

## MATHEMATICAL MODELING OF THE PROPAGATION OF COVID-19 PANDEMIC WAVES IN THE WORLD

E. S. Kurkina<sup>1,2</sup> and E. M. Koltsova<sup>2</sup>

UDC 51–7

We develop a mathematical model of the coronavirus propagation in different countries (Brazil, India, US, Japan, Israel, Spain, Sweden), in the city of Moscow, and across the world. The pandemic spreads by a highly complex dynamics because it occurs in open nonhomogeneous systems where new infection foci erupt from time to time, triggering new transmission chains from infected to susceptible people. In general, statistical data collected as cumulative and epidemic curves are a superposition of many distinct local pandemic waves. In our modeling, we use the system of Feigenbaum's discrete logistic equations (a logistic map) that describes the variation of the total number of infected over time. We show that this is the optimal model for the description of pandemic propagation in open nonhomogeneous systems with large errors in statistical data. We develop a procedure for isolating local waves, determining their model parameters, and predicting further evolution of each wave. We show that this model provides a good description of the statistical data and makes realistic forecasts. The forecast horizon depends on the degree of system closure and homogeneity. We calculate the start and end times of each wave, the peak, and the total number of infected in the current wave.

**Keywords:** mathematical modelling, COVID-19 pandemic, open systems, discrete logistic equation, superposition of epidemiological waves, forecasting.

### Introduction

The coronavirus pandemic has demonstrated that mankind is not prepared for the serious challenges associated with the spread of epidemics. Neither the economy, nor the health-care system, not administrative government organs in both rich and poor countries could control the spread of a moderately dangerous disease. Whole sectors of the economy have been paralyzed, international trade and manufacturing relations have collapsed. This is the deepest recession in a decade, and it will become worse because the pandemic is not over yet. By the end of February 2021, we have more than 110 million infected and about 2.5 dead from the coronavirus. Indirect losses are much greater and they remain to be assessed.

To successfully resist the spread of infection, we have to analyze the dynamics of morbidity, calculate the load it places on the health-care system, and make realistic forecasts. This requires the development of an adequate mathematical model of pandemic propagation with parameters that can be easily chosen from statistical data and easily adjusted when the propagation conditions change.

The mathematical modeling of human epidemics has a long history. It studies the mathematical properties and possibilities of models, and how to model and control the evolution of specific epidemics [1–11]. Mathematical modeling of the COVID-19 pandemic in various countries and regions began with the first outbreak of the disease in China and continues to be vigorously pursued to this day. Various models are used for this

---

<sup>1</sup> Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia; e-mail: elena.kurkina@cs.msu.su.

<sup>2</sup> Mendeleev Russian Chemical-Engineering University, Moscow, Russia; e-mail: kolts@muctr.ru.

purpose: Verhulst's logistic growth models and their modifications [12–15], compartmental models such as SIR [10, 16, 17, 18], SEIR, SIRS, SAIR [19–22], and more complex models [23–30] that allow for various refinements [17, 19, 20, 21, 25, 27]. All these models assume a homogeneous isolated system with adequate population mixing where the law of mass action applies. This is more or less true for large cities, where no health restrictions are imposed. However, compartmental models with constant parameters are inapplicable for the spread of the COVID-19 pandemic. They describe the propagation of the infection as a single epidemiological wave with a single peak.

In real life, systems are open and nonhomogeneous displaying highly complex dynamics of infection transmission. Smoothing the graphs of daily growth of the number of infected, we often obtain curves with several local maxima (peaks) or “shoulders” and “plateaus” – configurations that do not fit the classical single-peak model. This behavior can be explained by the appearance of several time-shifted local infection waves. Local waves may arise at different times in different parts of the country, or in one region but at different times if the infection is imported triggering a new outbreak, or some residents in a particular region begin to ignore quarantine restrictions. In general, the spread of an infection is a superposition of a very large number of local epidemic waves.

None of the constant-parameter mathematical models is capable of explaining this complex dynamics. It requires models with variable parameters.

Many efforts have been applied to adapt the SIR models to the spread of coronavirus [18, 21, 24] and to determine the model parameters. However, there are almost no studies that describe the spread of infection as a superposition of waves. Only one wave is typically considered. An attempt to describe two waves in Iran can be found in [31]. Ke Wu et al. [15], modeled the spread of the pandemic in a number of countries as a superposition of multiple waves. However, the wave parameters in the model were finally chosen in the decline phase near the end of the current wave. The model is thus appropriate for description but not prediction.

Other models considered for coronavirus propagation include network models [32], delayed models [33], models with fractal derivatives [34, 35], and more. All these models are largely theoretical. They reveal details of the mechanisms of infection transmission and enhance our understanding of the effect of various factors on the pandemic dynamics. However, none will work to describe and forecast the complex real-life dynamics of pandemic propagation.

Furthermore, although in this digital age we have access to an unprecedented database of disease transmission statistics for all the countries of the world, many studies have shown that the data are inaccurate and various authors estimate the error from 20% to 80%. Data inaccuracy is attributable to testing errors, insufficient coverage of tests in some countries, and the reluctance of people to see a physician if the disease symptoms are mild or moderate.

The determination of model parameters using inaccurate statistical data is an ill-posed inverse problem, which does not have a unique solution. The correct solution in these problems is isolated by regularization. The accuracy of the model parameters is determined by the accuracy of the statistical data. Since the error is large, it does not make sense to use complex multiparameter models, as the determination of the parameters requires much additional information. The difficulty with describing the complex evolution dynamics of a pandemic is neither in the complexity of the model nor in the large errors of the statistical data. The main difficulty is the openness and nonhomogeneity of the modeled systems (countries, cities, regions). Under these conditions, we need a simple mathematical model with a minimum number of parameters that are easy to determine and tune to changing conditions, a model that easily identifies new local waves and improves forecasts.

A suitable model is provided by the Feigenbaum discrete logistic equation, which contains only two parameters: the capacity of the system (associated with the potential number of people that may become infected) and the growth rate of the number of infected. Both affect the total number of infected people at the end.

We have developed a procedure for the identification of local waves and the determination of wave parameters in the model. Local waves are introduced into the model sequentially. A new wave is added when the calculated results begin to deviate from the actual data.

We have used the discrete logistic equation to model the spread of the COVID-19 pandemic in a number of countries in different regions of the world from the very beginning of the pandemic. We present the modeling results for the following countries: India, Brazil, Sweden, Spain, USA, Japan, Israel, the city of Moscow, and the entire world. The first notable conclusion is that the infection propagation dynamics and the model parameters differ strongly across countries. In some countries, the spread of infection was described for a long time by a single wave with a single peak (for instance, China, India, Brazil). In other countries, the mathematical description of epidemic evolution dynamics necessitated the introduction of several waves from the start. Thus, the spread of the pandemic in a single year in Japan and Spain required six waves, in Sweden nine waves, in USA and Moscow eleven local waves. Once each wave has passed its peak, a final forecast for each successive wave is made; we calculate when the wave will end and how many infected it will produce. Before the peak has been reached, it is impossible to determine the wave capacity exactly, and therefore different pre-peak wave evolution scenarios are constructed for different assumed capacities, and the timing of the peak and its parameters are predicted.

We show that the discrete logistic equation is the optimal choice for the modeling of infection spread in nonhomogeneous open systems with errors in the statistical data. The model has a minimum number of parameters, which are relatively simple to determine from statistical data. The model adequately describes the actual data and produces realistic forecasts. The forecast horizon is large if the propagation conditions do not change and the borders are closed, e.g., as for the first wave in China and Vietnam. In most countries and regions, the closure of the borders was not strict, quarantine measures were loosely enforced, and as a result new infection foci continued to erupt and the propagation conditions kept changing. Under these conditions, the model produced good 7-14-day forecasts (with error not exceeding 1%).

We have shown that constant monitoring is required to keep the pandemic under control. The model parameters should be frequently revised and forecasts should be made.

## 1. Mathematical Model of Epidemic Propagation

**1.1. Propagation of a Single-Wave Epidemic.** To describe and forecast the complex propagation dynamics of COVID-19 epidemics, we use the discrete logistic equation originally proposed by May in 1976 [36] and popularized by Feigenbaum [37], Sharkovskiy, and other researchers. It has the form

$$y_{n+1} = y_n \left( 1 - \frac{y_n}{N} \right)^\alpha, \quad y_1 = y_0 \quad \text{is known,} \quad (1)$$

where  $y_n$  is the total number of cases (the cumulative frequency of the population) in day  $n$ ;  $y_1$  is the initial number of cases in day 1,  $t_1 = t_0$ , the starting point for the time scale;  $\alpha$  is a parameter characterizing the population growth rate;  $N$  is the capacity, or the normalization factor. The solution of Eq. (1) is a numerical sequence  $\{y_n\}$ .

This equation is widely used in the modeling of various processes in biological, physical, economic, and other systems, but it has been hardly used to model epidemic propagation. It is known as a simple equation that demonstrates complex chaotic dynamics.

The system behavior may markedly change depending on the parameter  $\alpha$ . For  $0 < \alpha \leq 1$ , irrespective of the initial value  $y_0$ , the infected population goes to zero. Oscillating dynamics and chaos are produced by this

equation for  $\alpha > 3$ . In the range of parameter values  $1 < \alpha < 2$ , the equation displays logistic growth – the growth of a population in an environment with limited resources. It is this range that we use to describe the spread of an epidemic. The discrete logistic equation describes all phases of epidemic evolution: the exponential growth stage in the early stage of propagation when  $y_n \ll N$ ; then the decelerating growth of the number of new cases; the peak with the greatest number of cases; and the terminal stage when the epidemic declines. Toward the end of the epidemic, the total number of cases asymptotically goes to the steady-state value

$$\bar{y} = \left( \frac{\alpha - 1}{\alpha} \right) \cdot N, \quad y_n \rightarrow \bar{y}, \quad \text{as } n \rightarrow \infty. \quad (2)$$

Relationship (2) holds if  $\alpha$  remains constant during the entire course of the epidemic. If  $\alpha$  is variable, then the last observed value of the parameter should be substituted in (2).

We see from expression (2) that the terminal number of cases depends not only on the parameter  $\alpha$  but also on the system capacity  $N$ . The higher the value of  $\alpha$ , the greater the terminal number of cases. With  $\alpha = 2$ , this number reaches its maximum when it is equal to one-half the capacity  $\bar{y} = N/2$ . With  $\alpha$  close to 1 (i.e., slow spread of infection), the terminal number of cases is much smaller (by orders of magnitude) than the potential number of susceptible individuals  $N$ .

The value of the parameters  $N$  and  $\alpha$  depends on a whole range of factors, such as the openness of the country to viral infection (inflow of people from zones with epidemic foci), clustering and density of the population, the existence of megalopolises, resistance to infection, discipline of the population in the face of quarantine measures, etc. The quarantine measures imposed by the authorities reduce the infection probability, reducing the parameter  $\alpha$  and with it the total number of cases  $\bar{y}$ .

