2020/22

DP

Yurii Nesterov

Online prediction of COVID19 dynamics. Belgian case study

CORE
CENTER FOR OPERATIONS RESEARCH AND ECONOMETRICS

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/discussion-papers.html

CORE DISCUSSION PAPER

2020/22

Online prediction of COVID19 dynamics. Belgian case study

Yurii Nesterov *

June 10, 2020

Abstract

In this paper, we present a new axiomatic model of epidemic development, called HIT, which is consistent with the very special features of COVID19. This is a discrete-time linear switching model for predicting the dynamics of total number of infected persons and concentration of the asymptomatic virus holders in the population. A small number of its parameters can be tuned using the available real-time dynamic data on virus propagation. This model provides us with a rare possibility of *online prediction* of the future. As an example, we describe an application of this model to the online analysis of COVID19 epidemic in Belgium for eighty days in the period March - May, 2020. During this time, our predictions were exact, typically, within the accuracy of 0.5%. To the best of our knowledge, this is the first mathematical model for predicting the evolution of epidemics under containment measures, which prevent development of immunity in the population. From our analysis, we can easily detect a potential start of the second wave of infection. We get this warning for Belgium in the beginning of the second week of June 2020.

*Center for Operations Research and Econometrics (CORE), Catholic University of Louvain (UCL). E-mail: Yurii.Neterov@uclouvain.be.

Research results presented in this paper are obtained in the framework of Advanced Grant 788368 of the European Research Council.

1 Introduction

Motivation. During the last months, the pandemic of COVID19 remains the main factor influencing behavior of people over the world. Unprecedent security and quarantine measures changed the style of life of billions of people. Uncertain future raises the global interest to different predictions of propagation of this disease. In this paper we develop a new epidemic model, which fits well the special features of COVID19 pandemic. Before presenting it, let us discuss the classical approach.

It is clear that a reliable forecast should be based on some scientific model, explaining the internal mechanics of the pandemic. In Epidemiology, mathematical modelling has already quite a long history. The first models of that type were developed by Kermack and McKendrick in 1927 [5], most probably as a reaction on the extremely strong Spanish Flu pandemic of 1918-1919. The main variant of their model is called *SIR*, which describes the dynamic relations between susceptible, infected, and removed (by recovering or death) parts of the population. These parts, $S(t)$, $I(t)$, and $R(t)$, are differentiable functions of time, satisfying the following relations:

$$\begin{array}{lrcl} \text{a)} & S'(t) & = & -\frac{\beta}{N} I(t)S(t), \\ \text{b)} & I'(t) & = & \frac{\beta}{N} I(t)S(t) - \gamma I(t), \\ \text{c)} & R'(t) & = & \gamma I(t), \end{array} \tag{1.1}$$

where $N = S(t) + I(t) + R(t)$ is the total size of population (thus, $N'(t) \overset{(1.1)}{=} 0$). The parameter $\frac{1}{\beta}$ has interpretation of the typical time between the contacts, and $\frac{1}{\gamma}$ can be seen as the typical time until recovery. Its important characteristics is the value $R_0 = \frac{\beta}{\gamma}$ called the *basic reproduction ratio*. This model is successfully applied for analyzing the dynamics of the usual epidemics. It has many different variants (which can be found, for example, in [2, 4, 8]). However, being very useful for common diseases, all of them share the same drawbacks making these models hardly applicable for predicting the behavior of COVID19 pandemic (see, for example, models from the recent survey [1]). Let us mention the main features of this process

1. Small scale. The right-hand side of equations (1.1) depends on the product $I(t)S(t)$, which measures the number of contacts between infected and susceptible population. The product means that any infected person can meet any susceptible person. However, in the case of COVID19 this is clearly wrong. The main feature of this pandemic is that it touches a very small fraction of population. In many countries with seemingly finished epidemics, the total number of infected persons is counted by fractions of a percent. It is much less than the standard epidemic thresholds of the order 1% for population infected in one epidemic week. It is clear now that the main danger of COVID19 consists not in a high level of propagation, but in a very high *mortality rate* for certain age groups.

Let us rewrite SIR in terms of the *relative* characteristics

$$\hat{S}(t) \;=\; \tfrac{1}{N} S(t), \quad \hat{I}(t) \;=\; \tfrac{1}{N} I(t), \quad \hat{R}(t) \;=\; \tfrac{1}{N} R(t), \tag{1.2}$$

Then the equations (1.1) lead to the following dependencies:

$$\begin{aligned} \text{a)} \quad \hat{S}'(t) &= -\beta \hat{I}(t)\hat{S}(t), \\ \text{b)} \quad \hat{I}'(t) &= \beta \hat{I}(t)\hat{S}(t) - \gamma \hat{I}(t), \\ \text{c)} \quad \hat{R}'(t) &= \gamma \hat{I}(t), \end{aligned} \tag{1.3}$$

ensuring a nontrivial dynamics of these measurements. However, for COVID19, with very high accuracy we have

$$\hat{S}(t) \approx 1, \qquad \hat{I}(t) \approx 0. \tag{1.4}$$

In this case, the equation $(1.3)_b$ becomes $\hat{I}'(t) \approx (\beta - \gamma)\hat{I}(t)$. This looks as an over-simplification since the change in the number of infected persons depend only on this number.

2. Long latent period. It is well known that the world-wide distribution of COVID19 is related to a very long asymptomatic period, when the infected person can transmit the virus to others. The length of this period is estimated from five to fourteen days. At the same time, in the model (1.1) we can see an instantaneous interaction between the susceptive and infected persons $(1.1)_a$, between the infected and recovered persons $(1.1)_c$, etc. There is no time delay in this model.

3. No feedback by improved immunity. Again. in view of the small scale of the epidemic, the increasing size of recovered part of population does not seriously influences the propagation of the disease. For the standard flu-like epidemics, the development of immunity is one of the main reasons of recovering. However, at the current stage of development of COVID19 in the world, in most of the countries the main hopes are related to the containment regimes.

4. Discrete data. Any model describing a real-life situation is based on the available data. Hence, the model designer must take into account the type, periodicity, and accuracy of the data. During the pandemic of COVID19, we encountered a completely new situation, when the important information on the propagation of the virus over the world was reported daily, starting from the very beginning of the process. In this very comfortable situation, it is possible to construct a new type of models, which should be much more precise than the classical ones. In this paper, we are presenting one of them.

Contents. In this paper, we present a new axiomatic model of epidemic development, called *HIT*, which is consistent with above mentioned features of COVID19. This is a discrete-time linear switching model for predicting the dynamics of total number of infected persons and concentration of asymptomatic virus holders in the population. A small number of its parameters can be tuned using the available real-time dynamic data on virus propagation. Therefore, this model provides us also with *online prediction* of the future. As an application example, we present the conclusions of this model for epidemic of COVID19 in Belgium during in the period March-May, 2020. During eighty days, our predictions were exact, typically, within accuracy of 0.5%. This gives us a possibility to predict the start of the second wave of infection. We get this warning for Belgium in the beginning of the second week of June 2020 (see Table 3 in Appendix).

The paper is organized as follows. In Section 2, we describe the elements of HIT model and introduce our main assumptions on constant length of the latent period, described

by an integer parameter $\Delta \geq 1$. This gives us a possibility to prove the main theorem (Next-Day Low), ensuring predictability of the epidemic development. In Section 3, we establish the relations between the number of infected persons and the daily infection rate in form of a linear difference equation. Assuming that the infection rate is constant for some period of time, we characterize three different modes of epidemic development (propagation, stagnation, and recession). They can be distinguished by a trivial inequality between the infection rate and parameter Δ, which are easily observable during the real-life epidemic development. In Section 4, we study the mathematical properties of our solutions, represented in the form of Markov processes and consensus developments. The results of Section 5 are useful for understanding the asymptotic behavior of real-life epidemic processes.

Finally, in Section 6, we apply our model for describing the development of a real-life COVID19 epidemic in Belgium. We show how it is possible to estimate the parameters of the model and describe our day-to-day experience in predicting the behavior of main characteristics in eighty days of March - May 2020. We conclude the paper by Section 7 with a discussion of our results and perspectives of further development. All necessary statistics from [9] and the results of our model are presented in Appendix in Tables 1–3. Last lines of Table 3 are filled using the statistics of first ten days of June 2020.

Notation. Our notation is quite standard. We denote by $\mathbb{R}^n$ the n-dimensional space of real column vectors with origin $0_n \in \mathbb{R}^n$, and by $\mathbb{R}^n_+$ the cone of non-negative vectors in $\mathbb{R}^n$. The kth coordinate vector in $\mathbb{R}^n$ is denoted by e_k, and $\bar{e}_n \in \mathbb{R}^n$ denotes the vector of all ones. Notation I_n is used for the unit $n \times n$-matrix. The size of vector $x \in \mathbb{R}^n$ can be measured by different norms:

$$\|x\|_{(1)} \;=\; \sum_{i=1}^{n} |x^{(i)}|, \quad \|x\|_{(\infty)} \;=\; \max_{1 \leq i \leq n} |x^{(i)}|, \quad x \in \mathbb{R}^n.$$

For vector $x \in \mathbb{R}^n$, denote by $|x|$ the vector in $\mathbb{R}^n$ with coordinates $|x^{(i)}|$, $i = 1, \ldots, n$, and we write $x \geq y \in \mathbb{R}^n$ if $x^{(i)} \geq y^{(i)}$ for all $i = 1, \ldots, n$. The scalar product of two vectors $x, y \in \mathbb{R}^n$ is defined in the standard way:

$$\langle x, y \rangle \;=\; x^T y \;=\; \sum_{i=1}^{n} x^{(i)} y^{(i)}.$$

Finally, $\mathbb{C}$ denotes the complex plane, and $\mathbb{Z}$ is the set of all integer numbers.

2 HIT model

Our model is based on three dynamic characteristics of epidemic development. Current situation at day $d \in \mathbb{Z}$ is described by the following objects.

