Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A new modelling of the COVID 19 pandemic

Vladislav Soukhovolsky\textsuperscript{a}, Anton Kovalev\textsuperscript{a}, Anne Pitt\textsuperscript{a}, Boris Kessel\textsuperscript{b,*}

\textsuperscript{a} Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences"
\textsuperscript{b} Hillel Yaffe Medical Center, Rappaport Medical School, Technion, Haifa, Israel.

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 25 May 2020
Accepted 18 June 2020
Available online 20 June 2020

\textbf{Key words:}
epidemics
COVID-19
infection disease modelling
mathematics model

\textbf{A B S T R A C T}

A model of coronavirus incidence is proposed. Process of disease development is represented as analogue of first- and second order phase transition in physical systems. The model is very simple in terms of the data necessary for the calculations. To verify the proposed model, only data on the current incidence rate are required. However, the determination coefficient of model $R^2$ is very high and exceeds 0.95 for most countries. The model permits the accurate prediction of the pandemics dynamics at intervals of up to 10 days. The ADL\textsuperscript{(autoregressive distributed lag)}-model was introduced in addition to the phase transition model to describe the development of the disease at the exponential phase. The ADL-model allows describing nonmonotonic changes in relative infection over the time, and providing to governments and health care decision makers the possibility to predict the outcomes of their decisions on public health.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The COVID 19 pandemic definitively changed the face of the modern world. It emerged even the best national public health care systems before a huge trouble. COVID 19 is a mass causalities event per definition: a situation when the amount of the acute care patients overwhelmed emergency medical services resources, such as personnel and equipment. For "war plan", the intelligence is crucial.

Daily data about the coronavirus pandemic, including the number of new cases and deaths, is collected internationally. The use of these data makes possible the modelling of the international disease spread. However, due to the limited reliability of morbidity and mortality statistics, it is still questionable whether current databases may be used to assess the patterns of the epidemic and its prognosis.

Moreover, neither the significant percentage of asymptomatic virus-infected people nor the real number of recovered patients has not been well-established. Nevertheless, despite the limited and uncertain data, the mathematical modelling of the pandemic's progression is crucial for predicting the possible future spread, preparing the next steps to combat it, and correctly evaluating the efficacy of governmental public health measures already performed.

One of the most important information components is the possibility to develop the high-grade accuracy mathematics model of epidemics progression allowing the medical care providers to organize better the national resources.

Adequate modelling also establishes a framework for future evaluation of this and other pandemics.

This paper describes novel infectious disease modelling, covering many aspects of COVID-19 dynamics.

2. Material and methods

For statistical analysis of coronavirus incidence dynamics, data from the World Health Organization [1] and from Johns Hopkins University [2] from 01.01.2020 to 04.17.2020 was collected.

There are many infectious disease spread models, such as SIR, SEIRD, and SEIHR. Most of them use differential equation systems (continuous time) or difference equations (discrete time) for describing population dynamics with different characteristics in relation to infection [3-7].

In the simplest SIR-model, the entire population is divided into three groups: Susceptible (S), Infected (I) and Removed (R) \cite{8} The SIR model is a homogeneous Markov chain with a single absorbing state R, which can be imagined as a graph with the states of the S, I, R process and with the transitions between the states. The probability $p_{ji}$ defines the potential possibility of moving from state i to state j. A. Verification of the parameters of the SIR-model is based on daily medical statistics. However, in this case, the value of variable I (where patients with latent signs should be included) will not be the same as the I values from the statistics.

Since, at the early stages of an epidemic, the number of S susceptible patients is unproportionally much higher than the number...
of actually ill patients, and the rate of recovery is less than the rate of infection, the SIR-model defines the S value as constant. Thus, at the initial stage of an epidemic's development, the dynamics of infection can be represented as follows:

\[
\frac{dI}{dt} = mI(l - I) = \beta S \exp(mt)
\]

(1)

where \(m = \beta S = \text{const.}\)

However, model [1] is unable to recognize the nonmonotonic changes of process dynamics (for example, resulting from quarantine measures). In general, the SIR model describes the ending of the epidemic, as a process of exhausting the extent of group S, assuming in advance that the total population will be infected.