**1.2. The Discrete Equation for Increments.** The discrete logistic equation (1) directly describes the variation over time of the total number of infected  $y_n = y(n)$ , i.e., it specifies the *cumulative curve*. We introduce the equation for the daily number of new cases, i.e., the daily increments. The equation then describes the *epidemic curve*:

$$y_{n+1} = y_n + \Delta_{n+1}, \quad \Delta_{n+1} \text{ is the increase in the number of cases in day } n + 1;$$

$$y_n = y_{n-1} + \Delta_n, \quad \Delta_n \text{ is the increase in the number of cases in day } n, \text{ and so on.}$$

In other words, the cumulative function  $y_n$  equals the sum of all daily increments up to day  $n$ .

Consider the equation for increments:

$$\Delta_{n+1} = y_{n+1} - y_n = \alpha y_n \left( 1 - \frac{y_n}{N} \right) - y_n = b y_n \left( 1 - \frac{y_n}{\bar{y}} \right), \quad (3)$$

where  $b = (\alpha - 1) < 1$ ,  $\bar{y} = \left( \frac{\alpha - 1}{\alpha} \right) N$  is the stationary value (2).

The equation for increments (3) is the same discrete logistic equation as (1), but with different parameters. Since  $y_n \rightarrow \bar{y}$  as  $n$ , we obtain from Eq. (3)

$$\Delta_n \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

From equation (3) it follows that when the number of cases is relatively small, it increases from day to day until the maximum is reached; after that, the number of cases starts to decrease. The maximum value  $\Delta_{\max}$  is obtained by differentiating (3). It is attained for

$$y^* = \frac{\bar{y}}{2} = \left( \frac{\alpha - 1}{2\alpha} \right) N, \quad \Delta_{\max}(y^*) = \frac{(\alpha - 1)^2}{4\alpha} N. \quad (4)$$

The epidemic ends once the increments decrease to zero. The daily increments, however, do not necessarily drop to zero, as in the first wave of the coronavirus pandemic, but stick for a fairly long time at some low level. This implies that the epidemic has reached the endemic stage. Then a new outbreak may occur after some time.

Equation (3) may be interpreted differently. At the beginning of the epidemic, the number of daily new cases is proportional to the number of cases, and the new cases grow at a rate  $b$ . However, the number of recovered patients increases over time and a herd immunity develops reducing the propagation rate:

$$b(y_n) = b \left( 1 - \frac{y_n}{\bar{y}} \right). \quad (5)$$

From this expression it follows that the first term makes the main contribution to the rate of propagation in the early phase of the epidemic, while the second term kicks in toward the end of the epidemic.

**1.3. Comparison of the Discrete Logistic Equation with the Verhulst Equation and SIR.** Epidemics follow the logistic growth law. Many models describe logistic growth. The most common models for the description and forecasting of the propagation of the COVID-19 pandemic are differential equation models, such as the Verhulst equation and its various modifications [12–15] and SIR with its various extensions [16–30]. Let us compare the discrete logistic equation (1) with the Verhulst equation and the classical SIR model.

Note from the outset that the discrete logistic equation (1) is *not a difference analog* of the Verhulst differential logistic equation

$$\frac{dx}{dt} = \lambda x \left( 1 - \frac{x}{K} \right), \quad \lambda > 0, \quad x(t_0) = x_0. \quad (6)$$

The parameters of these equations have different interpretations. In the differential equation (6),  $K$  is the cumulative number of people who will certainly become infected by the end of the epidemic irrespective of the propagation rate  $\lambda$  and the initial value  $x_0$ . The coefficient  $\lambda$  does not influence the system dynamics qualitatively. In discrete equation (1), on the other hand,  $N$  is the capacity of the system associated with the potential number of people that may become infected. In *open* systems, the capacity  $N$  may exceed the population of the city, the country, or the region. The final number of cases depends on the coefficient  $\alpha$ . If quarantine restrictions are imposed, then  $\alpha$  diminishes and the total number of potential cases goes down.

Unlike Eq. (6), which has to be integrated to find the solution describing the logistic growth, Eq. (1) directly describes the logistic function by a numerical sequence. With appropriately chosen parameters, Eqs. (1) and (6) *produce close results*. The solution of Eq. (6) is the function

$$x(t) = \frac{x_0 K}{x_0 + (K - x_0) e^{-\lambda(t-t_0)}}. \quad (7)$$

In the exponential growth stage, the solutions of these equations are *identical*. Indeed, with  $x \ll K$ , the solution

of Eq. (6) is the exponential function

$$x(t) = x_0 e^{\lambda(t-t_0)}. \quad (8)$$

Equation (1) describes the exponential function (8) as the geometrical progression

$$y_n = \alpha y_{n-1}. \quad (9)$$

The parameters of Eqs. (1) and (6) are related by the expression

$$\alpha = e^{\lambda dt}, \quad dt = t_n - t_{n-1}. \quad (10)$$

The process of passing the peak and attaining the steady state is described somewhat differently for Eqs. (1) and (6) but on the whole the results are close (see below).

Let us now compare the discrete model (1) with the classical SIR model proposed by Kermack and McKendrick in 1927 [10]. SIR considers three group of individuals: susceptible to disease, infected, and recovered (or removed). Transmission is from infected to susceptible. Recovered individuals acquire immunity and cannot be re-infected. The model is described by a system of three differential equations:

$$\begin{aligned} \frac{dS}{dt} &= -rSI, \\ \frac{dI}{dt} &= rSI - \nu I, \\ \frac{dR}{dt} &= \nu I, \end{aligned} \quad (11)$$

$$S(t_0) = S_0, \quad I(t_0) = I_0, \quad R(t_0) = R_0.$$

Here  $S(t)$ ,  $I(t)$ ,  $R(t)$  is the number of susceptible, infected, and recovered individuals at time  $t$  respectively;  $r$  is the rate of infection transmission;  $\nu$  is the recovery rate of infected individuals.

We see from the system of equations that the size of the total population  $M$  is conserved:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0.$$

$$M(t) = S(t) + I(t) + R(t) = M = \text{const}.$$

The total number of infected in SIR is given by

$$Z(t) = M - S(t) = I(t) + R(t),$$

SIR can be used to model the propagation of diseases in homogeneous or weakly structured communities with a large well-mixed population. These conditions are adequately observed in large cities.

SIR has a stable steady-state solution for  $r > v$  that describes the end of the epidemic. As  $t \rightarrow \infty$ , all the phase variables go asymptotically to their steady-state values. The number of infected goes to zero. In SIR, the parameter  $M$  is often identified with the initial number of susceptible:  $M = S_0$ . The terminal number of infected is affected by the rates  $r$  and  $v$ , as in the discrete equation (1). The physical interpretation of the parameter  $M$  is almost the same as the parameter  $K$  in the Verhulst equation (6), but it is not the capacity  $N$  in the discrete logistic equation: it is the actual number of people that should get infected. The capacity  $N$  has a deeper physical meaning. It characterizes the size of the population in which the infection spreads, and even with the maximum possible propagation rate  $\alpha = 2$  the number of infected attains only one-half of the capacity. With the COVID-19 pandemic, the number of infected is much less than the capacity  $N$ .

There is no analytical formula in SIR for the steady-state solution. The steady-state solution  $\bar{S}$  for the number of susceptible that remain uninfected until the end is obtained from the transcendental equation [16]

$$\frac{r}{v} \left( (I_0 + S_0) - \bar{S} \right) = -\ln \left( \frac{\bar{S}}{S_0} \right). \tag{12}$$

Solving equation (12) and evaluating  $\bar{S}$  we find the total number of recovered  $\bar{R}$  and infected  $\bar{Z}$ :

$$\bar{R} = \bar{Z} = M - \bar{S}. \tag{13}$$

Let us compare the dynamics of the models (1), (6), and (11) for a concrete example. All three models initially describe exponential growth, then a peak, and in the final stages reach a steady-state total number of infected. We choose the model parameters so that (1) they equally describe exponential growth at the onset of the epidemic; (2) they attain the same stationary solution at the end of the epidemic. Assume that the SIR parameters are given. Then to satisfy desiderata (1) and (2), the differential logistic equation (6) should have the parameters

$$\lambda = r - v, \quad K = M - \bar{S}. \tag{14}$$

The corresponding parameters of the discrete logistic equation are obtained from (15) and (2):

$$\alpha = e^{\lambda dt}, \quad dt = 1 \text{ (one day)}, \quad N = (M - \bar{S}) \frac{\alpha}{\alpha - 1}.$$

**Example.** Consider SIR with some parameters that are characteristic for COVID-19:

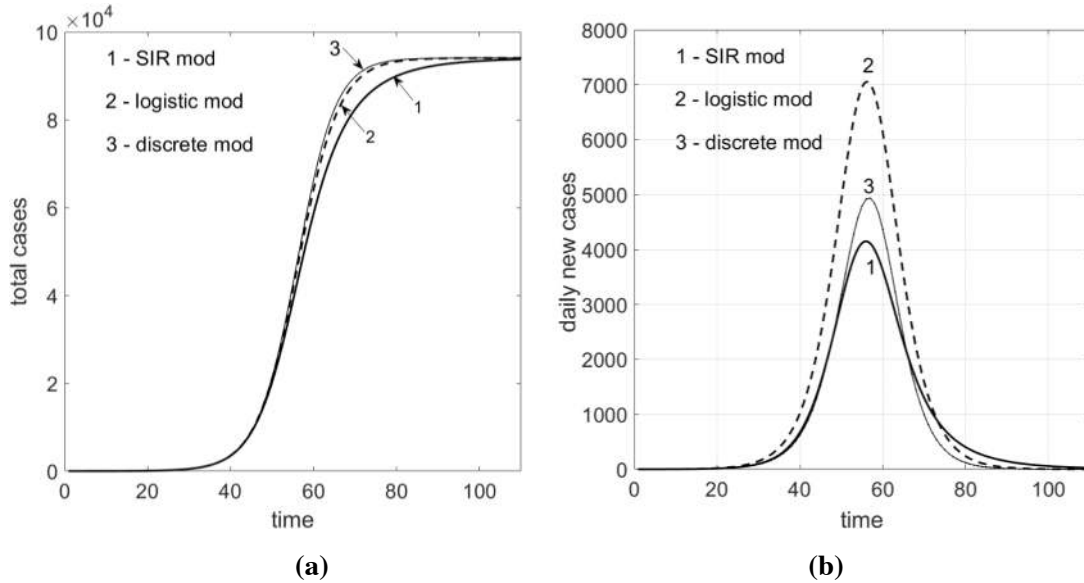
$$r = 0.3, \quad v = 0.1, \quad I_0 = 1, \quad S_0 = 10^5 - 1, \quad M = 10^5,$$

$$\lambda = r - v = 0.2, \quad \alpha = e^{b/2} = 1.1052, \quad N = 988290.$$

From the transcendental equation (12), we find the steady-state solution  $\bar{S}$ :

$$\bar{S} = 5951.9.$$

Then  $K = 94068$ .



**Fig. 1.** Comparative description of logistic growth in three models: SIR, Verhulst, and discrete equation (1); (a) cumulative curves, (b) epidemic curves.

The initial values are chosen so that the model solutions match on the exponential growth stage. Figures 1a and 1b show the cumulative and epidemic curves described by the three models. We see that the three curves coincide in the exponential growth stage and slightly diverge thereafter. All three models attain the same steady-state solution (this is how the parameters were chosen). The Verhulst model displays the fastest growth and the highest peak. The SIR model has the slowest decay, because allowing for the recovered individuals reduces the propagation rate. The discrete logistic model is closer to the SIR model for the given parameter values.