- $H(d)$, the number of asymptomatic virus holders in the beginning of day d. These holders infect other people with certain rate. The rate and duration of the latent period are also parameters of our model. They will be introduced later.
- $I(d)$, the number of newly infected persons during the day d. We call it Daily Infection Volume.

- $T(d)$, the total number of confirmed cases of infection up to the end of day d.

Following the classical examples, we will refer to our model as *HIT*. Among these characteristics, only function $T(\cdot)$ is observable, at least in a retrospective way. The two other functions will be computed as an output of our model. In our discussions, it is convenient to use also the following secondary characteristic:

- $C(d)$, the number of new cases discovered at day d. During the pandemic of COVID19, this information was reported every day by the official sources.

Thus, we want to predict the behavior of all these discrete-time dynamic characteristics of propagation of the disease. The modelling of other important characteristics like the recovery and mortality rates, and others needs additional input data. Hence, they are not included in our model.

In what follows, it is convenient to extend these characteristics also to all $d \in \mathbb{Z}$, having in mind that for the deep past all of them are equal to zero. The starting day of our model $d = 0$ corresponds to the first positive measurement of $T(\cdot)$.

The above characteristics satisfy two simple balance equations.

- **Conservation Low**

$$T(d) \;=\; T(d-1) + C(d), \quad d \in \mathbb{Z}. \tag{2.1}$$

- **Propagation Low**

$$H(d+1) \;=\; H(d) + I(d) - C(d), \quad d \in \mathbb{Z}. \tag{2.2}$$

  This means that all newly detected asymptomatic virus holders are isolated. The newly infected virus holders become active only at the next day.

Validity of the relations (2.1) and (2.2) follow from the physical meaning of our objects and the standard way of functioning of the health system.

In our model, it is convenient to introduce one more measurement.

Definition 1 *Function $\gamma(d) \geq 0$, defined by the equation*

$$I(d) \;=\; \gamma(d) H(d), \quad d \in \mathbb{Z}, \tag{2.3}$$

is called the Daily Infection Rate *of our process. If $H(d) = 0$, then $I(d)$ is also null, and we define $\gamma(d) = 0$.*

Parameters $\gamma(d)$ can be exogenous, or they can be computed by the model. These values are related to the communication mode in the society at day d, which influences the propagation rate of the virus. They correspond to the average number of persons infected during one day by one asymptomatic virus holder. Hence, they could be treated as expectations of random variables. However, in this paper we prefer to use a deterministic terminology.

Finally, for finishing our model, we need to introduce one key assumption.

Axiom 1 *Propagation of the virus has a* Constant Latent Period. *This means that*

$$C(d) \;=\; I(d - \Delta), \quad d \in \mathbb{Z}, \tag{2.4}$$

where $\Delta \geq 1$ is an integer parameter. It is called Contamination Delay.

Again, at this place we could introduce expectations. In many model this helps (e.g. [2]). However, as we will see in Section 6, our deterministic model is sufficient for a very precise description of the real-life epidemic processes. And it is clearly simpler than its probabilistic variants.

Since $C(d) = 0$ for $d \leq -1$, this axiom immediately leads to the following consequences:

$$\begin{array}{rcl} I(d) & = & 0, \quad d \leq -\Delta - 1, \\ H(d) & \stackrel{(2.2)}{=} & 0, \quad d \leq -\Delta. \end{array} \tag{2.5}$$

Under the Axiom 1, the knowledge of full historical data $T(\cdot)$ becomes sufficient for an *exact reconstruction* of all characteristics of our process, including the unobservable functions $H(\cdot)$ and $I(\cdot)$. This follows from two important relations.

Theorem 1 *For all $d \in \mathbb{Z}$, we have*

$$\begin{array}{rcl} H(d+1) & = & \sum\limits_{k=d-\Delta+1}^{d} I(k) \\ & = & T(d+\Delta) - T(d). \quad (\text{Next-Day Low}) \end{array} \tag{2.6}$$

Proof:
Indeed, for any k, we have

$$H(k+1) \stackrel{(2.2)}{=} H(k) + I(k) - C(k) \stackrel{(2.4)}{=} H(k) + I(k) - I(k-\Delta).$$

Let us fix some $m \geq 0$. Summing up these equalities for $k = d-m, \ldots, d$, we get

$$H(d+1) = H(d-m) - \sum_{i=0}^{\Delta-1} I(d-m+i) + \sum_{i=0}^{\Delta-1} I(d-i).$$

Taking now m big enough, say, $m = d + 2\Delta$ and applying equalities (2.5), we get the first equality in the representation (2.6). Further, we have

$$\begin{array}{rcl} T(d+\Delta) - T(d) & \stackrel{(2.1)}{=} & \sum\limits_{i=1}^{\Delta} C(d+i) \; = \; \sum\limits_{k=d+1}^{d+\Delta} C(k) \\ & \stackrel{(2.2)}{=} & \sum\limits_{k=d+1}^{d+\Delta} \Bigl(- H(k+1) + H(k) + I(k) \Bigr) \\ & = & -H(d+\Delta+1) + H(d+1) + \sum\limits_{k=d+1}^{d+\Delta} I(k) \\ & \stackrel{(2.6)_1}{=} & H(d+1). \end{array}$$

□

As we can see, the relations (2.6) do not depend at all on the daily infection rate $\gamma(\cdot)$.

Thus, knowing the history of epidemic $T(\cdot)$, we can reconstruct retrospectively all characteristics of the process. Indeed, assume that we know $T(d)$ for $d = 0, \dots D$. Then, by Conservation Low we have

$$C(d) \stackrel{(2.1)}{=} T(d) - T(d-1), \quad d = 0, \dots, D. \tag{2.7}$$

Daily infection volume $I(\cdot)$ can be found by Axiom 1:

$$I(d) \stackrel{(2.4)}{=} C(d+\Delta), \quad d = 0, \dots, D-\Delta. \tag{2.8}$$

Concentration of asymptomatic virus holders $H(\cdot)$ can be obtained by the Next-Day Low:

$$H(d) \stackrel{(2.6)}{=} T(d+\Delta-1) - T(d-1), \quad d = 0, \dots, D-\Delta+1. \tag{2.9}$$

Finally, daily infection rate $\gamma(\cdot)$ can be computed by its definition:

$$\gamma(d) \stackrel{(2.3)}{=} \frac{I(d)}{H(d)} = \frac{T(d+\Delta)-T(d+\Delta-1)}{T(d+\Delta-1)-T(d-1)}, \quad d = 0, \dots, D-\Delta. \tag{2.10}$$

Hence, all characteristics are reconstructible at least for $d = 0, \dots, D-\Delta$. For the comprehensive retrospective analysis, the only unknown parameter remains the length of the latent period $\Delta \geq 1$. However, this is a small integer value, and very often it is easy to estimate it by comparing several reasonable variants.

At the same time, one of the main goals of this paper is an *online analysis* of the developing epidemic processes. In the next sections, we will show how this can be done using the elements of our model.

3 Dynamics of epidemic development

Let us start from describing the dynamics of characteristics $T(\cdot)$ and $H(\cdot)$ in terms of the daily infection rate $\gamma(\cdot)$.

Theorem 2 *For any $d \in \mathbb{Z}$, we have*

$$T(d+1) = T(d) + \gamma(d-\Delta+1) \cdot \Big(T(d) - T(d-\Delta)\Big), \tag{3.1}$$

$$H(d+1) = (1+\gamma(d))H(d) - \gamma(d-\Delta)H(d-\Delta). \tag{3.2}$$

Proof:
Indeed,

$$\begin{aligned} T(d+1) &\stackrel{(2.1)}{=} T(d) + C(d+1) \stackrel{(2.4)}{=} T(d) + I(d-\Delta+1) \\ &\stackrel{(2.3)}{=} T(d) + \gamma(d-\Delta+1) \cdot H(d-\Delta+1) \\ &\stackrel{(2.6)}{=} T(d) + \gamma(d-\Delta+1) \cdot \Big(T(d) - T(d-\Delta)\Big), \end{aligned}$$

and this is the relation (3.1). For proving relation (3.2), note that

$$H(d+1) \stackrel{(2.2)}{=} H(d) + I(d) - C(d) \stackrel{(2.4)}{=} H(d) + I(d) - I(d-\Delta)$$

$$\stackrel{(2.3)}{=} H(d) + \gamma(d)H(d) - \gamma(d-\Delta)H(d-\Delta). \qquad \square$$

Note that, during some epidemics, the characteristic $T(\cdot)$ is observable in real time by the official daily statistics. On contrary, the characteristic $H(\cdot)$ is always hidden. However, it can be reconstructed either by the Next-Day Law (2.6), or by the dynamic relations (3.2) based on the hidden daily infection rate. At the same time, this rate can be learned by analyzing the behavior of $T(\cdot)$ (see (3.1)). Unfortunately, this can be done with the delay $\Delta \geq 1$, which introduces additional complications in the online analysis.

In our analysis, the evolution of daily infection rate plays a central role. It is clear that for some viruses it can be decreased by development of immunity in the population. This is typical for epidemics with a small contamination delay Δ, resulting in infecting sufficiently large portion of the population (several percents, at least). The main motivations for our research are the epidemics with large contamination delay, say $\Delta \in \{7, \ldots, 14\}$, having in mind COVID19 as the main example. In this case, the daily infection rate $\gamma(\cdot)$ can be constant during a long period of time. And it can be changed to another constant artificially, by introducing some behavioral restrictions in the society. This is the reason why we start from analyzing the development of epidemics subject to a constant daily infection rate.

