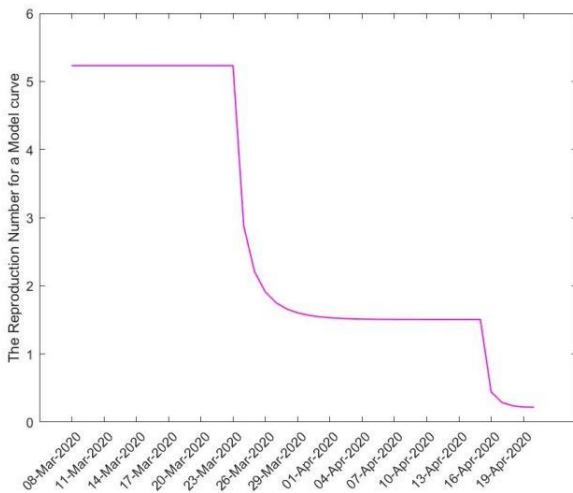


Figure 3.

8) On Figure 1 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 on Figure 2 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

$\exp(0.4 * (t - t_1))$ 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. $F(\theta)$ by varying the parameter set θ :

$$F(\theta) = \sum_{j=1}^n (cum(Idata)(t_j) - \sigma * cum(E)(t_j))^2 + (Rdata(t_j) + Deaths(t) - R(t_j))^2.$$

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 V.

14) The “curves” S_n, E_n, I_n, R_n of the discrete SEIR model are obtained by solving the system (2)

with initial conditions given by

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

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) On Figure 4 below we provide the daily data of infected cases $Idata(t)$ for Bulgaria:

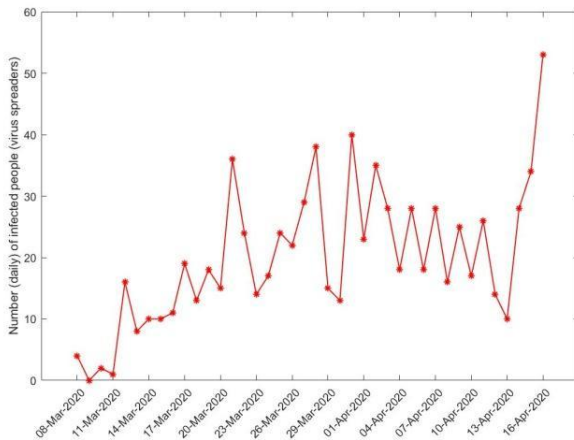


Figure 4.

16) On Figure 5 below the cumulative data for recovered and fatalities are provided. On Figure 6 we provide the fitting of the model curve $\sigma E(t)$ to the above data for Bulgaria, $Idata(t)$.

17) Figure 7 shows the fitting by the model curve $R(t)$ of the Recovered plus Fatalities data for Bulgaria.

18) It is important to remark that we have applied a “parsimonious” approach for constructing the spline model, by which one has to avoid 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.

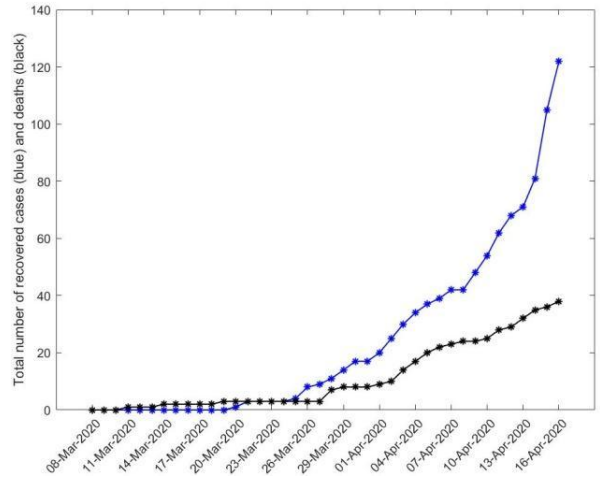


Figure 5.