**1.4. Determination of the Parameters of the Discrete Logistic Equation from Statistical Data.** As noted in the introduction, actual statistical data display a highly complex dynamic behavior. Smoothed endemic curves have several local maxima and a plateau. We may infer that these are a superposition of several local waves. The reason for the complex dynamics is clear. The point is that countries, cities, and regions are open nonhomogeneous systems in which new epidemic foci erupt randomly from time to time, generating new infection transmission chains from infected to susceptible individuals. In nonhomogeneous systems, outbreaks occur at different times in different locations and spread at different rates. All this gives rise to time-shifted local epidemic waves. As a result, an outbreak in a particular region is the sum of multiple local waves. Some of them merge into a common wave. Other strong local waves are time-shifted to such an extent that they do not merge into a common wave and actually produce a separate outbreak with its local peak even as the common wave decays.

Furthermore, the statistical epidemic-propagation data are prone to large errors, and the determination of the model parameters from data with errors is an ill-posed problem with a nonunique solution. This is an intrinsic difficulty, because a numerical solution cannot be more accurate than the statistical data used, in which the error reaches tens of percent. For this reason, we have chosen the simplest discrete logistic-growth model with variable parameters that can describe any dynamics. We find one of the possible solutions adequately describing the statistical data. The local waves can be separated out in different ways, and there is always a certain leeway in the definition of the parameters of model (1): the coefficient  $\alpha$  and the capacity  $N$ .



Let us first consider the calibration of a single-wave model.

Reduce model (1) to a nondimensional form by the change of variables

$$y_n = x_n N, \quad 0 < x_n \leq 1, \quad \alpha = \alpha. \tag{15}$$

Then model (1) takes the form

$$x_{n+1} = \alpha x_n (1 - x_n), \quad x_0 = \frac{y_0}{N}. \tag{16}$$

The initial condition  $y_0$  is typically set equal to 1 ( $y_0 = 1$ ).

An important feature of the equation is also the *choice of the discrete time increment* on which the exposed population is calculated. We took a 12-hour interval relying on Michael Levitt’s work for China [38]. We recalculate the number of cases every 12 hours and compare with the actual data after 24 hours, i.e., once daily. The chosen interval is the time during which one person on average may infect another. Calculations carried out for different countries during the entire past year confirm that this interval is optimal.

The parameter  $\alpha$  is determined at the start of the wave during the exponential growth stage from statistical data:

$$\frac{X_{s,n+1}}{X_{s,n}} \equiv \alpha_s, \tag{17}$$

where  $X_{s,n}$  is the observed number of cases in day  $n$ ;  $\alpha_s$  is the average of the ratio (17) over several days. Since the number of cases  $x_n$  is updated every 12 hours, the coefficient  $\alpha$  is given by

$$\frac{x_{n+1}}{x_n} = \alpha = \sqrt{\alpha_s}, \quad x_{n+1} = \alpha x_n, \tag{18}$$

where the factor  $\alpha$  may decrease over time. This behavior is monitored by analyzing the statistical data and linked with information about the introduction of restrictions. When modeling the spread of epidemic in different countries, the coefficient  $\alpha$  was kept unchanged for different waves or else changed in several jumps (usually 2–4) to achieve the best description of the actual data. The last change was usually made before the mid-peak point, and remained constant thereafter. The last value of  $\alpha$  was used for forecasting.

The parameter  $N$  was determined at the peak from (4)

$$N \equiv X_s^* \left( \frac{2\alpha}{\alpha - 1} \right), \quad \text{or} \quad N \equiv \Delta_s^* \frac{4\alpha}{(\alpha - 1)^2}, \tag{19}$$

where  $X_s^*$  and  $\Delta_s^*$  are the average number of cases and the average daily increases near the peak, respectively.

In some instances, the peak spreads into a fairly long plateau. Then  $N$  has to be changed, refined. Before the peak, starting with the exponential growth stage, we define various hypothetical values of the capacity  $N$ . They are borrowed from various estimates, from past experience, and from comparative analysis. For different  $N$  we construct different epidemic-evolution scenarios. For each scenario, we calculate when the peak will occur, what will be the increment in the peak, how many cases will be reported each day, when the wave will end,

and how many cases it will produce in total. If the coefficient  $\alpha$  decreases (restrictions are introduced), the scenario is again updated. In the exponential growth stage, the cumulative curves coincide for different  $N$ . The curves split closer to the peak. At the peak, we choose the final  $N$  from (19) and use the given wave to construct the final forecast for each of the following days until the wave's final decay.

**1.5. A Model of the Propagation of Several Epidemic Waves.** We treat the country, the city, the region, etc., where infection spreads as an *open system*. Open borders and moderate restrictions result in new infection foci appearing from time to time; these foci produce new infection-transmission chains thus increasing the potential number of people exposed to infection. New capacities of susceptible individuals and new local waves thus form. Different local waves form a superposition.

To calculate the spread of infection in the form of several waves, we use several discrete equations (1) (or (16)), each describing its own wave with its own capacity  $N^{(i)}$ , its own set of parameters  $\alpha^{(i)}$ , its starting time  $t_0^{(i)}$ , and its initial number of cases  $x_0^{(i)} = y_0^{(i)}/N^{(i)}$ :

$$x_{n+1}^{(i)} = \alpha^{(i)} x_n^{(i)} (1 - x_n^{(i)}), \quad x_0^{(i)} = \frac{y_0^{(i)}}{N^{(i)}}. \quad (20)$$

$$y_n^{(i)} = x_n^{(i)} N^{(i)}, \quad i = 1, \dots, M, \quad (21)$$

where  $M$  is the number of waves.

The total number of cases at time  $t_n$  is determined by the sum

$$Y_n^{(i)} = \sum_{i=1}^M x_n^{(i)} N^{(i)}. \quad (22)$$

The parameters in these equations are chosen to ensure the best description of the statistical data.

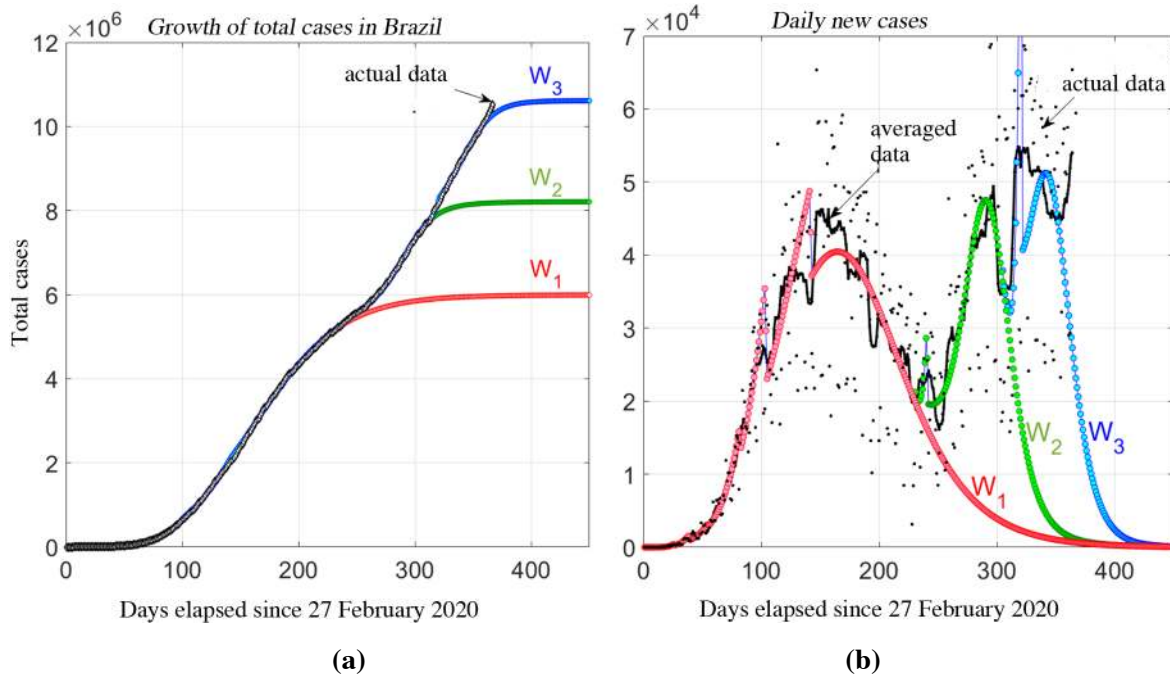
Let us briefly describe the algorithm that determines the parameters of the current local wave. Assume that we have already included  $j$  waves in the model describing the propagation of infection. The superposition of these waves up to some time  $t_n$  adequately described the observations  $X_{s,n}$ :

$$Y_n = \sum_{i=1}^j x_n^{(i)} N^{(i)} \approx X_{s,n}.$$

Then the combined wave  $j$  began to decay and the observations began to lead the calculated values. The differences between the observations and the calculated values give a sequence of  $Z_n$  for the determination of the next wave:

$$Z_n = X_{s,n} - Y_n, \quad n = n_1, n_1 + 1, n_1 + 2, \dots \quad (23)$$

The parameters in Eq. (16) for the description of wave  $j+1$  are chosen as for the first wave: first we choose the coefficient  $\alpha$  and take several hypothesized values of  $N$  describing various evolution scenarios for wave  $j+1$ . Then the capacity  $N$  is refined near the peak, and a final forecast is made for the subsequent evolution of the wave.



**Fig. 2.** The dynamics of the coronavirus pandemic in Brazil between 27 February 2020 and 27 February 2021. Actual data and modeling results:  $W_1$ ,  $W_2$ ,  $W_3$  are the modeled waves.

## 2. Modeling Results

The statistics of confirmed COVID-19 cases for our sample of countries (India, Brazil, Sweden, Spain, USA, Japan, Israel) and the world were taken from the Worldometer site [39]. The database with coronavirus cases for Moscow was taken from [40].

The spread of infection can be modeled by a single wave in two cases: 1) when weak restrictions virtually do not prevent free interpersonal contacts, 2) when very strict restrictions completely isolate the infected population. The first instance characterizes the pandemic propagation in India and Brazil until the end of October 2020. The second instance was observed in China and Vietnam. In most other countries, the mathematical models describe the spread of the epidemic by several waves. Local waves are apparent as local peaks and plateaus on the smoothed graphs of daily new cases. Since the statistics are not accurate and the inverse problem of wave reconstruction is ill-posed, it is impossible to determine the number of local waves and their characteristics. We accordingly approach the situation as a phenomenological problem and chose the local-wave parameters so as to ensure the best possible description of the COVID-19 data available at the time.

The modeling results for each sample country are shown in two graphs. Figures 2–10(a) present the growth of the total number of cases – the cumulative curves. Figure 2–10(b) are the daily number of new cases – the epidemic curves. The statistical data are marked by diamonds, calculated values by disks. The thick solid line in the (b) panels plots the seven-day moving average of daily new cases. We start the description of our coronavirus modeling results with Brazil.