Let us fix some initial interval $[a-\Delta, a]$ with some $a \in \mathbb{Z}$, and let γ be a positive constant. Consider the discrete-time function $q(\cdot)$, $k \geq a-\Delta$, which is a solution to the following difference equation of order $\Delta+1$ with $\Delta \geq 1$ (compare with (3.1)) :

$$\begin{array}{rcl} q(k+1) & = & (1+\gamma)q(k) - \gamma q(k-\Delta), \quad k \geq a, \\ q(k) & = & \phi(k), \quad k \in \{a-\Delta, \ldots, a\}. \end{array} \tag{3.3}$$

Here, the initial conditions at the interval $[a-\Delta, a]$ are given by function $\phi(\cdot)$. For corresponding trajectory $q(\cdot)$, we use notation

$$q \;=\; \mathcal{F}_{a,\Delta}(\phi, \gamma).$$

From the theory of difference equations, we know that all solutions of (3.3) can be expressed in terms of the roots (complex and real) of the following characteristic equation:

$$p_{\gamma,\Delta}(r) \stackrel{\text{def}}{=} r^{\Delta+1} - (1+\gamma)r^{\Delta} + \gamma \;=\; 0, \quad r \in \mathbb{C}. \tag{3.4}$$

Namely, let $r_1, \ldots, r_{\Delta+1}$ be the *simple* roots of this polynomial. Then any solution $q(\cdot)$ to (3.3) can be represented as follows:

$$q(k) \;=\; \sum_{i=1}^{\Delta+1} c_i r_i^k, \quad k \in \mathbb{Z}, \tag{3.5}$$

where the complex coefficients c_i can be found from the initial conditions in (3.3). Positions of these roots at complex plain $\mathbb{C}$ are essential for behavior of trajectory $q(\cdot)$. At the

same time, if some root r_i of polynomial (3.4) has multiplicity k, then the corresponding sum in (3.5) must be replaced by $p(k)r_i^k$, where $p(\cdot)$ is a polynomial of degree $k-1$ with coefficients defined by the boundary conditions.

Note that the equation (3.3) always has a unit root $r_1 = 1$. However, it also has another positive real root $r(\gamma, \Delta)$, which is responsible for the asymptotic behavior of the trajectory.

Lemma 1 *Equation (3.4) has two positive real roots, one at unity, and the second one $r = r(\gamma, \Delta) > 0$ satisfying the equation*

$$r^{\Delta} \;=\; \gamma(1 + \ldots + r^{\Delta-1}). \tag{3.6}$$

Its position with respect to unity is defined by the following conditions.

- $\gamma\Delta > 1$ *if and only if* $r(\gamma, \Delta) > 1$.
- $\gamma\Delta < 1$ *if and only if* $r(\gamma, \Delta) < 1$.
- $\gamma\Delta = 1$ *if and only if* $r = 1$ *is the root of equation (3.4) with multiplicity two.*

Proof:
Since zero does not satisfy equation (3.4), we can rewrite it in an equivalent form by dividing by r^{Δ}:

$$r + \gamma r^{-\Delta} \;=\; 1 + \gamma. \tag{3.7}$$

This means that

$$r - 1 \;\stackrel{(3.7)}{=}\; \gamma\,(1 - r^{-\Delta}) \;=\; \gamma\, r^{-\Delta}(r^{\Delta} - 1),$$

and (3.6) follows.

Further, denote by $\xi(r) = r + \gamma r^{-\Delta}$ the left-hand side of equation (3.7). Note that it is satisfied by $r = 1$ since $\xi(1) = 1 + \gamma$. On the other hand,

$$\xi'(1) \;=\; 1 - \gamma\Delta.$$

Therefore, if $\xi'(1) < 0$, then there exists a unique other root $r(\gamma, \Delta) > 1$. If $\xi'(1) > 0$, then this root also does exist, but now $r(\gamma, \Delta) < 1$.

Finally, if $\xi'(1) = 0$, then

$$p'_{\gamma,\Delta}(r) \;\stackrel{(3.4)}{=}\; (\Delta + 1)r^{\Delta} - \Delta(1 + \gamma)r^{\Delta-1} \;=\; (\Delta + 1)r^{\Delta-1}(r - 1).$$

Hence, $r = 1$ is a simple root of the derivative. Consequently, it is a double root of the polynomial $p_{\gamma,\Delta}(\cdot)$. □

Sometimes we need to use simple bounds for the position of the root $r(\gamma, \Delta)$. Denote

$$r_* \;\equiv\; r_*(\gamma, \Delta) \;\stackrel{\rm def}{=}\; (\gamma\Delta)^{\frac{1}{1+\Delta}}.$$

This is the minimum of function $\xi(\cdot)$ introduced in the proof of Lemma 1.

Corollary 1 *For any $\Delta \geq 1$ and $\gamma > 0$ we have*

$$r_*^2 \;\leq\; r(\gamma,\Delta) \;\leq\; \max\{r_*, \gamma\Delta\}. \tag{3.8}$$

Moreover, if $\gamma\Delta \geq 1$, then we have a simpler lower bound:

$$r \;\geq\; 1 + \tfrac{1}{\Delta}(\gamma\Delta - 1). \tag{3.9}$$

Proof:
Indeed, using the inequality with arithmetic and geometric mean, for $r = r(\gamma,\Delta)$ we have

$$r^\Delta \overset{(3.6)}{=} \gamma \sum_{j=0}^{\Delta-1} r^j \;\geq\; \gamma\Delta \cdot r^{\frac{\Delta-1}{2}},$$

and this is the first inequality in (3.8). Further, if $\gamma\Delta \leq 1$, then $r \leq r_* = \max\{r_*, \gamma\Delta\}$. If $\gamma\Delta \geq 1$, then $r \geq 1$ and

$$r^\Delta \overset{(3.6)}{=} \gamma \sum_{j=0}^{\Delta-1} r^j \leq\; \gamma\Delta \cdot r^{\Delta-1}.$$

Thus, $r \leq \gamma\Delta = \max\{r_*, \gamma\Delta\}$, and this finishes the proof of the second inequality in (3.8).

Finally, if $\gamma\Delta \geq 1$, then by Lemma 1 we have $r \geq 1$ and $r^\Delta \overset{(3.6)}{\geq} \gamma\Delta$. Consequently,

$$\tfrac{1+\gamma}{r} \overset{(3.7)}{=} 1 + \tfrac{\gamma}{r^{\Delta+1}} \;\leq\; 1 + \tfrac{1}{r\Delta},$$

and this is (3.9). □

Remark 1 *For numerical evaluation of $r = r(\gamma,\delta)$, it is better to rewrite equation (3.6) in the following form (we assume that $\Delta \geq 2$):*

$$\sum_{j=1}^{\Delta-1} r^{-j} \;=\; \tfrac{1}{\gamma}. \tag{3.10}$$

The left-hand side of this equation is a convex decreasing function of r. Hence, it can be solved very efficiently by univariate Newton method starting from the initial approximation $r \approx r_*^2$ (see, for example, Appendix A1 in [6]). □

We will see that the asymptotic behavior of any solution to (3.3) is defined by the position of the positive root $r(\gamma,\Delta)$. Thus, in view of Lemma 1, the value $\gamma\Delta$ describes the *modality* of the process (3.3).

Definition 2

- *If $\gamma\Delta > 1$, we say that the process (3.3) is in a* propagation *mode.*
- *If $\gamma\Delta < 1$, we say that the process (3.3) is in a* recession *mode.*
- *If $\gamma\Delta = 1$, we say that the process (3.3) is in a* stagnation *mode.*

Thus, the role of the product $\gamma\Delta$ is similar to the role of basic reproduction ratio R_0 in SIR model [5]. However, in our case the conclusion on modality of process (3.3) is derived from the *observable* characteristic γ (Δ is usually fixed). This is important for the online analysis of the epidemic processes.

In order to support the above terminology, we need to study asymptotic behavior of the solutions of equation (3.3). For that, let us introduce one more characteristic.

Definition 3 *Function*

$$s(k) \stackrel{\text{def}}{=} q(k+1) - r(\gamma,\Delta)\, q(k), \quad k \geq a, \tag{3.11}$$

is called the sliding residual *of the trajectory* $q = \mathcal{F}_{a,\Delta}(\phi,\gamma)$.

Lemma 2 *For any* $k \geq a$ *and* $r = r(\gamma,\Delta)$, *we have*

$$s(k) \;=\; \tfrac{1}{1+\ldots+r^{\Delta-1}}\Bigl(s(k-1) + rs(k-2) + \ldots + r^{\Delta-1}s(k-\Delta)\Bigr). \tag{3.12}$$

Proof:
Indeed,

$$q(k+1) \stackrel{(3.3)}{=} (1+\gamma)q(k) - \gamma q(k-\Delta) \stackrel{(3.7)}{=} (r+\gamma r^{-\Delta})q(k) - \gamma q(k-\Delta).$$

Hence,

$$\begin{aligned} s(k) &\stackrel{(3.11)}{=} \gamma r^{-\Delta}(q(k) - r^{\Delta}q(k-\Delta)) \\ &= \gamma r^{-\Delta}\Bigl(q(k) - rq(k-1) + r(q(k-1) - rq(k-2)) + \\ &\qquad \ldots + r^{\Delta-1}(q(k-\Delta+1) - rq(k-\Delta)\Bigr) \\ &= \gamma r^{-\Delta}\Bigl(s(k-1) + rs(k-2) + \ldots + r^{\Delta-1}s(k-\Delta)\Bigr). \end{aligned}$$

It remains to note that $\frac{1}{\gamma}r^{\Delta} \stackrel{(3.6)}{=} 1 + \ldots + r^{\Delta-1}$. □

Lemma 2 allows us to bound the trajectory $q(\cdot)$ using its initial conditions. First of all, let us look at deviations from the main trend, defined by the rate $r = r(\gamma,\Delta)$. Denote

$$\begin{aligned} \ell_{a,\Delta}(\phi,\gamma) &= \min_{a-\Delta\leq k\leq a-1}\{\phi(k+1) - r\phi(k)\}, \\ M_{a,\Delta}(\phi,\gamma) &= \max_{a-\Delta\leq k\leq a-1}\{\phi(k+1) - r\phi(k)\}, \end{aligned} \tag{3.13}$$

where $r = r(\gamma,\Delta)$. Then we can establish the following useful statement, which is a simple corollary of Lemma 2.