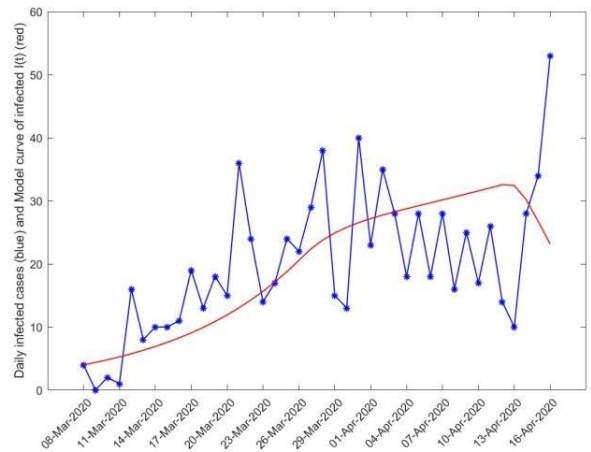


Figure 6.

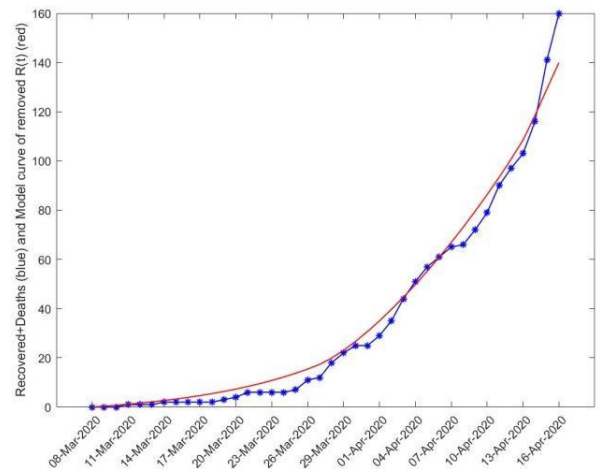


Figure 7.

19) As we said, the result of the minimization of the quadratic functional $F(\theta)$ is a set of parameters θ for which the minimum is attained, whereby there may be multiple solutions. Once we have found some model based on the parameters θ , we proceed to constructing “prediction scenarios”. We choose some date T_6 , which we call Horizon, say at most 2 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

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

The coefficients $Coef1$, $Coef11$, $Coef2$, $Coef22$ are used further to control and represent our scenario building in the SBT-COVID19 Tool. Their meaning and choice is explained in detail in Section VIII, where we introduce the Tool.

V. COVID-19 SPREAD IN BULGARIA IN OCTOBER-DECEMBER, 2020: APPLICATION TO BULGARIAN COVID-19 DATA AND SCENARIOS GENERATED BY THE SBT-COVID19 TOOL

In the SBT-COVID19 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, (Kounchev et al., (2020), [17]). There we have shown the possibility for a next wave of the infection.

Here we demonstrate how to generate prediction scenarios 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 25th, 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-COVID19 Tool will be described in detail in Section VII. We provide the visualizations of the model fitting which are available in the SBT-COVID19 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. 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 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 IV 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 θ (which define *Model1*) for which the minimum $F(\theta)$ 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 $\underline{\theta} = \theta^1, \theta^2, \theta^3$, 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. Hastie and Tibshirani (2009), [16]. Thus, it will provide us also alternative to finding the Confidence intervals for the obtained results.

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.

3) 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 8 below 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 $28 - Jan - 2021$.