**Brazil.** Brazil is the third in the world by the total number of cases – 10.5 million as of the end of February 2021. Statistics of daily new cases display strong volatility (Fig. 2b, points), stronger than in the other countries. This is associated with mass testing difficulties. The modeling results are presented in Figs. 2a and 2b.

The pandemic in Brazil flared up in the second half of March 2020, when the daily number of new cases reached hundreds of people. A local peak was observed in Brazil at the beginning of June 2020, and the epidemic began to decline thereafter (from 7 July to 19 July 2020). But another jump in new cases began on 20 July 2020. The absolute maximum of the daily number of new cases (70,869) occurred on 29 July 2020, after which the pandemic began to decline on average. We modeled the pandemic in Brazil starting in May 2020. The normalization factor was updated several times. The respective wave capacities are shown in Fig. 2a. The growth rate indicator was reduced several times as the pandemic spread. This can be attributed to various restrictions that lowered the probability of contact with infected individuals.

The model adequately described a local wave with the 7 July peak. But in August 2020, we again had to update the parameters. Until the end of October 2020, the pandemic in Brazil was declining and after the September 2020 update of the parameters ( $N = 450$  million,  $\alpha_7 = 1.0135$  from 16 July) the model adequately described the observations to the end of October 2020. From 29 July to 26 October 2020, the deviation in the total number of cases between the model and the actual statistics was less than 1% (except for three days, when the deviations exceeded 1%).

In November 2020, a new outbreak began in Brazil, and this required the introduction of a second wave into the model. Two scenarios were proposed. In the first scenario, the parameters of the second-wave peak were required to be of the same order as in the first wave. In the second scenario, the second outbreak was required to be stronger than the first, as in many European countries. In the realization of the first scenario, the peak occurred on 14 December 2020. Calculations show that the first wave in Brazil will end in June 2021 (when the daily number of new cases will have dropped to tens), after producing 6 million cases in total. The second wave will end in April 2021, adding 2.2 million cases.

The third local wave in Brazil began to form in December 2020. Its peak was observed at the beginning of February 2021, followed by a decline. Calculations suggest that the third wave will end in May 2021 with 2.4 million additional cases. A new fourth local wave began to rise at the end of February 2021, but it was too early to estimate its parameters.

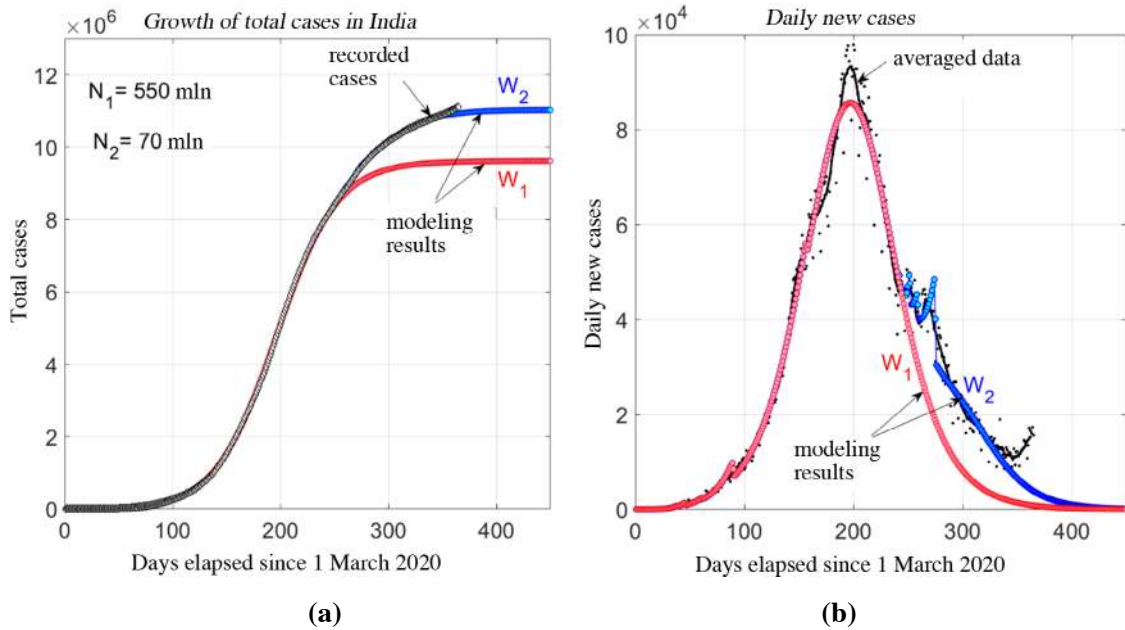
**India** – the second most populous country in the world ( $\approx 1.4$  billion people) with the population clustered in scattered settlements. Despite the efforts to introduce strict restrictions in the initial stage of infection spread, control was lost and the infection began to spread rapidly in the densely populated country. India overtook Brazil by the number of cases and moved up to the second place in the world after USA.

Figure 3 shows the daily record of the total number of cases (a) and the daily new cases (b), superimposing the modeling calculations on the statistical data. We see that up to the beginning of November 2020 the single-wave model adequately fits the observations. The peak of this single wave passed in mid-September 2020. The first-wave normalization factor updated in September 2020 was  $N = 550$  million. The growth rate went down as the pandemic spread, and the last (sixth) value of the growth rate indicator was  $\alpha_6 \cong 1.018$ , having remained steady since 3 August 2020. In November 2020, the growth slowed down and the model began to lag behind the observations, which necessitated the introduction of a second wave.

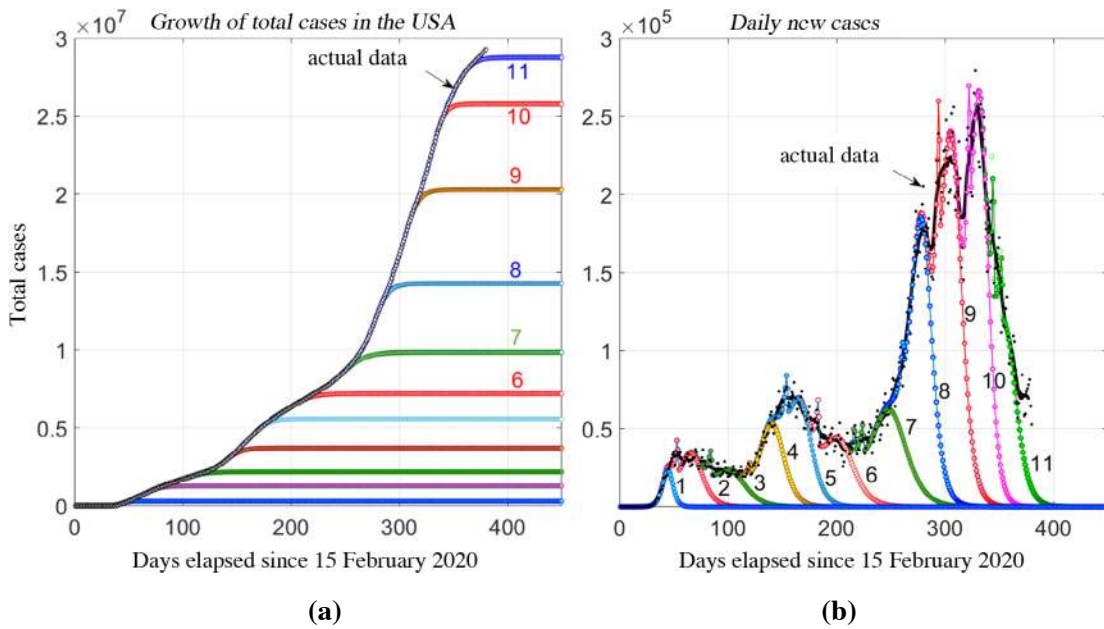
The second wave was much weaker than the first. Before long, it passed the peak in the end of November 2020 and the decline began thereafter. Forecasts adequately fit the observations until mid-February 2021, but in the second half of February 2021 a third wave began to rise. It is unclear if it is going to decay or evolve into a new powerful outbreak. More time is needed to model the third wave and forecast its evolution.

The model shows that the first wave in India will end in June 2021, having produced about 9.6 million cases. The second wave will end approximately at the same time adding 1.4 million cases.

**USA** is a huge country spanning diverse regions. Regarding the spread of pandemic, we clearly differentiate between large cities, such as New York, Los Angeles, Chicago, San Francisco, and others, with high population



**Fig. 3.** The dynamics of the coronavirus pandemic in India between 1 March 2020 and 28 January 2021. Actual data and modeling results:  $W_1, W_2$  are the two modeled waves.



**Fig. 4.** The dynamics of the coronavirus pandemic in the USA between 15 February 2020 and 1 March 2021. Actual data and modeling results. Curve labels identify the modeled waves.

and a crazy tempo of life. Statistical analysis reveals three large outbreaks in the USA (Figs. 4a and 4b): spring and summer of 2020, and then autumn-winter of 2020 which began to decline in February 2021. We see from the smoothed graphs of daily new cases (Fig. 4b) that these major waves have several local maxima on both the upward and the downward shoulders, suggesting that they are themselves superpositions of multiple local waves.

The first outbreak began a year ago, in February 2020. Modeling the spread of the pandemic in the USA from the very beginning, we successively added local waves to the model and made forecasts for each wave. Poorly organized restrictions, civil protests, open borders, etc., led to unpredictable changes in the situation. The forecasts based on the current wave worked well for a week or at most two. Then the model results began to lag behind the observations, and the next wave had to be added. During the elapsed 12 months, we introduced 11 local waves describing the spread of the pandemic in the USA, i.e., on average one local wave was added into the model each month. The model provides a marvelous fit of the observations, with deviations below 1%, except on some days. We stress again that the problem of determining the model parameters from data with errors produces a nonunique solution. Initially, we achieved a good fit of the spring outbreak with six waves, but in retrospect we combined them in pairs and reduced the total number to three waves (Fig. 4). The first spring outbreak ended only in August 2020, with an estimated 2,185 cases.

While in most European countries, the coronavirus morbidity went down significantly in June–July 2020, street protests in the USA initially triggered a second large wave. Its peak was reached by the end of July 2020. The height of the second-wave peak was more than double the peak of the first large wave. Modeling calculations show that the second wave ended only in November 2020, adding slightly more than 5 million cases. We modeled it by three local waves: the fourth, the fifth, and the sixth counting from the start of the pandemic (Figs. 4a, 4b). Even before the summer wave had subsided substantially, the next wave began to emerge in the USA, as in many European countries. Its appearance was apparently attributable to the fact that large parts of the population ignored the restrictions and the borders remained open. The autumn-winter outbreak was much stronger than the spring and summer outbreaks. It has already lasted for more than half a year and is just now beginning to subside. So far, however, the number of new cases remains as it was in the peak of the summer outbreak.