Theorem 3 *Let* $q = \mathcal{F}_{a,\Delta}(\phi,\gamma)$. *Then for its sliding residuals we have the following bounds:*

$$\ell_{a,\Delta}(\phi,\gamma) \;\leq\; s(k) \;\leq\; M_{a,\Delta}(\phi,\gamma), \quad k \geq a-\Delta. \tag{3.14}$$

Proof:
Indeed, for $k \in \{a-\Delta, \dots, a-1\}$, these bounds are valid by definition (3.13). Further, for $k = a$, relations (3.14) are valid in view of representation (3.12). And we can continue by induction. $\square$

In fact, it is possible to prove that the sequence (3.11) always converges to a single value, defined by the initial conditions $\phi(\cdot)$. For proving this important result, we need to represent the sequence of sliding residuals in terms of stochastic matrices and prove for them several auxiliary statements. This is the subject of the next section.

4 Markov processes and consensus development

Let us choose an arbitrary vector c from the standard simplex:

$$c \;\in\; \mathbb{S}_n \stackrel{\rm def}{=} \{x \in \mathbb{R}^n_+ : \; \langle \bar{e}_n, x \rangle = 1\}.$$

Then we can define the following *companion matrix:*

$$A(c) \;=\; \begin{array}{|c|c|}\hline \multicolumn{2}{|c|}{c^T} \\ \hline I_{n-1} & 0_{n-1} \\ \hline \end{array} \;\in\; \mathbb{R}^{n \times n}.$$

Note that matrix $A(c)$ is *row-stochastic*:

$$A(c)\bar{e}_n \;=\; \bar{e}_n. \tag{4.1}$$

This means, that the transposed matrix $A^T(c)$ is column-stochastic. By Perron-Frobenius Theorem, it has a nontrivial eigenvector $y_*(c)$ for the unit eigenvalue:

$$A^T(c) y_*(c) \;=\; y_*(c) \;\in\; \mathbb{S}_n. \tag{4.2}$$

In our case, the components of this vector have a simple closed-form representation:

$$y_*^{(i)}(c) \;=\; \frac{1}{B} \sum_{j=i}^{n} c^{(i)}, \quad i = 1, \dots, n, \quad B \;=\; \sum_{i=1}^{n} i\, c^{(i)}. \tag{4.3}$$

Our interest to the matrix $A(c)$ is explained by the following representation. Let us choose $r = r(\gamma, \Delta)$, and $n = \Delta$. Then for the vector $c \in \mathbb{S}_n$ with the components

$$c^{(i)} \;=\; \frac{r^{i-1}}{1 + \dots + r^{\Delta - 1}}, \quad 1 \le i \le \Delta, \tag{4.4}$$

we can define the companion matrix $A(c)$. Consider now the following *consensus process*:

$$x_0 \;\in\; \mathbb{R}^n, \quad x_{k+1} \;=\; A(c) x_k, \quad k \ge 0. \tag{4.5}$$

Then, for the choice $x_0^{(i)} = s(a-i)$, $i = 1, \dots, \Delta$, all components of the vectors x_k will be equal to the corresponding elements of the sliding window of length Δ from the sequence $\{s(k)\}_{k \ge a-\Delta}$ (see representation (3.12)).

Our current goal is to prove that, for process (4.5), the components of vectors $\{x_k\}_{k\geq 0}$ converge to the same value ξ_* with a linear rate. Note that

$$\langle y_*(c), x_{k+1}\rangle \;=\; \langle y_*(c), A(c)x_k\rangle \;=\; \langle A^T(c)y_*(c), x_k\rangle \;\stackrel{(4.2)}{=}\; \langle y_*(c), x_k\rangle. \tag{4.6}$$

Hence, the only possible common target for the values $x_k^{(i)}$, $i = 1, \ldots, n$, is the following weighted average of the components of vector x_0:

$$\xi^* \;\stackrel{(4.2)}{=}\; \langle y_*(c), x_0\rangle. \tag{4.7}$$

We acjieve our goal in several steps. Firstly, let us introduce vectors b_k with the following structure:

$$b_k \;=\; (c^{(k)}, \ldots, c^{(1)}, 1, 0, \ldots, 0)^T \in \mathbb{R}^n, \quad k = 1, \ldots, n-1. \tag{4.8}$$

Lemma 3 *For any $k = 1, \ldots, n-1$, we have*

$$A^k(c)e_1 \;\geq\; b_k. \tag{4.9}$$

Proof:
Indeed, for $k = 1$, we have $A(c)e_1 = b_1$, and therefore (4.9) is valid. Assume it is valid for some $k \geq 1$. Then, in view of presence in matrix $A(c)$ the down-shifting block I_{n-1}, we have

$$[A^{k+1}(c)e_1]^{(i+1)} \;=\; [A^k(c)e_1]^{(i)} \;\stackrel{(4.9)_k}{\geq}\; b_k^{(i)} \;=\; b_{k+1}^{(i+1)}, \quad i = 1, \ldots, n-1.$$

It remains to note that

$$[A^{k+1}(c)e_1]^{(1)} \;=\; \langle c, A^k(c)e_1\rangle \;\stackrel{(4.9)_k}{\geq}\; \langle c, b_k\rangle \;\stackrel{(4.8)}{\geq}\; c^{(k+1)}.$$

Thus, we have proved that $A^{k+1}(c)e_1 \geq b_{k+1}$. □

Let us assume that

$$c_{\min} \;=\; \min_{1\leq i\leq n-1} c^{(i)} \;>\; 0. \tag{4.10}$$

Then, in view of Lemma 3, we have the following inequality:

$$A^{n-1}(c)e_1 \;\geq\; c_{\min}\bar{e}_n. \tag{4.11}$$

Now, let us consider one iteration of the *Markov process*

$$\bar{y} \;\in\; \mathbb{S}_n, \quad y_+ \;=\; B\bar{y}, \tag{4.12}$$

defined by a column-stochastic matrix $B \in \mathbb{R}^{n\times n}$. Since B is column stochastic, there exists a vector $y_* \in \mathbb{S}_n$ such that $By_* = y_*$. For the reader's convenience, we present here the following result from [7] with simple proof.

Lemma 4 *Let $B^T e_1 \geq \sigma\bar{e}_n$ with some $\sigma \in (0, 1)$. Then, for the iteration (4.12), we have*

$$\|y_+ - y_*\|_{(1)} \;\leq\; (1-\sigma)\|\bar{y} - y_*\|_{(1)}. \tag{4.13}$$

Proof:
Consider the matrix $\hat{B} = \frac{1}{1-\sigma}(B - \sigma e_1 \bar{e}_n^T) \geq 0$. Then $B = (1-\sigma)\hat{B} + \sigma e_1 \bar{e}^T$. Note also that

$$\hat{B}^T \bar{e}_n \;=\; \tfrac{1}{1-\sigma}(B^T - \sigma \bar{e}_n e_1^T)\bar{e}_n \;=\; \bar{e}_n.$$

Therefore, for any vector u with $\langle \bar{e}_n, u \rangle = 0$, we have

$$\begin{aligned} \|Bu\|_{(1)} \;&=\; (1-\sigma)\|\hat{B}u\|_{(1)} \;\leq\; (1-\sigma)\langle \hat{B}|u|, \bar{e}_n \rangle \\ &=\; (1-\sigma)\langle |u|, \hat{B}^T \bar{e}_n \rangle \;=\; (1-\sigma)\|u\|_{(1)}. \end{aligned}$$

Hence, $\|y_+ - y_*\|_{(1)} \stackrel{(4.12)}{=} \|B(\bar{y} - y_*)\|_{(1)} \leq (1-\sigma)\|\bar{y} - y_*\|_{(1)}$. □

We are going to use Lemma 4 for analyzing the process (4.5) with $B^T = A^{n-1}(c)$. Note that by (4.11), the condition of this lemma is satisfied.

Theorem 4 *Let sequence $\{x_k\}_{k\geq 0}$ be generated by the consensus process*

$$x_0 \;\in\; \mathbb{R}^n, \quad x_{k+1} \;=\; Ax_k, \quad k \geq 0, \tag{4.14}$$

where A is a row-stochastic matrix with positive left-dominant eigenvector:

$$A^T y_* \;=\; y_* \;\in\; \operatorname{rint} \mathbb{S}_n. \tag{4.15}$$

Suppose that for some integer $m \geq 1$ we have

$$A^m e_1 \;\geq\; \sigma \bar{e}_n \tag{4.16}$$

with certain $\sigma \in (0,1)$. Then the process (4.14) has the linear rate of convergence,

$$\|x_k - \xi_* \bar{e}_n\|_A \;\leq\; \tfrac{4}{1-\sigma}(1-\sigma)^{k/m} \cdot \|x_0\|_{(\infty)}, \quad k \geq 0, \tag{4.17}$$

with respect to the norm $\|x\|_A = \sum\limits_{i=1}^n y_*^{(i)} |x^{(i)}|$, where $\xi_* = \langle y_*, x_0 \rangle$.