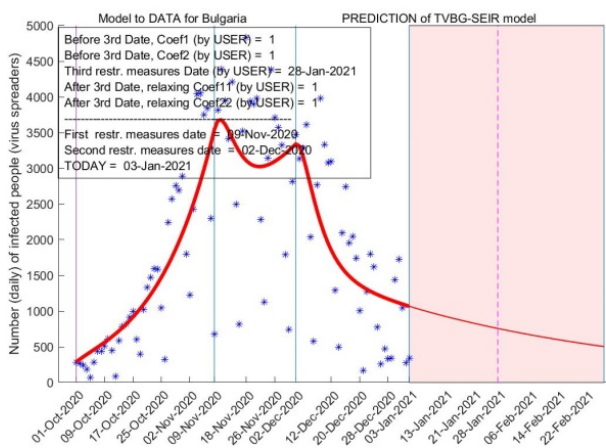


Figure 8.

4) 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 the Figure 9 below:

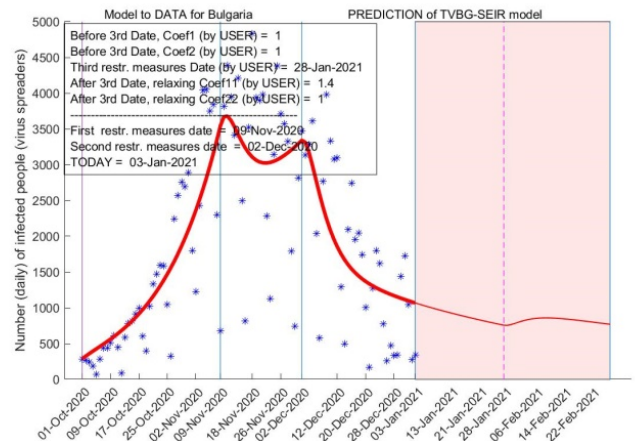


Figure 9.

From Figure 10 below we see that too much relaxing one of the measures, e.g. the first one, with $Coef11 = 1.8$, a “second wave” of infectious cases will arise in February 2021:

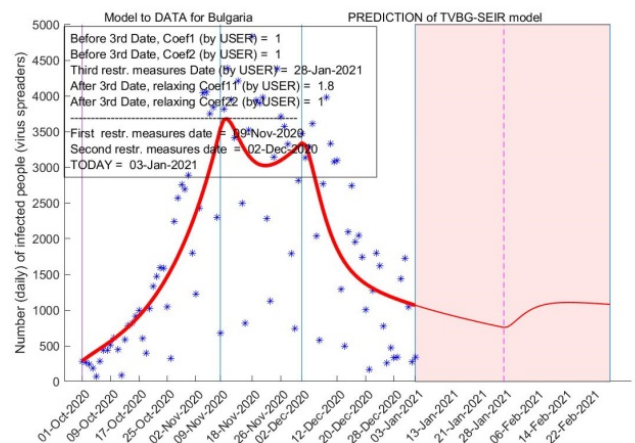


Figure 10.

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 11.

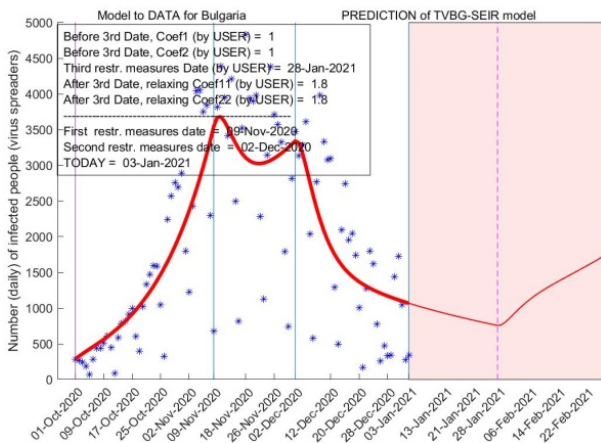


Figure 11.