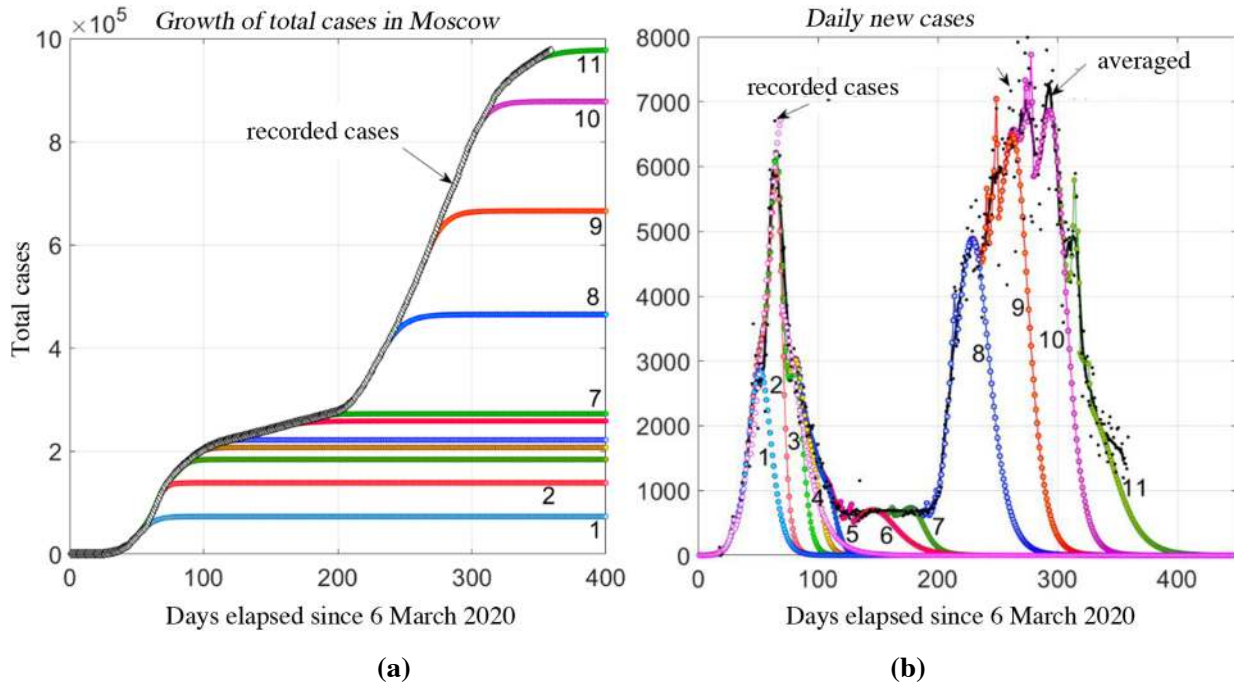
We see from the graph of daily new cases (Fig. 4b) that the autumn outbreak has three well-defined maxima. In other words, it began declining on three separate occasions, but then a new local wave pulled it up. The maximum daily increments were observed in January 2021. Five local waves were added to the model to describe this outbreak: one moderate wave on the ascending shoulder, three large waves in the peak, and one moderate was on the descending shoulder. We see from the graphs that the last (eleventh) local wave began to lag behind the observations, and a next local wave emerged. The model indicates that the eleventh wave ended in April 2021. The five waves (seventh to eleventh) describing the autumn-winter outbreak added a total of some 21.6 million cases.

**Moscow** is one of the largest megalopolises in the world, a center of international business and Russia's main transportation hub. The city is an open system: domestic travel to all parts of the country did not stop even under the most severe restrictions. As a result, new infection foci kept randomly appearing in Moscow at all times, triggering new local waves. The coronavirus situation in Moscow is continuously monitored since the end of March 2020. As of today, Moscow had contributed a touch less than one-quarter of the total number of cases in Russia, while its population of 12.7 million is approximately 1/11th of Russia's population.

Figures 5a and 5b present the statistical data and the modeling results for Moscow. Coronavirus spread in Moscow in two large waves: a spring outbreak and an autumn-winter outbreak, connected by a lower plateau. In other words, the spring outbreak in Moscow had not dropped to zero; it passed into an endemic stage that lasted for two months.

From the beginning of July 2020 to mid-September 2020, the average daily number of new cases was at the level of 600–700. The first outbreak was modeled with seven local waves: two waves on the ascending shoulder, three waves on the descending shoulder, and two wide waves to describe the endemic stage.

The first large wave (superposition of five waves) ended in July 2020 according to the modeling results. It produced 226,000 cases. Echoes of the endemic finally vanished in October 2020, having added 50,000 cases.



**Fig. 5.** The dynamics of the coronavirus pandemic in Moscow between 6 March 2020 and 27 February 2021. Actual data and modeling results. Curve labels identify the modeled waves.

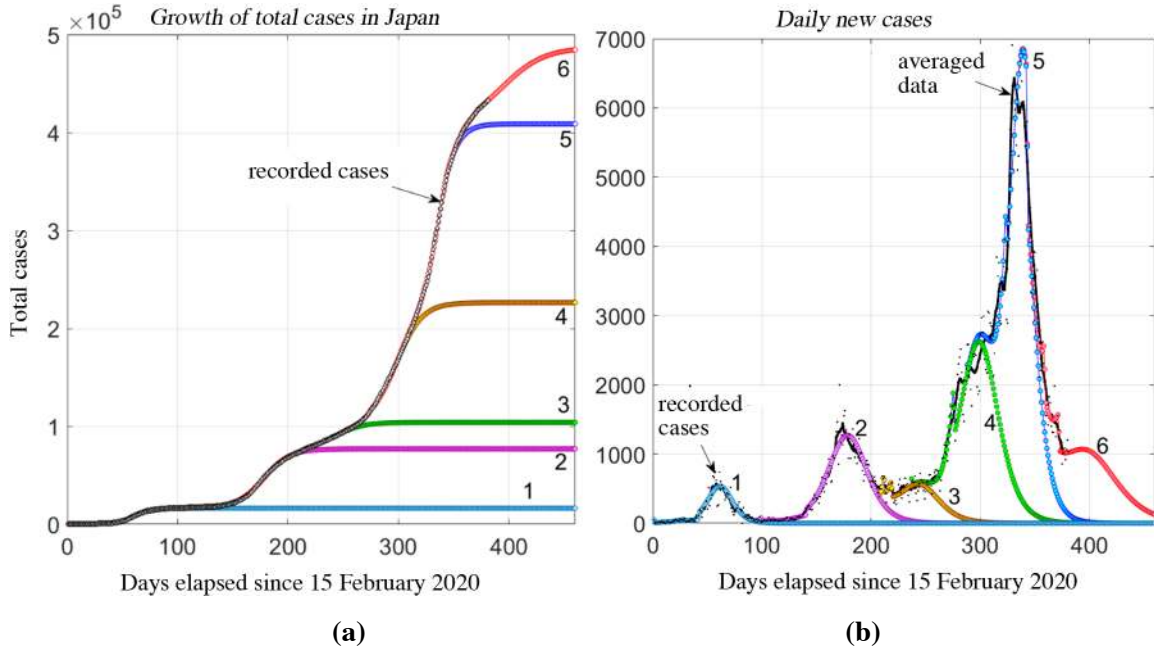
The growth rate indicator  $\alpha$  of the first local wave dropped to one-half, its decrease associated with the introduction of restrictions. The growth rates of the other five local waves after the introduction of restrictions were at the level of the first local wave.

The second large outbreak in Moscow began in mid-September 2020. In the second half of October 2020, the daily number of new cases exceeded 5,000. Restrictions began to be tightened every day in Moscow, which slowed down the propagation of the infection. As a result, for two weeks from 16 to 29 October 2020 the daily number of new cases remained unchanged, averaging 4,713. But then the daily increments started creeping up, and in the second half of November 2020 they exceeded 7,000 new cases daily.

The next local maximum emerged at the end of December 2020, and after that the pandemic declined with fluctuations. The autumn-spring outbreak in Moscow is adequately modeled by four local waves: one on the ascending shoulder (eighth local wave), one on the descending shoulder (eleventh local wave), and two waves (ninth and tenth) describing the two largest local maxima. Calculations show that this outbreak ended in April 2021, and the four waves (8 to 11) added 706,000 cases. The 11 waves combined added 982,000 cases. The eleventh wave is obviously not the last in the model. Already now it has begun to lag behind the actual data. On 20 February 2021, the modeled results deviated by a mere 0.02% from the observations, whereas after that the deviation began to increase monotonically, reaching 0.49% on 27 February 2021.

**Japan.** Three large outbreaks were observed in Japan during the 12 months of the coronavirus spread. These outbreaks are modeled as a superposition of six local waves. The modeled results and the actual observations are shown in Figs. 6a and 6b. The model adequately fits the observations. Japan is more closed and more homogeneous than USA, France, Spain, and other European countries. The first wave in Japan is adequately described by a single-wave model characteristic for a closed system. For a whole month after the peak, the observations closely tracked the forecast.





**Fig. 6.** The dynamics of the coronavirus pandemic in Japan between 15 February 2020 and 1 March 2021. Actual data and modeling results. Curve labels identify the modeled waves.

The capacity of the first wave in Japan was  $N = 270,000$ , and the last growth rate parameter was  $\alpha = 1.065$ . The theoretical steady-state total number of cases (2) is

$$\bar{y} = \left( \frac{\alpha - 1}{\alpha} \right) \cdot N = \left( \frac{1.065 - 1}{1.065} \right) \cdot 270\,000 = 16478 \text{ cases.}$$

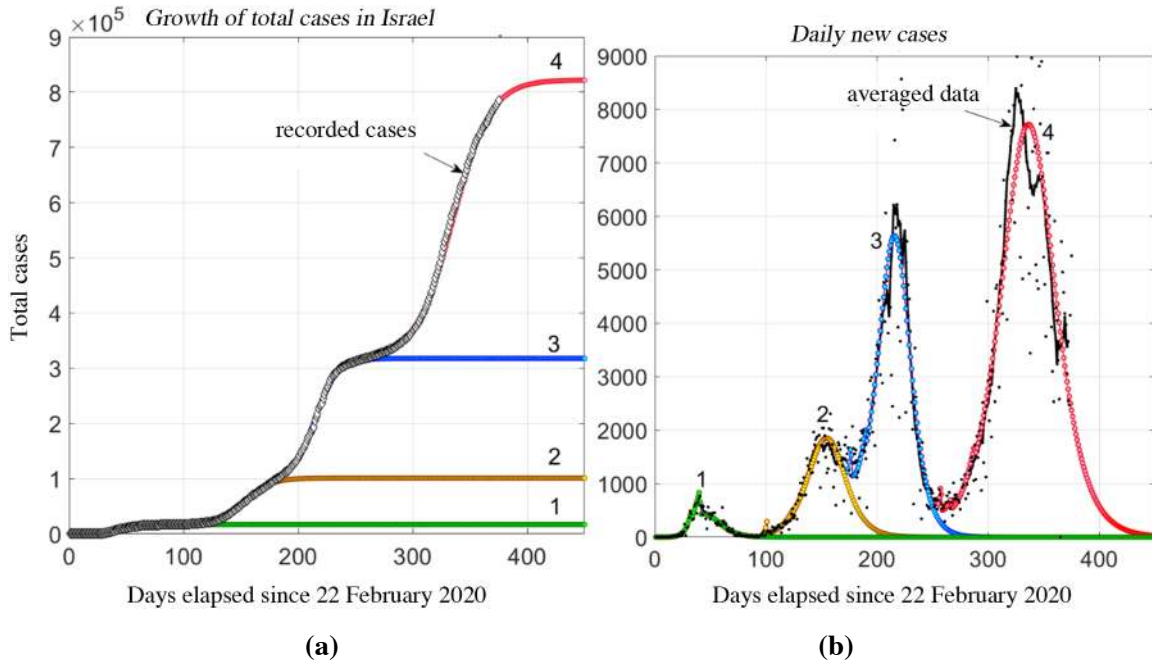
The model suggests that the first wave should end in the fourth week of May 2020, attaining a steady-state level of about 16,500 cases. The actual data for Japan at the end of the first wave (23 May 2020) revealed 16,536 cases, which provides a convincing confirmation of our model.