Proof:
Note that all points of the process (4.14) belong to the same hyperplane:

$$\langle y_*, x_k \rangle \;=\; \langle y_*, A^k x_0 \rangle \;=\; \langle [A^T]^k y_*, x_0 \rangle \;\stackrel{(4.15)}{=}\; \langle y_*, x_0 \rangle \;=\; \xi_*. \tag{4.18}$$

Let us fix an arbitrary $y_0 \in \mathbb{S}_n$, and define the sequence $y_k = [A^T]^k y_0$, $k \geq 0$. Then by Lemma 4, for $B^T = A^m$ we have

$$\|y_{k+m} - y_*\|_{(1)} \;=\; \|B(y_k - y_*)\|_{(1)} \;\stackrel{(4.13)}{\leq}\; (1-\sigma)\|y_k - y_*\|_{(1)}.$$

At the same time, for $0 \leq k \leq m$, we have

$$\|y_k - y_*\|_{(1)} \;\leq \|y_k\|_{(1)} + \|y_*\|_{(1)} \;\leq\; 2.$$

Hence, by induction, it is easy to check that

$$\|y_k - y_*\|_{(1)} \;\leq\; \tfrac{2}{1-\sigma}(1-\sigma)^{k/m}, \quad k \geq 0. \tag{4.19}$$

This means that

$$\begin{aligned} \langle y_0, x_k\rangle - \xi_* &\overset{(4.18)}{=} \langle y_0 - y_*, x_k\rangle \;=\; \langle [A^T]^k (y_0 - y_*), x_0\rangle \;=\; \langle y_k - y_*, x_0\rangle \\ &\overset{(4.19)}{\leq} \tfrac{2}{1-\sigma}(1-\sigma)^{k/m} \cdot \|x_0\|_{(\infty)}. \end{aligned}$$

Since y_0 is an arbitrary point from the standard simplex $\mathbb{S}_n$, we have

$$\max_{1\leq i\leq n} x_k^{(i)} - \xi_* \;\leq\; \tfrac{2}{1-\sigma}(1-\sigma)^{k/m} \cdot \|x_0\|_{(\infty)} \overset{\text{def}}{=} \delta_k.$$

Thus, we conclude that

$$\begin{aligned} \|x_k - \xi_* \bar{e}_n\|_A &= \sum_{i=1}^n y_*^{(i)} \cdot \left|\xi_* - x_k^{(i)} + \delta_k - \delta_k\right| \\ &\leq \sum_{i=1}^n y_*^{(i)} \cdot \left(\xi_* - x_k^{(i)} + \delta_k + \delta_k\right) \overset{(4.18)}{=} 2\delta_k. \qquad \square \end{aligned}$$

5 Bounds for different epidemic developments

Let us consider three examples of application of Theorem 3 and Theorem 4 to the analysis of epidemic dynamics.

1. Propagation mode. We assume that $\gamma\Delta > 1$. Hence, by Lemma 1, we have $r = r(\gamma, \Delta) > 1$. Consider a propagation mode with constant daily infection rate:

$$\gamma(d) \;\equiv\; \gamma > 0, \quad d \geq -\Delta.$$

Let us assume that $T(d) = 0$ for $-\Delta \leq d \leq -1$ and $T(0) = 1$. Then

$$\begin{aligned} T(1) &\overset{(3.1)}{=} (1+\gamma)T(0) - \gamma T(-\Delta) \;=\; 1+\gamma, \\ &\cdots \\ T(\Delta) &= (1+\gamma)T(\Delta-1) - \gamma T(-1) \;=\; (1+\gamma)^\Delta. \end{aligned} \tag{5.1}$$

This is the end of the initial propagation phase, after which we have to switch onto a slower growth defined by the rate $r = r(\gamma, \Delta) \overset{(3.7)}{\leq} 1+\gamma$.

Let us write down the lower and upper bounds for the dynamics of total number of cases $T(\cdot)$. For that, we need to define the initial conditions

$$\phi(d) \;=\; (1+\gamma)^d, \quad 0 \leq d \leq \Delta,$$

and compute the bounds (3.14) for its sliding residuals. Note that

$$\phi(d+1) - r\phi(d) \;=\; (1+\gamma)^d(1+\gamma-r), \quad 0 \leq d \leq \Delta - 1.$$

Therefore,

$$\begin{array}{rcl} \ell & \stackrel{\text{def}}{=} & \min\limits_{0\le d\le\Delta-1}[\phi(d+1)-r\phi(d)] \;=\; 1+\gamma-r \;\stackrel{(3.7)}{=}\; \gamma r^{-\Delta}, \\ M & \stackrel{\text{def}}{=} & \max\limits_{0\le d\le\Delta-1}[\phi(d+1)-r\phi(d)] \;=\; (1+\gamma)^{\Delta-1}(1+\gamma-r) \\ & \stackrel{(3.7)}{=} & \gamma(1+\gamma)^{\Delta-1}r^{-\Delta} \;=\; \frac{\gamma}{1+\gamma}\left(\frac{1+\gamma}{r}\right)^{\Delta} \;\stackrel{(3.9)}{\le}\; \frac{\gamma}{1+\gamma}\left(\frac{1+\gamma}{1+\gamma-\Delta^{-1}}\right)^{\Delta} \\ & \le & \frac{\gamma}{1+\gamma}\left(\frac{1}{1-(1+\gamma)^{-1}}\right) \;=\; 1. \end{array} \tag{5.2}$$

Denoting now $s(d) = T(d+1) - rT(d)$ for $d \ge \Delta$, by Theorem 3, we have

$$\ell \;\le\; s(d) \;\le\; M, \quad d \ge \Delta. \tag{5.3}$$

This means that

$$T(d+1) \;\le\; rT(d) + M \;\le\; \ldots \;\le\; r^{d+1-k}T(k) + M\frac{r^{d+1-k}-1}{r-1}, \quad \Delta \le k \le d.$$

Taking $k = \Delta$, for any $d \ge \Delta$, we obtain

$$T(d) \;\le\; r^{d-\Delta}T(\Delta) + M\frac{r^{d-\Delta}-1}{r-1} \;\le\; r^{d-\Delta}\left(T(\Delta) + \frac{M}{r-1}\right).$$

Similarly, we get

$$T(d) \;\stackrel{(5.3)}{\ge}\; r^{d-\Delta}T(\Delta).$$

Thus, we come to the following bounds.

Lemma 5 *In propagation mode, for any $d \ge \Delta$, we have*

$$\begin{array}{rcl} r^{d-\Delta}T(\Delta) & \le & T(d) \;\le\; r^{d-\Delta}\left(T(\Delta) + \frac{M}{r-1}\right) \\ & \stackrel{(3.9)}{\le} & r^{d-\Delta}\left(T(\Delta) + \frac{\Delta}{\gamma\Delta-1}\right). \end{array} \tag{5.4}$$

At the same time, using Theorem 4, we can compute the limiting value of the sliding residual:

$$\lim_{d\to\infty}[T(d+1) - rT(d)] \;=\; \xi_* \;=\; \langle y_*, x_0\rangle, \tag{5.5}$$

where y_* can be obtained by (4.3), (4.4), and the initial sliding residuals are defined as

$$x_0^{(i)} \;=\; T(\Delta+1-i) - rT(\Delta-i) \;\stackrel{(5.1)}{=}\; \frac{1-(-\tau)^{n+2-i}}{1+\tau}(1+\gamma)^{n+1-i}, \tag{5.6}$$

for $i = 1, \ldots, \Delta$, with $\tau = \frac{r}{1+\gamma} \in (0,1)$. Note that $x_0^{(i)} > x_0^{(i+1)}$, $i = 1, \ldots, \Delta-1$.

2. Stagnation mode. In this case $\gamma\Delta = 1$ and $r(\gamma,\Delta) = 1$. Hence, the growth of the total number of cases $T(\cdot)$ is defined by the sliding residuals

$$s(d) \;=\; T(d+1) - T(d), \quad d \ge 0.$$

Let us compute its limiting value. In view of equations (4.4) and (4.3), we have $c^{(i)} = \frac{1}{\Delta}$, $i = 1, \ldots, \Delta$, $B = \frac{\Delta+1}{2}$, and

$$y_*^{(i)} \;=\; \tfrac{2(\Delta+1-i)}{\Delta(1+\Delta)}, \quad i = 1, \ldots, \Delta. \tag{5.7}$$

Applying these weights to the initial sliding residual, we obtain the limiting value

$$\begin{aligned} \xi_* \;&=\; \sum_{i=1}^{\Delta} y_*^{(i)} \Bigl[T(1-i) - T(-i) \Bigr] \\ &\overset{(5.7)}{=}\; \tfrac{2}{\Delta+1} \Bigl[T(0) - \tfrac{1}{\Delta} \sum_{i=1}^{\Delta} T(-i) \Bigr]. \end{aligned} \tag{5.8}$$

Thus, in this mode we observe a stabilization of our main characteristics:

$$\lim_{d \to \infty} C(d) \;=\; \xi_*, \quad \lim_{d \to \infty} H(d) \;=\; \xi_* \Delta, \quad \lim_{d \to \infty} I(d) \;=\; \xi_*. \tag{5.9}$$

3. Recession mode. In this mode, $\gamma\Delta < 1$ and $r = r(\gamma, \Delta) < 1$. Hence, it leads to the end of epidemics. The good news is that this can happen without development of immunity in the population.

Let us compute the main characteristic of this mode, the limiting value for the total number of cases. Assume that the mode started at day $d = 0$. Define the sliding residuals

$$s(d) \;=\; T(d+1) - rT(d), \quad d \geq -\Delta. \tag{5.10}$$

In accordance to Theorem 3, they are bounded from above:

$$s(d) \;\leq\; M \;\overset{\text{def}}{=}\; \max_{1 \leq i \leq \Delta} s(-i), \quad d \geq 0. \tag{5.11}$$

Therefore, for $d \geq 0$ and $0 \leq k \leq d$ we have

$$\begin{aligned} T(d+1) \;&\leq\; rT(d) + M \;\leq\; r(rT(d-1) + M) + M \\ &\leq\; \ldots \;\leq\; r^{k+1} T(d-k) + M(1 + r + \ldots + r^k). \end{aligned}$$

Since $T(\cdot)$ is an increasing function, taking $k = d$, we prove existence of the limit

$$T_* \;\overset{\text{def}}{=}\; \lim_{d \to \infty} T(d),$$

which satisfies the inequality $T_* \leq \frac{M}{1-r}$. Moreover, in view of Theorem 4, we can compute the exact value T_*, using prehistory $\{s(-i)\}_{i=1}^{\Delta}$. Indeed, taking the limit in (5.11), we get

$$T_* \;=\; \tfrac{1}{1-r} \lim_{d \to \infty} s(d) \;=\; \tfrac{\xi_*}{1-r} \;=\; \tfrac{1}{1-r} \langle y_*, x_0 \rangle, \tag{5.12}$$

where y_* is defined by (4.3), (4.4), and $x_0^{(i)} = s(-i)$, $i = 1, \ldots, \Delta$. The rate of this convergence is given by (4.17) with $m = n - 1$ and $\sigma \overset{(4.10)}{=} \frac{1}{1 + \ldots + r^{\Delta-1}} \min\limits_{0 \leq i \leq \Delta-2} r^i \overset{(3.6)}{=} \frac{\gamma\Delta}{r^2}$.