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

6) 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 12 below:

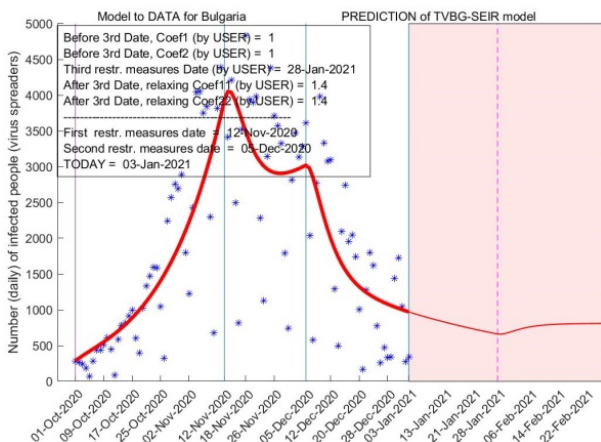


Figure 12.

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.

VI. APPLICATION TO ITALIAN COVID-19 DATA IN OCTOBER-DECEMBER, 2020, AND SCENARIOS GENERATED BY THE SBT-COVID19 TOOL

In the present section we provide similar results obtained by our scenario SBT-COVID19 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 $Fval = F(\theta) = 153.09$.

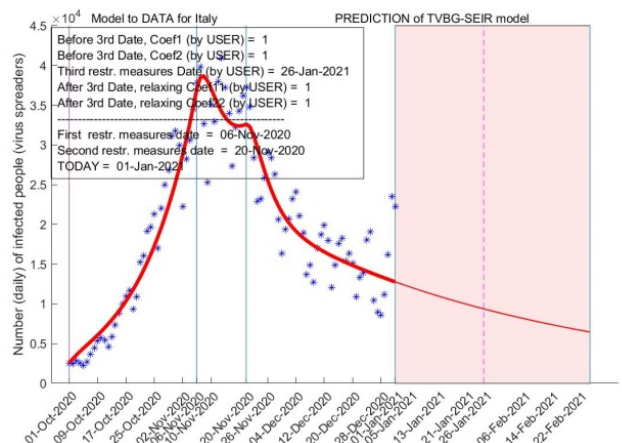


Figure 13.

A strong relaxation after the Horizon date 26-Jan-2021 results in a strong “next wave” seen in Figure 14

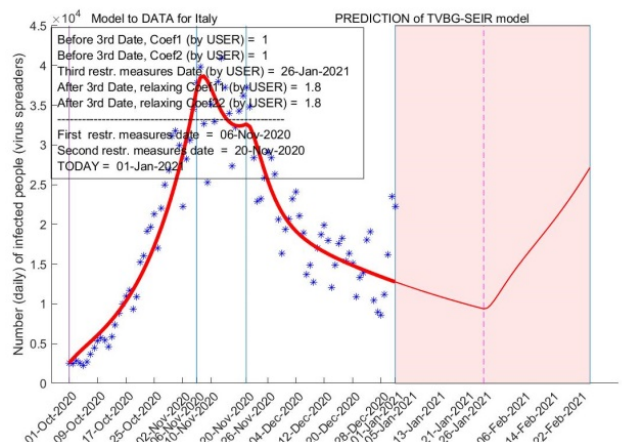


Figure 14.

2) *Model2* has $T_2 = 3 - Nov - 2020$, $T_3 = 17 - Nov - 2020$, and $Fval = F(\theta) = 159.77$. For it we obtain the following scenario:

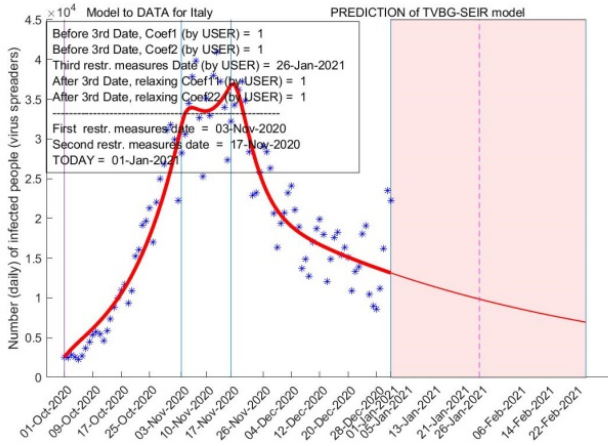


Figure 15.