The second and third waves erupted, apparently, due to the relaxation of the restrictions. Modeling results suggest that the second wave began to emerge at the end of May 2020, producing significant daily increments (more than 100 new cases) in the first days of July 2020. It proved to be almost four times stronger than the first wave, adding about 60,500 cases. Its peak occurred on 10 August 2020. The second-wave normalization factor in the model was  $N = 1.5$  million, and the growth parameter  $\alpha$  was of the same order as in the first wave.

The autumn-winter outbreak started with the emergence of the third wave in the beginning of September 2020, before the dissipation of the second wave. The third local wave was relatively small. It passed the peak in mid-October 2020 and began to decline. Had strict restrictions been introduced, as in China, this wave could have decayed. But lax restrictions and open borders triggered a strong growth in the number of cases, as in many other countries. This growth is modeled by two local waves: the fourth and the fifth, which was the largest (see Fig. 6b). The fifth wave reached its peak in mid-January 2021, when it began to decrease abruptly, tracking the observations.

Toward 10 February 2021, the rate of decline had slowed down and the observations began to lead the model. This necessitated the introduction of the sixth wave in the model. Calculations show that the sixth wave ended in June 2021, when the number of cases added by all waves in aggregate had reached 486,000.





**Fig. 7.** The dynamics of the coronavirus pandemic in Israel between 22 February 2020 and 1 March 2021. Actual data and modeling results. Curve labels identify the modeled waves.

**Israel.** Israel experienced four outbreaks during the first year, which are described in the model by four waves with  $N_1 = 340,000$ ,  $N_2 = 2$  million,  $N_3 = 4.4$  million, and  $N_4 = 17$  million, respectively. The modeling results and the actually observed number of cases are presented in Figs. 7a and 7b. We see that the model adequately fits the statistical data, especially the first two waves that evolved as in a closed system.

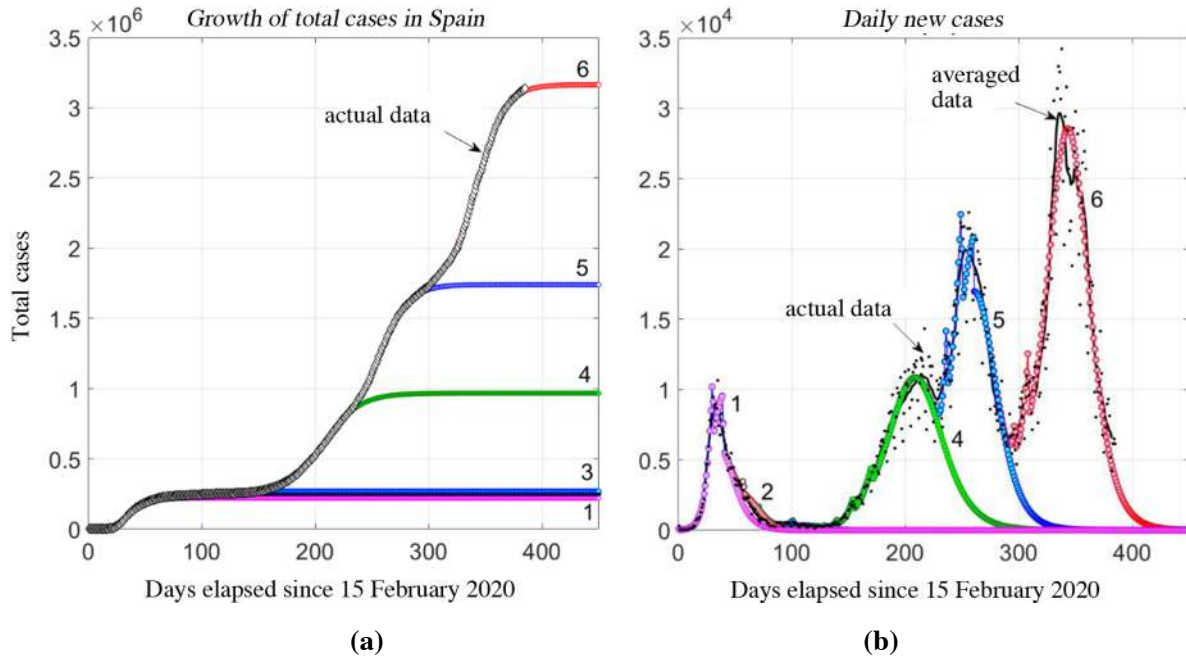
The first wave in Israel reached its peak on 1 April 2020. It is at that time that the first-wave capacity was updated to 340,000 cases. The last growth indicator was  $\alpha = 1.052$ . The model shows that the first wave should end at the end of May 2020 reaching a steady-state value calculated from (2):

$$\bar{y} = \left( \frac{\alpha - 1}{\alpha} \right) \cdot N = \left( \frac{1.052 - 1}{1.052} \right) \cdot 340\,000 = 16806 \text{ cases.}$$

This fully matched the modeling results (16,804 cases in the end of May 2020) and was corroborated by the pandemic statistics which on 27 May 2020 gave 16,872 cases in total. The model forecast thus worked well for two whole months, until the next summer outbreak.

The coronavirus waves in Israel progressed upward. The second wave ended in the beginning of October 2020, having added 84,200 cases. The third wave emerged as the second wave just began to decay. The third-wave peak was observed at the end of September 2020. The government of Israel imposed strict restrictions at the end of September 2020, which slowed down the spread of the infection. In the model, this is manifested in the decrease of the growth rate factor from 18 and 29 September 2020. This third wave ended at the beginning of December 2020 after adding 216,600 cases. We calculated that if the growth rate factor had remained as before 18 September 2020, the predicted total number of cases produced by the third wave should have been 30,000 higher.

The fourth wave was the strongest, with capacity exceeding the total population of Israel. This is an indication of an open system, with cases imported from other countries. The fourth wave has one prominent peak and

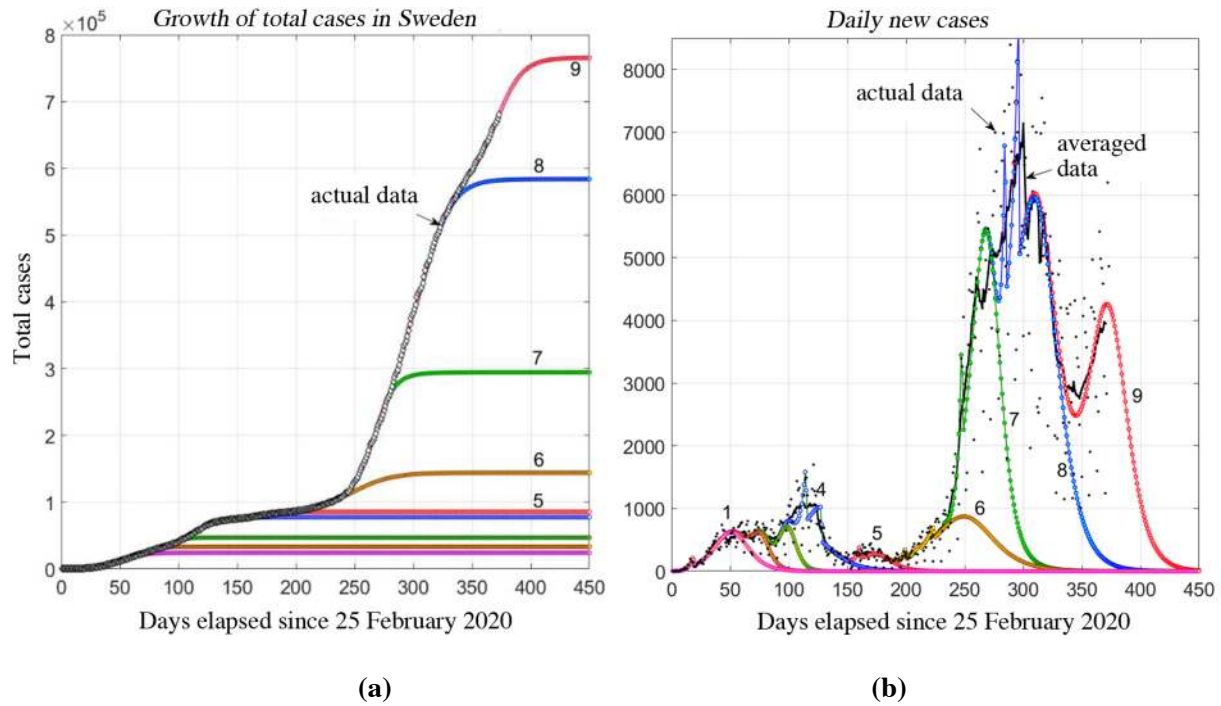


**Fig. 8.** The dynamics of the coronavirus pandemic in Spain between 15 February 2020 and 3 March 2021. Actual data and modeling results. Curve labels identify the modeled waves.

another small local peak on the descending shoulder suggesting a local wave. A more exact description could have been achieved by introducing this local wave into the model, but since the peak was small and tightly adjoined the principal wave, we described these two waves by one averaged wave. The deviation of modeled results from observations is shown in the graph of daily new cases. The error in the total number of cases in the vicinity of substantial deviations averaged 2%–2.5%, but not more than 3%. In most of the remaining cases, the error was less than 1%. The fourth-wave peak was observed in mid-January 2021; now at the end of February 2021 the fourth wave is decaying. The daily count of new cases displays strong volatility. The model suggests that the fourth wave will decay in the end of June 2021 after adding 504,700 cases. The total number of cases produced by all four waves will be 822,500.

**Spain.** Spain is one of the European countries that suffered most from the coronavirus pandemic. Modeling results and statistical observations on the number of cases are shown in Figs. 8a and 8b. The propagation dynamics of coronavirus in Spain is similar to the dynamics in Israel. Spain also experienced from the beginning four large waves, with each successive wave stronger than its predecessor. The first spring outbreak passed the peak in the twenties of March 2020 and then began to decline. But the decay phase stretched for a long time, accompanied by several small bursts. While the ascending phase took about one month, the decay continued for three months. To achieve a better description of the slow decay of the first wave, we introduced in the model two additional local waves with small capacities:  $N_2 = 0.52$  million and  $N_3 = 0.5$  million, respectively; the capacity of the principal first local wave was  $N_1 = 4.3$  million. The total number of cases produced by the first wave was approximately 271,700.