6 Online prediction. Belgian case study

In Section 2, we have seen that availability of the full historical data $\{T(d)\}_{d\geq 0}$ is sufficient for reconstructing all other hidden elements of our model, namely

$$\{C(d)\}_{d\geq 0}, \quad \{H(d)\}_{d\geq 0}, \quad \{I(d)\}_{d\geq 0}. \tag{6.1}$$

Moreover, we can compute also the daily infection rates

$$\gamma(d) \overset{(2.10)}{=} \frac{T(d+\Delta)-T(d+\Delta-1)}{T(d+\Delta-1)-T(d-1)}, \quad d \geq 0. \tag{6.2}$$

This possibility is very useful for a retrospective analysis of epidemics. At the same time, it can be used also for *online prediction* of its development.

Indeed, having the full history of epidemics available up to some day $k \geq 0$, that is the data $\{T(i)\}_{i=0}^{k}$, we can compute the *exact value* of the infection rate $\gamma(k-\Delta)$ by the expression (6.2). Then, assuming that this rate was not changed for all $d \geq k-\Delta$, we are able to form a forecast for the development of the epidemic in the future.

Of course, we should have serious reasons for this constancy assumption. The following two ones are the most common.

- The total number of infected persons is not sufficient for developing a significant immunity level of the population.
- The level of infection rate is mostly related to the behavioral habits of population. It can be changed only by restrictions introduced by the central government.

Note that both reasons are well present in the COVID19 epidemic.

For the practical verification of our model, we choose Belgium, a medium-size European country, which was affected very seriously by the epidemic. At the moment of releasing this paper (end of May 2020), it takes the third position in the number of cases per one million of population in the Western World (after Spain and the USA). Moreover, it has the highest relative mortality rate in the world. At the same time, its population has a high level of social responsibility, which helps to control the infection rate and maintain it on a constant level. It is important that, contrary to some other countries, the official statistics on development of epidemics in Belgium was reported very accurately daily around noon.[1]

To the best of our knowledge, we present here the first attempt of online analysis of epidemic development. Hence, we explain all our steps in details, sufficient for implementing the same strategy for other countries.

For applying our model to a real-life situation, we need to choose properly its parameters. They have different importance and influence on the quality of predictions.

- **Contamination delay Δ.** This is the key parameter of the model. A priori, it could be any integer greater than one. However, in our situation, from the external information it was clear that the right number is something between seven and twelve.

[1] We discuss several exceptions and their consequences later.

- **Daily infection rate $\gamma(\cdot)$.** This parameter is responsible for the rate of development of the epidemic. The rate of growth of the number of cases is extremely sensitive to its values. Hence, even a small inaccuracy in its evaluation immediately leads to a quite visible difference between the predictions and the real data. At the same time, due to the presence of contamination delay, our prediction abilities heavily depend on the volatility of $\gamma(\cdot)$. The predictions are good if $\gamma(\cdot)$ remains constant for sufficiently long periods of time. Thus, one of the goals of our case study was verification of the hypothesis that this function can have indeed a piece-wise constant behavior.

Let us describe our online prediction strategy. It is based on information $\{T(d)\}_{d\geq 0}$, reported daily at the same hour. We construct in parallel our predictions $T_p(\cdot)$, which should approximate the real data $T(\cdot)$.[2] These predictions are computed by a piece-wise constant daily infection rate $\gamma_p(\cdot)$. If we decide that this function takes some value γ for all days after day a, then we generate the trajectory

$$T_p(d+1) \stackrel{(3.1)}{=} T_p(d) + \gamma(T_p(d) - T_p(d-\Delta)), \quad d \geq a. \tag{6.3}$$

Using these predictions, we can compute all other elements of our model:

$$\begin{aligned} C_p(d) &= T_p(d) - T_p(d-1), \quad H_p(d) = T_p(d+\Delta-1) - T_p(d-1), \\ I_p(d) &= C_p(d+\Delta), \quad d \in \mathbb{Z}. \end{aligned} \tag{6.4}$$

We have two criterions for checking the quality of our predictions. This verification can be done at every day $d \geq 0$ after receiving a new measurement $T(d)$. Firstly, we can compare our prediction with the actual data:

$$T_p(d) \approx T(d). \tag{6.5}$$

Secondly, we can compare γ with the actual value of $\gamma(\cdot)$:[3]

$$\gamma \approx \gamma(d-\Delta) \stackrel{(3.1)}{=} \frac{T(d)-T(d-1)}{T(d-1)-T(d-\Delta-1)}. \tag{6.6}$$

If the quality of approximation is good enough, we continue with the same γ. However, if we see sufficient divergence with the predictions for several consecutive days,[4] then the parameter γ has to be changed in order to fit the new regime.

Clearly, the above strategy provide us with interesting predictions only if the periods of constancy of function $\gamma(\cdot)$ are long enough as compared with the contamination delay. Let us present the results of our analysis, which show that for COVID19 epidemics this is indeed true.

Let us show how all of that was implemented for the Belgian case. All our conclusions are based on the real-life data presented in Appendix.

[2] For all $d < 0$, we suppose that $T(d) = 0$ and $T_p(d) = 0$.

[3] We assumer that Axiom 1 is valid. Therefore, the right-hand side of this equation is the *exact* infection rate for the corresponding day.

[4] Usually, inaccurate results for two-three days is a sufficient reason for updating γ.

The first step in our modelling consists in estimating the contamination delay $\Delta \geq 1$. From the external sources, it was known that this is a number between seven and twelve. Using these values and the initial statistics for the first days of March, we generated several variants of the curve $T_p(\cdot)$, using the recurrence (6.3). For each curve, we tried to adjust manually the parameter $\gamma > 0$ in order to fit well the avalinable date on function $T(\cdot)$. For several values of Δ, the results are shown on Figure 1

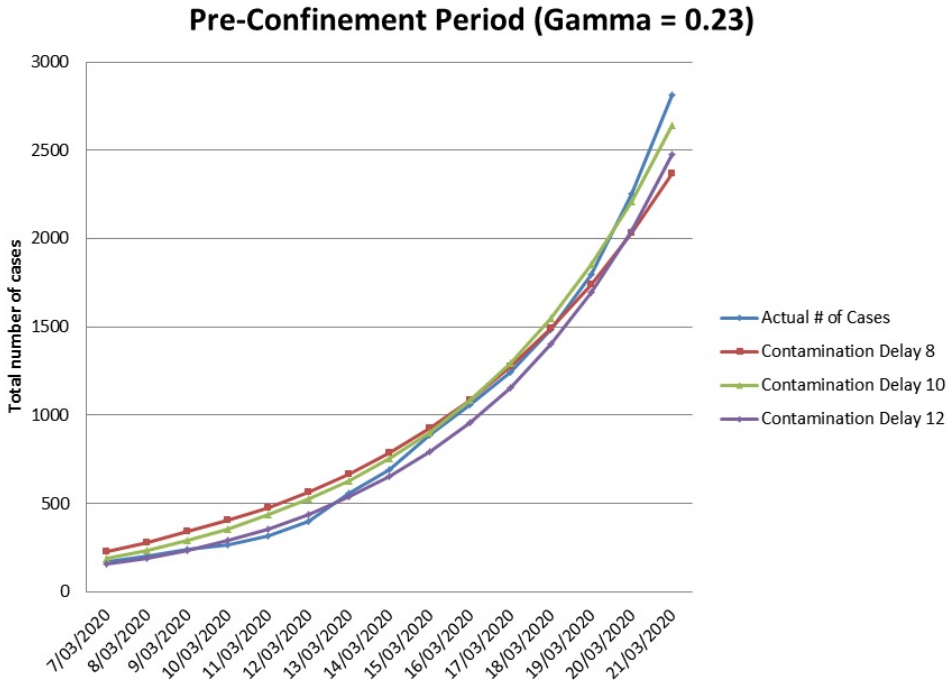

Figure 1: Initial stage: first estimates for Δ and γ.

As we can see, different variants of Δ give very close results, approximating the actual curve within accuracy of $10 - 15\%$. Surprisingly enough, the appropriate infection rate γ is quite low. With this rate, the asymptomatic virus holder infects one new person approximately in four days. Thus, the main danger of this epidemics consists in the long contamination delay. At that moment, it became clear that for estimating it accurately we need more information.

We got a good chance for that in the end of March. Indeed, a fair estimate of the delay can be obtained by observing an influence of a sudden change in the infection rate. This happened in Belgium on March 19, when the whole country was obliged to follow a very strict containment regime.[5] Hence, the delay in reflecting this by statistics, gives us a good value of Δ. For implementing this idea, we generated several estimates of function $\gamma(\cdot)$ defined by (6.6), which correspond to different contamination delay. The delay ensuring the abrupt change of γ *exactly* at March 19 was chosen as the right one. You can see our results at Figure 2.

[5]In fact, the regime officially started at the noon of March 18. However, we consider March 19 as the first full day of containment.

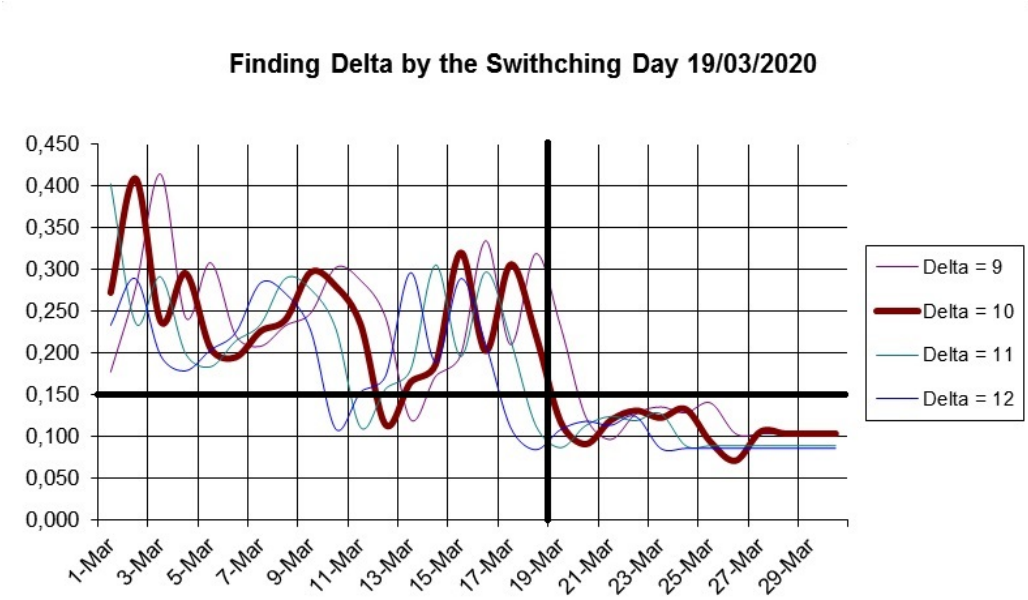

Figure 2: Dynamics of different infection rates around the Switching Day.