Just as in *Model1*, further strong relaxation gives a strong “next wave” provided in Figure 16.

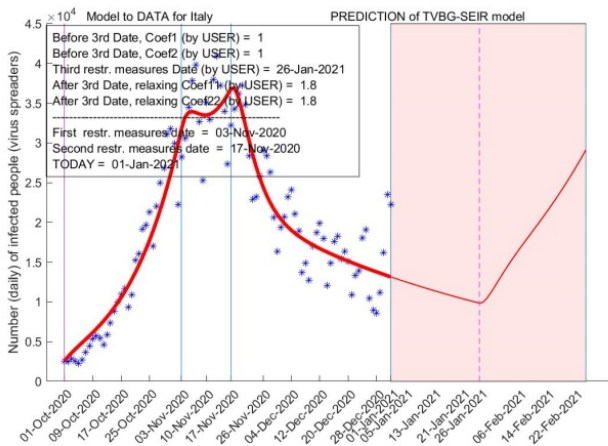


Figure 16.

Model3 has $T_2 = 6 - Nov - 2020$, $T_3 = 23 - Nov - 2020$, and $Fval = F(\theta) = 169.72$. It gives the following Figure 17.

Further relaxation after 26 - Jan - 2021 shows a bigger “next wave” in the Figure 18.

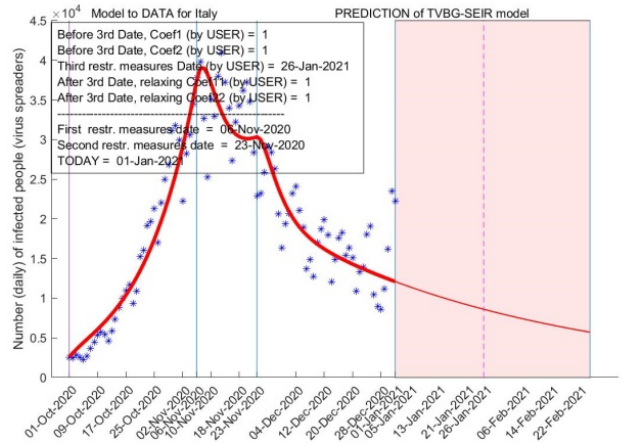


Figure 17.

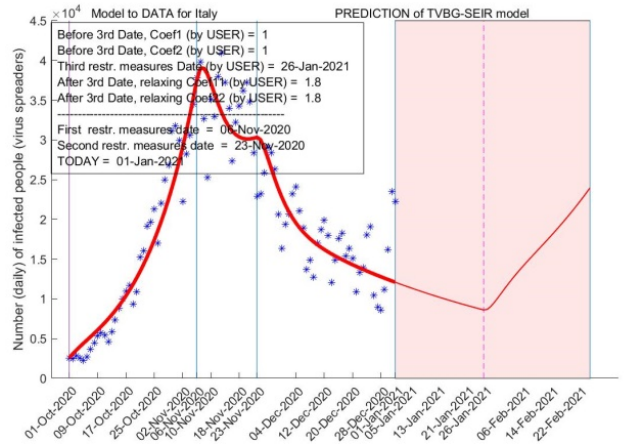


Figure 18.

VII. APPLICATION OF THE MODEL TO GERMAN COVID-19 DATA IN MARCH-MAY, 2020, AND SCENARIOS GENERATED BY THE SBT-COVID19 TOOL

We provide the results for the German data. The overall feeling is, that unlike the data for Bulgaria and Italy, they show 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 $Fval = F(\theta) = 160.96$.