A second large wave began to emerge in Spain according to our model in the twenties of June 2020. Its emergence, as in other countries, is associated with the relaxation of the restrictions and the opening of the borders. In the model, the second wave has  $N_2 = 23$  million, which is much greater than the first-wave capacity. The second-wave peak in the model passed on 10 September 2020, adding 10,855 cases at the maximum.



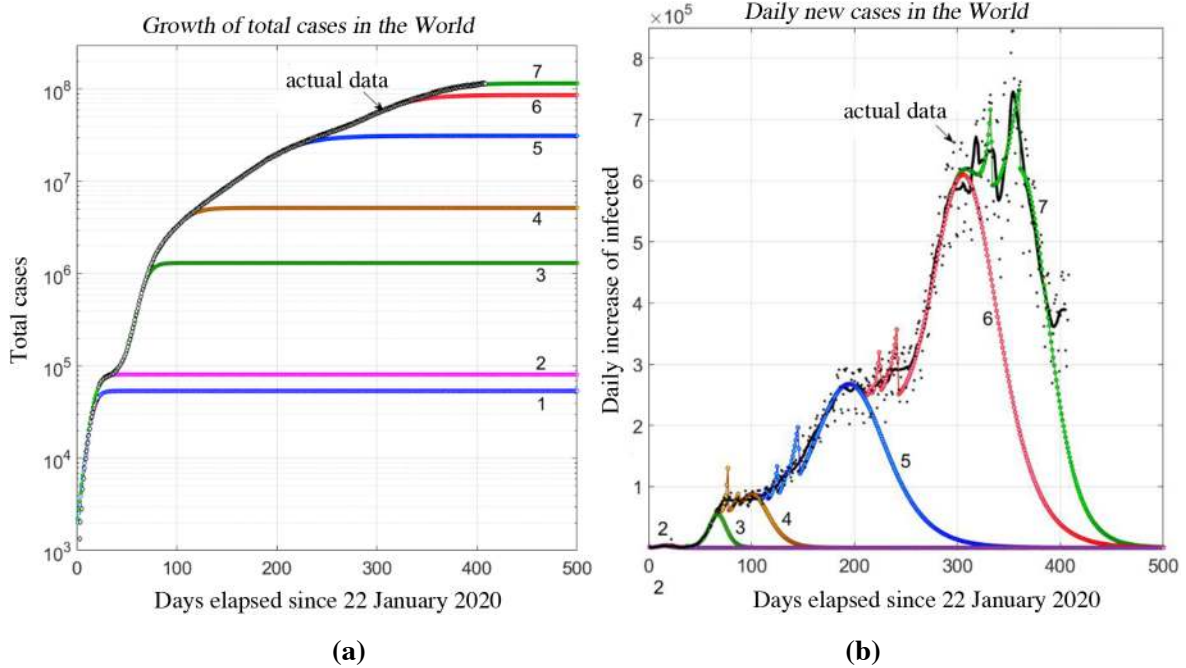
**Fig. 9.** The dynamics of the coronavirus pandemic in Sweden between 25 February 2020 and 3 March 2021. Actual data and modeling results. Curve labels identify the modeled waves.

The second large wave in the model ended in January 2021, adding about 696,000 cases in total. The restrictions that the government of Spain began to introduce in August and September 2020 reduced the growth parameter  $\alpha$  and lowered the infection rate. These restrictions, however, did not stop the next wave, and after a brief decline the number of new cases in Spain again started increasing. The new third outbreak is described in the model by a fifth local wave. It began in the end of August 2020. Its peak was observed in the very end of October 2020. The model shows that it ultimately decayed in February 2021, after contributing 773 new cases.

The fourth disease outbreak in Spain is described in the model by the sixth wave with  $N_6 = 37$  million. Its peak passed in the twenties of January 2021 and the wave is currently declining. The sixth wave is projected to disappear in the end of May 2021, after adding 1.423 million cases. In total, these waves will have contributed 3.164 million cases.

**Sweden.** Sweden is the only European country, other than Belarus, that did not impose strict restrictions and allowed all enterprises, schools, cafes, restaurants, theaters, etc., to continue operating. Figure 9 presents the actual observations and the modeling results for the total number of COVID-19 cases and the daily number of new cases in Sweden. We see that the statistical data are highly volatile, more so than in other countries.

The first outbreak in Sweden began in the end of February 2020 and settled to a plateau at the end of April 2020. The plateau persisted for two months, until the end of May 2020. The daily new cases in this period fluctuated from 261 to 808, and the daily average new cases reached 560. In June 2020, there was a local coronavirus outbreak which peaked on 24 June 2020 with 1,698 new cases daily. After that, the pandemic began to decline rapidly, reaching its minimum on 12 July 2020 with 106 new cases daily. The lower plateau with relatively small local maxima persisted until the twenties of September 2020. The average number of new cases was 236 daily.



**Fig. 10.** The dynamics of the coronavirus pandemic in the World between 22 January 2020 and 3 March 2021. Actual data and modeling results. Numbers attached to curves denote the modeled waves. (a) log scale; (b) linear scale

To achieve adequate modeling of the long plateaus and the local maxima of the spring 2020 outbreak in Sweden, we used a superposition of five local waves with combined capacity of 1,780,000. The last local wave relating to the spring outbreak formed in the second half of July 2020 and decayed in the end of September 2020. The spring outbreak in Sweden took longer and produced a greater loss of life than in the neighboring European countries, which had imposed restrictions. The model shows that the spring 2020 outbreak produced 86,000 cases.

The autumn-winter outbreak in Sweden was much stronger than the spring-summer outbreak, as in other European countries. It began to rise in Sweden in the second half of September 2020 and reached its absolute maximum on 23 December 2020 with seven-day average of 7,136 new cases daily. We successively introduced three local waves (seventh, eighth, and ninth) into the model for the description of the autumn-winter outbreak, but an even greater number of local waves could have been introduced to describe exactly all the local maxima. In mid-January 2021, the pandemic in Sweden began to decline, but this was a brief phase. Already after 20 February 2021, the daily count of new cases began to increase on average. The average increments rose from 2,796 new cases daily on 8 February 2021 to 3,962 in the peak of 4 March 2021. The ninth wave was introduced into the model to describe this outbreak.

The seventh wave ended in mid-January 2021, after producing 150,000 cases. The eighth wave ended in the model in mid-April 2021, after adding 290,000 cases. The ninth wave is projected to end in the beginning of June 2021; it will add another 182,000 cases. The nine waves in total will have produced approximately 765,000 cases. It is impossible to predict today whether the pandemic will decline, continue to fluctuate near the peak for a long time, or flare up again.

**The World.** Figures 10a and 10b present the actual data and the modeling results for the total number of COVID-19 cases in the world. The spread of the pandemic in the world is a superposition of many epidemiological waves originating from various foci in various regions. Even a superficial analysis of statistical data

**Table 1**  
**Parameters of the local waves describing the pandemic spread in the world**

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 4	Wave 6	Wave 7
Capacity $N$	0.5 mln	0.25 mln	10 mln	90 mln	1.3 bln	2.6 bln	1 bln
Start date	Dec. 2019	End Jan. 2020	Feb. 2020	Mid-Mar. 2020	End Apr. 2020	Aug. 2020	Oct. 2020
End date	End Feb. 2020	Begin Mar 2020	Mid-May 2020	End July 2020	End Mar 2021	Aug. 2021	Aug. 2021
Number of cases	53,900	27,400	1.232 mln	3.876 mln	26.1 mln	55.2 mln	29.6 mln

shows that the coronavirus spread in the world in three distinct global waves that raised the pandemic to new levels. The first global wave raised the infection rate to 90,000 cases daily. Its peak occurred on 10 April 2020 with 90,180 new cases daily (actual observations). This peak settled to a plateau that persisted for more than a month, until mid-May 2020. The average number of new cases in this period was 80,669 daily. A new global wave began in mid-May 2020, with the daily count of new cases increasing. Peak counts reached about 290,000 new cases daily at the very end of July 2020. Then the pandemic began to decline and settled to a plateau that persisted for a month-and-a-half until mid-September 2020. The daily number of new cases in this period did not exceed the peak value, and the average daily number of new cases was of the order of 265,000.

In September 2020, a new global autumn-winter outbreak flared up. It was much stronger than the first two outbreaks. The average daily count of new cases began to increase, at first slowly and then at an accelerating rate. The peak values were reached in mid-November 2020 (about 590,000 new cases daily). Over the next two months, the pandemic evolved slowly. The absolute maximum was observed on 11 January 2021 with the average of 745,000 new cases daily. Then a decline started, which continued for approximately 40 days. The daily count of new cases dropped to 360,000. In the last three weeks, a slow emergence of a new global wave is observed.

In our model, the first global outbreak is adequately described by a sum of four local waves, the second global outbreak by a single wave (the fifth local wave), the third autumn-winter outbreak by two local waves (the sixth and the seventh). The waves are shown in Fig. 10. The capacities, the start and end dates, and the total number of infected produced by these waves are shown in Table 1.

The first two global waves describe the spread of the epidemic in China. In the beginning of March 2020, they jointly produced about 81,000 cases, i.e., a little more (by 500 cases) than the number of infected in China at that time.

The third and the fourth waves mainly describe the spring outbreak in Europe, USA, and Russia. The third wave forming in mid-February 2020 and ending in mid-May 2020 produced 1,232,000 cases. The fourth wave reached its peak in the beginning of May and finally decayed at the end of July. The model shows that the fourth wave produced 3,876,000 cases.

The fifth wave in the model described the second global outbreak, when the pandemic reached and began to spread rapidly in Latin America (Brazil, Peru, Chile, and others), in South Asia (India, Pakistan, Bangladesh, and others), and Africa. This wave received a major boost from the summer outbreak in the USA – a result of

protest demonstrations. The peak of the daily new cases in the fifth wave was reached in the model on 3 August 2020, with 268,000 new cases daily. The fifth wave ended in the model in the fourth week of March 2021, after adding about 26.1 million cases in the world.

The sixth wave in the model describing the autumn global outbreak covers the spread of the disease in multiple countries, including the accelerated growth of morbidity in Europe, Russia, and USA. In October 2020, we considered two scenarios for the development of the sixth wave:  $N = 2$  billion and  $N = 3$  billion, respectively. Both scenarios provided a good fit of the observations. In the second half of November 2020 it became clear that an intermediate scenario with  $N = 2.6$  billion had been realized. The sixth wave passed its peak on 20 November 2020. The model shows that it ended in August 2021, having produced 55.2 cases.

The seventh waves began to form in October 2020. It describes a broad peak with several local maxima corresponding to the winter outbreak (Fig. 10b). The absolute maximum of the seventh wave is observed in the model on 14 January 2021 with 747,700 new cases daily. This is consistent with the statistical data, which reached a peak on 12 January 2021 with 741,400 new daily cases (averaged value) The seventh wave is now declining. It is projected to end in August 2021, after adding 29.6 million cases.

The daily count of new cases has been increasing for three weeks in a row and the seventh wave now lags behind the observations. This implies the onset of a new outbreak. Only time will show how strong the new outbreak is. In the model, it will be described by an eighth local waves, with parameters to be determined somewhat later.

## CONCLUSION

For more than a year now, mankind has been monitoring the spread of COVID-19 pandemic in different countries and regions of the world, collecting data and trying to make predictions based on mathematical models. The dynamics of pandemic propagation is very complex. Smoothed epidemic curves in many countries and regions have several local maxima and contain plateaus.