From this picture, it is clear that the right variant corresponds to $\Delta = 10$. This delay was fixed in our model for all consequent forecasts. We will see that this constant explains properly the evolution of statistics for the whole period of our analysis.

Thus, starting from March 30, for our online analysis we are in much better conditions: the contamination delay is fixed and it is possible to concentrate only on the correct estimates of the daily infection rate. Therefore, we introduced in our model the *exact data* for the period March 1-29:

$$T_p(d) \quad = \quad T(d), \quad d \leq 29/03/2020.$$

For the consequent days, we apply our prediction strategy (6.3) - (6.6). The results of our predictions are shown at Figure 3. It displays four curves. Three of them are the predictions of our model for corresponding days in the future:

- Number of new cases $C_p(\cdot)$ (blue line),
- Number of asymptomatic virus holders $H_p(\cdot)$ (grey line),
- Total number of infected persons $T_p(\cdot)$ (pink line).

The quality of our prediction is compared with the actual data $T(\cdot)$ (light blue crosses). As we can see, the predictions are *very exact*, usually the fit the range of 0.5%. So, the difference is almost invisible at the picture.

Let us discuss now day-to-day evolution of our model in this period of time, shown at Figure 3. It was filled progressively, using the regular flow of new data. All estimates for daily infection rates were computed in the beginning of the corresponding periods using the data generated by the model. The only exception is the pre-containment infection rate $\gamma = 0.243$ computed by averaging the estimates (6.6) for the period March 1–29. Recall that for the latter period we use the actual statistical data.

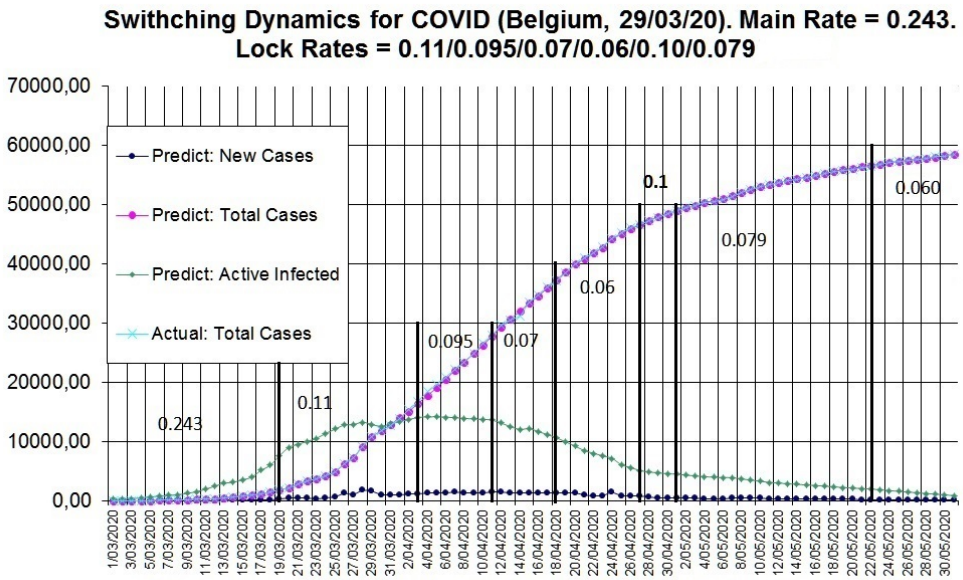

Figure 3: Online prediction for the asymptomatic virus holders

1. Stagnation Phase. It corresponds to the dates March 19 - April 10. This phase is visible in the statistics with a shift of ten days, from March 29 to April 20. Its initial infection rate $\gamma = 0.11$ was selected at April 1st. Note that for Contamination Regime, this starting daily infection rate is surprisingly high. Indeed, in accordance to Definition 2, despite to the strict quarantine measures, we still have a propagation of the epidemic. This propagation is very slow since $\gamma\Delta \approx 1$. So, it is similar to a stagnation mode, which means that every day the number of new cases is approximately constant. This constant plateau can be seen in the statistics with a shift in ten days, in the period from March 30 to April 12.

Starting from April 3, for better fitting the actual data, we change the daily infection rate to $\gamma = 0.095$ (this was done on April 15). Thus, the process got a light recession bias.

This period of three weeks in total corresponds to the highest level of asymptomatic virus holders. In accordance to our estimates, it was almost constant, varying in the interval from thirteen to fourteen thousands.

It is interesting to understand why the quarantine regime can coexist with a propagation mode of epidemic, and for how long this situation could survive. Our main hypothesis is that during this time the locked asymptomatic virus holders were infecting their cohabitants. This gives us a possibility to estimate the length of this period. Indeed, in accordance to our model, on March 19 we had 7648 asymptomatic virus holders and 1795 detected cases. The average size of a family in Belgium is 2.3. However, the size of the standard cohabitation cell must be larger (parents, friends, other cohabitants, etc.). We estimate it roughly by three, and compute the total number of presumable victims of these locked virus holders, that is

$$1795 + 7648 \times 3 \quad = \quad 24739.$$

So, we could expect a drop in the infection rate when the total number of cases will go beyond this number. This happened on April 9. And indeed, starting from April 11 we got a new recession phase.

Note that from the mathematical point of view, the stagnation mode is very exceptional. Indeed, it corresponds to existence of a double root at unity of the algebraic polynomial (3.4). This is an unstable situation, which disappears after an infinitely small variation of the coefficients of this polynomial. However, in our case, this mode managed to survive during three weeks. As compared with other regimes in our analysis, this is almost the maximal duration.

2. First Recession Period. This period starts on April 11 with daily infection rate $\gamma = 0.07$, and terminates on April 26 with $\gamma = 0.06$. It is one of the most successful recession processes of this epidemic. During these two weeks, our estimate of the asymptomatic virus holders was decreased in 2.4 times, from 13658 to 5642.

3. First Exceptional Day. The efficiency of the First Recession Period was slightly disturbed by the celebration of Catholic Easter. Firstly, the official statistics on the number of newly detected cases for April 13-15 was clearly very inaccurate. Secondly, despite to all stay-at-home demands of the government, many people decided to arrange big family meetings. This is clear from our estimate of the coefficient γ for April 14, which is based on the jump in the number of newly detected cases for April 24-25. This gives us one more confirmation of correctness of our estimate for contamination delay $\Delta = 10$.

It is important that this celebration did not change the whole dynamics of the period. For the next days, we used the same infection rates as before. However, since the jump in the infection rate was very big, we introduced manually the corresponding estimate of γ_p directly in our model. This was done on April 25.

3. Four Exceptional Days. Very successful First Recession Period was suddenly terminated by Four Exceptional Days from April 27 to April 30. The importance of these days became clear ten days later, after observing a big jump in the daily statistics for the period May 7-10. The increase in the daily rate between May 6 and May 7 was more than 200%. For the next three days it remained at more or less the same high level, and after that it dropped back on the level of May 4. The reason of this disturbance is not clear yet. Its influence was not local, since after these days the infection rate was stabilized on a higher level of $\gamma = 0.079$, which survived for the next three weeks. It does not look as a consequence of a nation-wide celebration of certain event (like Easter). The power of this shock is comparable only with a massive return to the country of many citizens infected in these days somewhere abroad. But we do not have any definite proof of that. Thus, the actual reason remains interesting for further investigations.

4. Second Recession Period. This recession period started from May 1st and ended on May 21st. So, its consequences were visible in statistics up to the end of the month. It was much less efficient period as compared to the first one (our estimate for its infection rate is $\gamma = 0.079$). However, it still ensured a significant progress in the number of asymptomatic virus holders.

During this period, we observed a higher volatility in our estimates for $\gamma(\cdot)$, computed by (6.6). For several days, like March 12–14 and March 18-19, these estimates were close

to the threshold value $\gamma = 0.1$. However, they returned to the lower level very quickly. Maybe this is a result of de-containment measures gradually introduced by the government starting from May 4. In accordance to our model, the effect of these measures must be visible in the statistics ten days later.

In the end of May, for four good days May 22-25 (reflected by statistics of June 1-4), we even decreased our forecast for the infection rate up to $\gamma = 0.055$. However, starting from May 26 (June 5), statistics becomes worse and worse, indicating either switching to a new stagnation phase, or even the beginning of the second wave of infection. This conclusion will be justified (or not) in the nearest future.

7 Conclusion

In this paper, we presented a new model HIT for analyzing development of epidemics, for which we cannot count on the fast development of immunity in the population. Hence, the classical epidemic models (like SIR) do not work. Nevertheless, we demonstrate that such epidemics can be stopped by quarantine measures. Moreover, using a real-life example of Belgium, we show that the available online information on epidemic development is crucial for an efficient fighting against such diseases.

The main features of our model are as follows.