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 19 below:

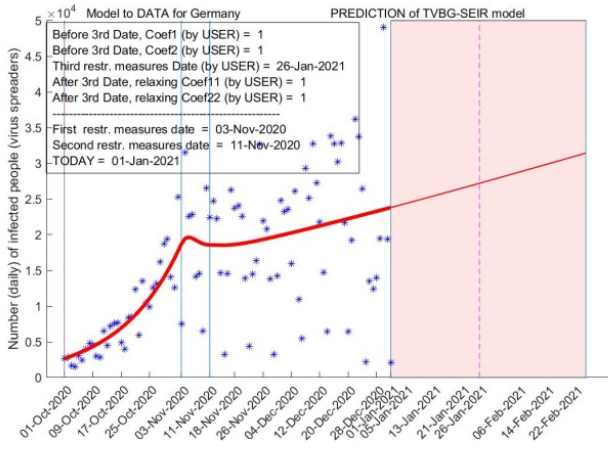


Figure 19.

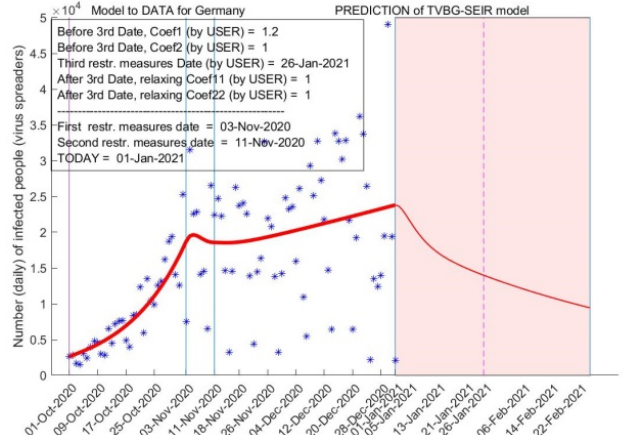


Figure 21.

2) However, a moderate tightening of the second measures ($Coef2 = 1.2$) will result in a calming (Figure 20).

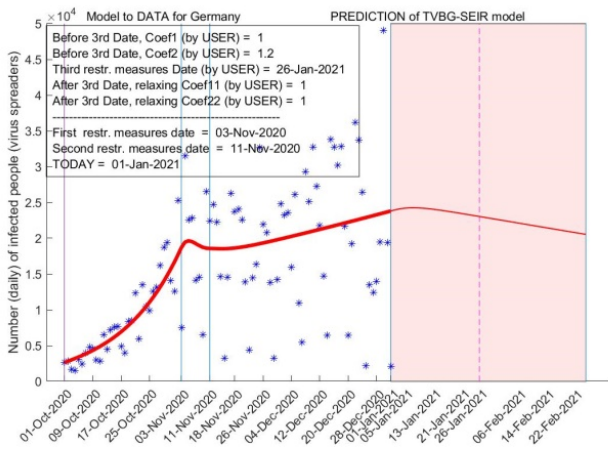


Figure 20.

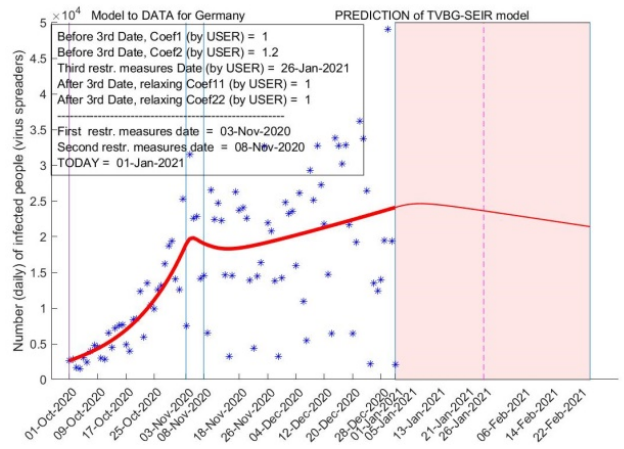


Figure 22.

But a much better result will bring the tightening of the measures influencing the coefficient β , i.e. $Coef1 = 1.2$ as seen in the Figure 21.