This is attributable to the fact that cities, countries, and region are open nonhomogeneous systems with new infection sources appearing at random in various locations. These sources trigger new infection transmission chains from infected to susceptible individuals. If the chains affect a large number of people and are strongly shifted in time, they manifest as separate epidemic waves that spread differently in the nonhomogeneous environment. In general, the statistical data collected in the form of cumulative and epidemic curves are superpositions of many different local epidemic waves.

Moreover, we should keep in mind during modeling that the statistical data contain large errors, which may reach tens of percent. Determining the model parameters on data with errors is an ill-posed inverse problem with non-unique solutions.

To model such complex propagation dynamics in nonhomogeneous open systems with errors in data, we have used the relatively simple Feigenbaum discrete logistic equation, which in a certain range of parameters describes logistic population growth in an environment with scarce resources. This equation has only two parameters: the capacity and the growth rate of the number of infected. Both affect the final total number of cases. The two parameters are chosen so as to obtain the best fit of the observations. The growth rate parameter is variable. Its first value is chosen in the exponential growth stage. Subsequently, it may be reduced, for instance, due to the introduction of restrictions. The capacity is finally determined in the peak. Several values of capacity are chosen to this end, and several scenarios of pandemic evolution are accordingly considered.

We have developed a procedure for the separation of local waves and choice of parameters for their modeling. Local waves are added to the model successively, as the modeling results begin to lag behind the observations.

The difference between modeling results and observations produces a sequence of values for a new local wave on which the parameters are determined.

Local waves often decay slowly, and their influence on the overall picture of infection propagation persists for a long time. This implies that we should start modeling the evolution of the pandemic from the very beginning. We should not start modeling a new wave if the previous wave is still active.

We modeled the pandemic dynamics and forecast the spread of infection in various countries and regions from March 2020 [41–44] and through February 2021. In this article, we present the modeling results for the spread of the coronavirus pandemic in Brazil, India, USA, Japan, Israel, Spain, Sweden, the megalopolis of Moscow, and the World. The mathematical description of the growth of the total number of cases in India required a single wave, in Brazil three waves, in Israel four waves, in Japan and Spain six wave each, in Sweden nine waves, in USA and Moscow 11 local waves, and in the World seven waves. For each wave, we prepared daily forecasts and updated the wave parameters until the peak was reached. Once the wave had passed the peak, we immediately updated the wave capacity as needed. After the peak, we made a forecast for the particular wave, predicting when the wave would end and how many cases it would produce.

We have shown that our model adequately describes the pandemic propagation dynamics in open systems with uncertainty. The model produces realistic forecasts. The forecast horizon depends on the degree of closure and homogeneity of the system. Thus, in China, Japan, and Israel, where strict restrictions were imposed during the first wave, the forecasts produced an error less than 1% for a month-and-a-half or two months. Theoretical, modeling, and statistical data were all consistent. In India and Brazil, where the spread of infection for a long time had been modeled with a single wave, long forecasts were also possible. On average, a model-based forecast worked for 10–14 days, in the worst case a week. The difficulty is not the accuracy of the model, but rather the openness and unpredictability of the system.

The coronavirus pandemic has not reached its end. In some countries the infection rates are declining, but in other countries new outbreaks occur. The total number of new cases in the World has been increasing in the last three weeks. This may signal the beginning of a new coronavirus outbreak. To contain the situation, we have to continue making model-based forecasts, comparing modeling results with statistical observations, updating the model parameters, or adding new waves and recalculating previous forecasts. Model forecasts should be used as a basis for making decisions to strengthen or relax the restrictions.

## REFERENCES

1. M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, Princeton University Press, NJ (2008).
2. J. D. Murray, *Mathematical Biology: An Introduction*, Springer (2002).
3. W. Liu, H. W. Hethcote, and S. A. Levin, “Dynamical behavior of epidemiological models with nonlinear incidence rates,” *J. Math. Biology*, **25**, No. 4, 359–380 (1987).
4. R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press (1991).
5. W. H. Hamer, “Epidemic disease in England,” *Lancet*, **I**, 733–739 (1906).
6. W. Liu, S. Tang, and Y. Xiao, “Model selection and evaluation based on emerging infectious disease data sets including A/H1N1 and Ebola,” *Computational and Mathematical Methods in Medicine* (2015).
7. Y. H. Hsieh, J. Y. Lee, and H. L. Chang, “SARS epidemiology modeling,” *Emerging Infectious Diseases*, **10**, No. 6, 1165 (2004).
8. S. Kingsland, “The refractory model: The logistic curve and the history of population ecology,” *The Quarterly Review of Biology*, **57**, No. 1, 29–52 (1982).
9. H. W. Hethcote, “The mathematics of infectious diseases,” *SIAM Review*, **42**, No. 4, 599–653 (2000).
10. W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character*, **115**, No. 772, 700–721 (1927).
11. M. Y. Li, J. S. Muldowney, “Global stability for the SEIR model in epidemiology,” *Mathematical Biosciences*, **125**, No. 2, 155–164 (1995).
12. Gabriele Martelloni and Gianluca Martelloni, “Analysis of the evolution of the Sars-Cov-2 in Italy, the role of the asymptomatics and the success of logistic model,” *Chaos, Solitons and Fractals*, **140**, 110150 (2020).



13. P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in COVID-19 with logistic model and machine learning techniques," *Chaos, Solitons and Fractals*, **139**, 110058 (2020); doi.org/10.1016/j.chaos.2020.110058.
14. J. Guo, "Theoretical epidemic laws based on data of COVID-19 pandemic," medRxiv (2020).
15. Ke Wu, D. Darcet, Q. Wang, and D. Sornette, "Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world," medRxiv Preprint; doi: <https://doi.org/10.1101/2020.03.11.20034363>.
16. N. S. Barlow and S. J. Weinstein, "Accurate closed-form solution of the SIR epidemic model," *Physica D*, **408**, 132540 (2020).
17. M. Cadoni and G. Gaeta, "Size and timescale of epidemics in the SIR framework," *Physica D*, **411**, 132626 (2020).
18. T. T. Marinov and R. S. Marinova, "Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models," *Chaos, Solitons and Fractals*, X **5**, 100041 (2020).
19. T. Odagaki, "Analysis of the outbreak of COVID-19 in Japan by SIQR model," *Infectious Disease Modelling*, **5**, 691e698 (2020).
20. A. G. M. Neves and G. Guerrero, "Predicting the evolution of the COVID-19 epidemic with the A-SIR model: Lombardy, Italy and São Paulo State, Brazil," *Physica D*, **413**, 132693 (2020).
21. A. Comunian, R. Gaburro, and M. Giudici, "Inversion of a SIR-based model: A critical analysis about the application to COVID-19 epidemic," *Physica D*, **413**, 132674 (2020).
22. M. Agrawal, M. Kanitkar, and M. Vidyasagar, "SUTRA: An approach to modelling pandemics with asymptomatic patients, and applications to COVID-19"; <https://arxiv.org/abs/2101.09158>.
23. Y. Ding and L. Gao, "An evaluation of COVID-19 in Italy: A data-driven modeling analysis," *Infectious Disease Modelling*, **5**, 495–501 (2020).
24. A. M. Ramos, M. R. Ferrández, M. Vela-Pérez, A. B. Kubik, and B. Ivorra, "A simple but complex enough  $\theta$ -SIR type model to be used with COVID-19 real data. Application to the case of Italy," *Physica D*, **412**, 132839 (2020).
25. C. H. Nkwayep et al., "Short-term forecasts of the COVID-19 pandemic: A study case of Cameroon," *Chaos, Solitons and Fractals*, **140**, 110106 (2020).
26. H. P. C. Cintra and F. N. Fontinele, "Estimative of real number of infections by COVID-19 in Brazil and possible scenarios," *Infectious Disease Modelling*, **5**, 720–736 (2020).
27. S. Contreras, H. A. Villavicencio et al., "A multi-group SEIRA model for the spread of COVID-19 among heterogeneous populations," *Chaos, Solitons and Fractals*, **136**, 109925 (2020).
28. U. Avila-Ponce de León, Á. G. C. Pérez, and E. Avila-Vales, "An SEIARD epidemic model for COVID-19 in Mexico: Mathematical analysis and state-level forecast," *Chaos, Solitons and Fractals*, **140**, 110165 (2020).
29. M. Higazy, "Novel fractional order SIDARTHE mathematical model of COVID-19 pandemic," *Chaos, Solitons and Fractals*, **138**, 110007 (2020).
30. P. Wang, X. Zheng, G. Ai, D. Liu, and B. Zhu, "Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran," *Chaos, Solitons and Fractals*, **140**, 110214 (2020).
31. B. Ghanbari, "On forecasting the spread of the COVID-19 in Iran: The second wave," *Chaos, Solitons and Fractals*, **140**, 110176 (2020).
32. L. F. S. Scabini, L. C. Ribas, M. B. Neiva et al., "Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil," *Physica A*, **564**, 125498 (2021).
33. K. Yew Ng and M. Mei Gui, "COVID-19: Development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and susceptibility," *Physica D*, **411**, 132599 (2020).
34. U. Sahin and T. Sahin, "Forecasting the cumulative number of confirmed cases of COVID-19 in Italy, UK and USA using fractional nonlinear grey Bernoulli model," *Chaos, Solitons and Fractals*, **138**, 109948 (2020).
35. B. S. T. Alkahtani and S. S. Alzaid, "A novel mathematics model of COVID-19 with fractional derivative. Stability and numerical analysis," *Chaos, Solitons and Fractals*, **138**, 110006 (2020).
36. R. May, "Simple mathematical models with very complicated dynamics," *Nature*, **261**, 59–467 (1976).
37. M. J. Feigenbaum, "The universal metric properties of nonlinear transformations," *J. Statistical Physics*, **21**, 669–706 (1979).
38. <https://www.calcalistech.com/ctech/articles/0,7340,L-3800632,00.html>.
39. <https://www.worldometers.info/coronavirus/>.
40. <https://coronavirus-monitor.ru/coronavirus-v-rossii>.
41. E. M. Koltsova, E. S. Kurkina, and A. M. Vasetsky, "Mathematical modeling of COVID-19 coronavirus spread in Moscow," *Computational Nanotechnology*, **7**, No. 1, 99–105 (2020).
42. E. M. Koltsova, E. S. Kurkina, and A. M. Vasetsky, "Mathematical modeling of COVID-19 coronavirus spread in several European, Asian countries, Israel and Russia," *Problems of Economics and Law Practice, Moscow*, No. 2, 154–165 (2020).
43. E. M. Koltsova, E. S. Kurkina, and A. M. Vasetsky, "Mathematical modeling of the spread of COVID-19 in Moscow and Russian Regions," arXiv Preprint, arXiv:2004.10118 (submitted 21 Apr 2020).
44. E. M. Koltsova, E. S. Kurkina, and A. M. Vasetsky, "Mathematical modeling of the spread of waves of the COVID-19 coronavirus epidemic in different regions of the world," arXiv Preprint, arXiv:2007.02283 (submitted 5 Jul 2020).