- For each day of epidemics, it forms predictions for the total number of infected persons and the number of asymptomatic virus holders.
- This is an axiomatic model. All its conclusions are derived rigorously from a single axiom, assuming a constant contamination delay Δ.
- The development of an epidemic can be characterized by one of three different modes (propagation, stagnation, and recession). This modality can be easily learned from the observable parameters of a real-life epidemic process.
- In our model, we have an unusual *Stagnation Mode*, when the number of detected cases $C(\cdot)$ and number of infected persons $I(\cdot)$ are constant. In SIR, this regime is just impossible (see (1.1)).[6] However, from the history of COVID19 epidemics in Belgium, it is clear that the country was in this mode during several weeks.
- Assuming that the infection rate is changing only in discrete moments of time, we can predict the future development of epidemics by the observed data.
- Our experience with online prediction of COVID19 epidemics shows that the daily reported data on the number of infected persons is necessary and sufficient for the correct prediction. Its accuracy is proportional to the accuracy of the data.
- In our experiments with Belgian data, our predictions of the future were usually exact within 0.5% of accuracy.
- Our model is able to predict the important dynamics of asymptomatic virus holders. This information is crucial at the final stage of epidemics. At the same time, it is usually missing in the classical epidemic models, which do not fit well the specifics of COVID19 developments.

[6] It is impossible also in a newer version of SIR, so called SEIR model [3].

- Since our predictions are very accurate, we can predict the beginning of the second wave of infection much better than other models.

References

[1] F. Brauer. Mathematical epidemiology: Past, present, and future. *Infectious Desease Modelling*, **2**, 113-127 (2017).

[2] O. Diekmann, H. Heesterbeek, and T. Britton. Mathematical Tools for Understanding Infectious Disease Dynamics. *Princeton University Press* (2013)

[3] H.W. Hethcote. THe mathematics of infectious deseases. *SIAM Review*, **42**(4), 599-653 (2000).

[4] M. Martcheva. An Introduction to Modern Epidemiology. *Texts in Applied Mathematics, 61*, Springer (2015)

[5] W.O. Kermack and A.G. McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London A: Mathematcal, Physical, and Engineering Sciences*, **115**, 700–721 (1927).

[6] Yu. Neserov. *Lectures on Convex Optimization*. Springer (2018)

[7] Yu. Nesterov, and A. Nemirovski. Finding the stationary states of Markov chains by iterative methods. *Applied Mathematics and Computation*, **255**, 58-65 (2015).

[8] M. Tibayrenc. *Encyclopedia of Infectious Diseases: Modern Methodologies.* J.Willey & Sons, Inc. (2007)

[9] WORLDOMETER. https://www.worldometers.info/coronavirus/

Appendix. Epidemic statistics for Belgium for the period March 1 - May 31, 2020

In the three tables below, we present an official statistics from [9] on the number of COVID19 cases in Belgium for the period March 1 - May 31, 2020. It is given in the columns $T(\cdot)$ and $C(\cdot)$. Column $\gamma(\cdot)$ contains the estimates of the infection rate computed by this data using the formula (6.6).

Up to March 29, the column $T_p(\cdot)$ is filled by the official data. Starting from March 30, it is evaluated by (6.3), using the estimated infection rate $\gamma_p(\cdot)$. The remaining columns are computed by (6.4).

Table 1: Statistics and forecasts for March 2020

Date	$T(\cdot)$	$C(\cdot)$	$\gamma(\cdot)$	$\gamma_p(\cdot)$	$C_p(\cdot)$	$T_p(\cdot)$	$H_p(\cdot)$
1/03	2	2	0.243	0.243	2	2	267
2/03	8	6	0.243	0.243	6	8	312
3/03	13	5	0.243	0.243	5	13	391
4/03	23	10	0.243	0.243	10	23	546
5/03	50	27	0.243	0.243	27	50	666
6/03	109	59	0.243	0.243	59	109	836
7/03	169	60	0.243	0.243	60	169	949
8/03	200	31	0.243	0.243	31	200	1074
9/03	239	39	0.243	0.243	39	239	1286
10/03	267	28	0.243	0.243	28	267	1556
11/03	314	47	0.243	0.243	47	314	1990
12/03	399	85	0.243	0.243	85	399	2501
13/03	559	160	0.243	0.243	160	559	3002
14/03	689	130	0.243	0.243	130	689	3184
15/03	886	197	0.243	0.243	197	886	3580
16/03	1058	172	0.243	0.243	172	1058	4051
17/03	1243	185	0.243	0.243	185	1243	5177
18/03	1486	243	0.243	0.243	243	1486	6041
19/03	1795	309	0.223	0.110	309	1795	7648
20/03	2257	462	0.118	0.110	462	2257	9041
21/03	2815	558	0.091	0.110	558	2815	9574
22/03	3401	586	0.119	0.110	586	3401	10069
23/03	3743	342	0.131	0.110	342	3743	10590
24/03	4269	526	0.123	0.110	526	4269	11413
25/03	4937	668	0.133	0.110	668	4937	12142
26/03	6235	1298	0.093	0.110	1298	6235	12810
27/03	7284	1049	0.083	0.110	1049	7284	12921
28/03	9134	1850	0.102	0.110	1850	9134	13294
29/03	10836	1702	0.093	0.110	1702	10836	12906
30/03	11899	1063	0.126	0.110	995	11831	12624
31/03	12775	876	0.129	0.110	1053	12884	13018

In accordance to our terminology, the recession phase of epidemics has started from April 11. After this moment, the dates with a dangerously high infection rate are underlined. This is done, for example, for April 14, which corresponds to Catholic Easter. Note that its celebration is the most probable reason for a significant disturbance in the number of detected cases for the period April 13-15.

Table 2: Statistics and forecasts for April 2020

Date	$T(\cdot)$	$C(\cdot)$	$\gamma(\cdot)$	$\gamma_p(\cdot)$	$C_p(\cdot)$	$T_p(\cdot)$	$H_p(\cdot)$
1/04	13964	1189	0.097	0.110	1108	13991	13397
2/04	15348	1384	0.116	0.110	1165	15156	13763
3/04	16770	1422	0.066	0.095	1255	16411	14112
4/04	18431	1661	0.038	0.095	1336	17747	14197
5/04	19691	1260	0.193	0.095	1409	19156	14210
6/04	20814	1123	0.089	0.095	1421	20578	14151
7/04	22194	1380	0.095	0.095	1462	22040	14074
8/04	23403	1209	0.075	0.095	1420	23460	13948
9/04	24983	1580	0.095	0.095	1389	24848	13854
10/04	26667	1684	0.110	0.095	1432	26280	13781
11/04	28018	1351	0.073	0.070	1474	27754	13658
12/04	29647	1629	0.072	0.070	1514	29268	13141
13/04	30589	942	0.074	0.070	1341	30608	12547
14/04	31119	530	0.123	0.123	1349	31957	12084
15/04	33573	2454	0.078	0.070	1350	33307	12222
16/04	34809	1236	0.069	0.070	1344	34651	11728
17/04	36138	1329	0.049	0.070	1337	35988	11204
18/04	37183	1045	0.061	0.060	1325	37313	10652
19/04	38496	1313	0.052	0.060	1316	38629	9966
20/04	39983	1487	0.070	0.060	1309	39938	9248
21/04	40956	973	0.060	0.060	956	40895	8493
22/04	41889	933	0.060	0.060	920	41814	8047
23/04	42797	908	0.051	0.060	878	42693	7610
24/04	44293	1496	0.051	0.060	1486	44179	7188
25/04	45325	1032	0.041	0.060	856	45035	6133
26/04	46134	809	0.052	0.060	821	45856	5645
27/04	46687	553	0.138	0.100	784	46640	5163
28/04	47334	647	0.125	0.100	639	47279	4895
29/04	47859	525	0.125	0.100	598	47877	4746
30/04	48519	660	0.102	0.100	555	48432	4622

In Table 3, the estimates for $\gamma(\cdot)$ of the last days of the month are computed by the statistics of June 1-10, which is not included in this table. In our estimates, there is a clear indication of the danger of the second wave of infection, or of a new stagnation stage, at least. We will see the answer in the nearest future.

Table 3: Statistics and forecasts for May 2020

Date	$T(\cdot)$	$C(\cdot)$	$\gamma(\cdot)$	$\gamma_p(\cdot)$	$C_p(\cdot)$	$T_p(\cdot)$	$H_p(\cdot)$
1/05	49032	513	0.081	0.079	510	48941	4530
2/05	49517	485	0.075	0.079	483	49424	4378
3/05	49906	389	0.047	0.079	457	49881	4241
4/05	50267	361	0.075	0.079	431	50312	4119
5/05	50509	242	0.089	0.079	368	50680	4014
6/05	50781	272	0.083	0.079	339	51019	3963
7/05	51420	639	0.069	0.079	516	51535	3937
8/05	52011	591	0.072	0.079	490	52025	3732
9/05	52596	585	0.065	0.079	475	52499	3537
10/05	53081	485	0.060	0.079	462	52961	3342
11/05	53449	368	0.087	0.079	358	53319	3143
12/05	53779	330	0.099	0.079	346	53665	3034
13/05	53981	202	0.109	0.079	335	54000	2928
14/05	54288	307	0.100	0.079	325	54326	2824
15/05	54644	356	0.089	0.079	317	54643	2722
16/05	54989	345	0.042	0.079	313	54956	2620
17/05	55280	291	0.056	0.079	311	55267	2514
18/05	55559	279	0.111	0.079	295	55562	2401
19/05	55791	232	0.093	0.079	279	55841	2296
20/05	55983	192	0.055	0.079	264	56105	2198
21/05	56235	252	0.089	0.079	248	56353	2108
22/05	56511	276	0.063	0.055	240	56593	2026
23/05	56810	299	0.049	0.055	231	56824	1898
24/05	57092	282	0.039	0.055	223	57047	1771
25/05	57342	250	0.051	0.055	215	57262	1645
26/05	57455	113	**0.098**	0.095	207	57469	1520
27/05	57592	137	**0.114**	0.095	199	57668	1458
28/05	57849	257	**0.104**	0.095	190	57858	1398
29/05	58061	212	**0.089**	0.095	181	58039	1341
30/05	58186	125	**0.069**	0.095	174	58213	1287
31/05	58381	195	**0.106**	0.095	167	58379	1236