3) For *Model2* we have $T_2 = 3 - Nov - 2020$, $T_3 = 8 - Nov - 2020$, and $Fval = F(\theta) = 169.16$, and for *Model3* - $T_2 = 3 - Nov - 2020$, $T_3 = 20 - Nov - 2020$, and $Fval = F(\theta) = 171.61$. The application of the two models gives a result similar to that of *Model1*.

On the Figures 22 and 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.

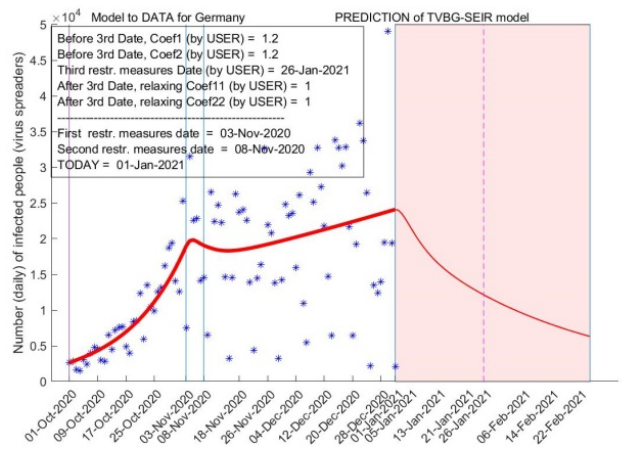


Figure 23.

And for *Model3*: we have a slight difference on Figure 24 below:

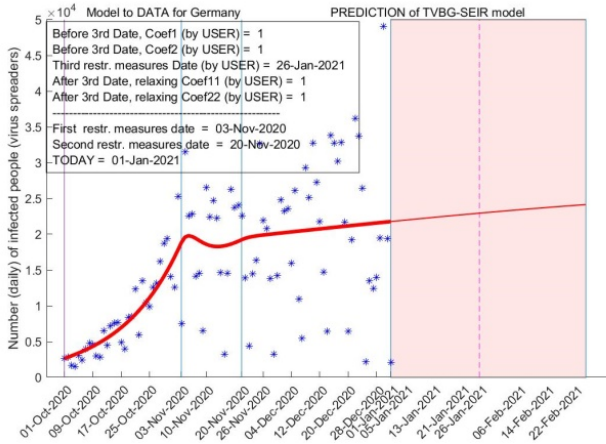


Figure 24.

VIII. DESCRIPTION OF THE SBT-COVID19 TOOL FOR CONTROLLED SCENARIOS

1) We have designed a SBT-COVID19 Tool for the Model predictions of the Corona virus (and similar infectious diseases) spread. The SBT-COVID19 Tool is based on the fitting of the TVBG-SEIR model to the official data available on a daily basis as described in Section II, B). The online tool is available at the site

<http://213.191.194.141:8888/notebooks/>

TVBG-SEIR-Spline-model_v3.ipynb?token=

b5d97bfbfd7dd062e47ee7ab51837e470a8c226743a4667ee

2) 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 .

3) 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 $T_5 = T_4 + 5, T_4 + 15, T_4 + 25$.

4) 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 .

5) 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*.

6) 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.

7) 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/> (Initially, 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.

IX. RECENT RESEARCH ON TIME-VARYING TRANSMISSION RATES

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

Let us mention some research about solving an inverse problem for finding time-varying $\beta(t)$, in a SIR model, for a fixed removal rate γ , [10], from the number of infectious cases. In [1], the authors do research and provide further references of research on specific models for the transmission rate $\beta(t)$.

X. ACKNOWLEDGEMENTS

All authors acknowledge the support by grants KP-06-N32-8 and KP-06N42-2 of Bulgarian NSF. The first-named author acknowledges the partial support by Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program (2014-2020) and co-financed by the European Union through the European structural and Investment funds.

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

The authors would like to thank the anonymous referee for the positive criticism which has essentially improved the paper.

REFERENCES

- [1] 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, Id: hal-01677886, <https://hal.inria.fr/hal-01677886>
- [2] Kermack, W. O. McKendrick, A. G., "A Contribution to the Mathematical Theory of Epidemics", Proceedings of the Royal Society A, 115 (772): 700–721, 1927.
- [3] S. A. Lauer, K. H. Grantz, Qifang Bi, F. K. Jones, Qulu Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, (2020), "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application", Annals of Internal Medicine, Vol. 172, No. 9, May 2020, pp. 577-582, doi:10.7326/M20-0504.
- [4] Keeling, M. J. and Rohani, P., (2008), "Modelling Infectious Diseases", Princeton University Press.
- [5] E. Brooks-Pollock, G. O. Roberts and Matt J. Keeling, "A dynamic model of bovine tuberculosis spread and control in Great Britain", Nature 511 228-231, 2014, doi:10.1038/nature13529.
- [6] Michael J. Tildesley, Thomas A. House, Mark C. Bruhn, Ross J. Curry, Maggie O'Neil, Justine L. E. Allpress, Gary Smith, and Matt J. Keeling, (2010), "Impact of spatial clustering on disease transmission and optimal control", Proc Natl Acad Sci USA 107 1041-6, doi:10.1073/pnas.0909047107
- [7] M Woolhouse, M Chase-Topping, D Haydon, J Friar, L Matthews, G Hughes, D Shaw, J Wilesmith, A Donaldson, S Cornell, M Keeling and B Grenfell, (2001), "Foot-and-mouth disease under control in the UK", Nature 414, 258-258.
- [8] A. J Kucharski, Timothy W. Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, Rosalind M. Eggo, (2020), "Early dynamics of transmission and control of COVID-19: a mathematical modelling study", The Lancet, published online Published: 11 March 2020, doi:10.1016/S1473-3099(20)30144-4
- [9] Marc Lipsitch, Ted Cohen, Ben Cooper, James M. Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H. Samore, David Fisman, and Megan Murray, (2003), "Transmission Dynamics and Control of Severe Acute Respiratory Syndrome", Science, 20 June 2003, Vol. 300, pp. 1966-1970, doi:10.1126/science.1086616
- [10] Mark Pollicott, Hao Wang, Howard Howie Weiss, (2012), "Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem", 2012, doi:10.1080/17513758.2011.645510
- [11] Stoer, J.; Bulirsch, R. (2002), "Introduction to Numerical Analysis", 3rd ed., Berlin, New York: Springer-Verlag.
- [12] Biao Tang, Xia Wang, Qian Li, Nicola Luigi Bragazzi, Sanyi Tang, Yanni Xiao, Jianhong Wu, (2020), "Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions", J. Clinical Medicine, 2020, 9, 462; doi:10.3390/jcm9020462
- [13] Biao Tang, Nicola Luigi Bragazzi, Qian Li, Sanyi Tang, Yanni Xiao, Jianhong Wu, (2020), "An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV)", Infectious Disease Modelling, Vol. 5, 2020, pp. 248-255, ISSN 2468-0427, doi:10.1016/j.idm.2020.02.001
- [14] Wallinga, J. and M. Lipsitch, (2007), "How generation intervals shape the relationship between growth rates and reproductive numbers", Proceedings: Biological Sciences, Vol. 274, No. 1609, (Feb. 22, 2007), pp. 599-604
- [15] Jane M Heffernan, R. J. Smith and L.M. Wahl, (2005), "Perspectives on the basic reproductive ratio", J. R. Soc. Interface (2005) 2, Published online (7 June 2005), pp. 281–293, doi:10.1098/rsif.2005.0042
- [16] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, 2009.

- [17] O. Kounchev, G. Simeonov, Zh. Kuncheva, The TVBG-SEIR spline model for analysis of COVID-19 spread, and a Tool for prediction scenarios, <https://arxiv.org/abs/2004.11338>