Most other modelling systems use Markov models that have all the above-mentioned disadvantages. [9,10]. The proposed bifurcation models that include the sudden increases in incidence at certain stages of the epidemic [11] are regression models and do not explain the reasons for such bifurcations.

We propose that the prevalence of coronavirus infected population can exist in one of two states. The number of cases is low and is kept constant in time at the initial low-infection stage of the disease. The viral epidemic goes to an exponential stage with very high incidence later. The coronavirus pandemic may be interpreted as the biological analogue of a well-known physics phenomenon, first-order phase transition. An example of such a phase transition is the condensation of a substance from gaseous to liquid state, when the density of gas molecules increases sharply from some critical temperature or pressure. In general, the function of thermodynamic potential (free energy) may be presented as a bimodal function with two local minimums and a maximum between these lows - a potential barrier.

Theory and Calculations: Theory:

With the bimodal function of the \(F(x)\) state, the system could be near one of two stable states - low \(x_1\) or high \(x_2\). A "soft" system with a small height and a small half-width of a potential barrier easily alters its phase state. On the contrary, a "tough" system with a high potential barrier and a high half-width value of a potential barrier will last quite some time in this condition.

The transition from a stable state with low incidence to an epidemic state will occur when some critical value of \(x_c\) of incidents is reached. The ecological potential of the population morbidity \(F(x)\) will be defined as follows: consider a long time series of coronavirus cases since the first patient. In this series, there will be a significant proportion of events (days in time series) with cases close to \(x_1\). These events will be responsible for the state of the low-incidence population. The other part of the events will have values close to \(x_2\), the state of epidemic. Finally, a small proportion of the events will have values close to the critical value of \(x_c\). When, rarely, there are many cases close to this value, there is a high rate of system transfer from \(x_1\) to \(x_2\).

If the proportion of people with this disease is low, and the proportion of time when the population is observed with a slightly higher incidence compared to \(x_1\) is small, then condition \(x_1\) can be said to be stable. The potential function of \(F(x)\) can be represented as the reverse function of the daily infection incidents' density distribution function. In a state of initial stage of epidemic with low incidence, the number of cases will be near the \(x_1\) and value of \(F(x_1)\) will be small. If we consider the epidemic development as a first-order phase transition, this population will be near the critical value of \(x_c\) number of patients for a very short time, and the value of the function \(F(x_c)\) will be large.

The process of the pandemic's growth can be considered as an analogue to a second order phase transition, in which the connection of the system's integral characteristics (the order parameter \(w\), varying from 0 to 1) to the impact of the external factor \(Y\) is described by the equation \(w^2(Y) = A - BY\) [12,13]. In this version of the second order phase transition model, the role of the external factor \(Y\) is significant from the pandemic's absolute beginning on the planetand \(w(T)\) is the portion of countries with coronavirus incidents at time \(T\).

To simulate the processes of coronavirus dynamics in an epidemic state, we will use data on the current daily incidence of disease \(f(t) - f(t-1)\). Since the \(n\) population in different countries can vary greatly - from billions to tens of thousands, we obtained, as comparable data for each country, indicators of relative current incidence introduced as \(p(t) = \frac{f(t)}{n} - f(t-1)\) per inhabitant. To reduce the variance, the value of \(\ln p(t)\) will be considered instead \(p(t)\).

Under the assumption that the reported incidence on day \(t\), after the onset of the epidemic, depends both on the duration of the \(t\) epidemic since the beginning of its exponential stage and on the disease in \(k\) of the previous days. Then, in the logarithmic scale, you can use the ADL \((k,1)\) (autoregressive distributed lag) - model [14].

\[
\ln p(t) = a_0 + \sum_{j=1}^{k} a_j \ln p(t-j) + bt + \varepsilon(t)
\]

(2)

where \(p(t)\) - relative morbidity, \(t\) - duration of the exponential phase of the epidemic (number of days since its beginning), \(a_0, a_2, b\) - odds, \(k\) - the order of autoregressing, \(\varepsilon(t)\) - noise.

Eq. (2) can be considered as a regression with the available values of the current relative morbidity, the relative morbidity during the previous days, and the duration of time since the beginning of the exponential epidemic stage. To calculate regression coefficients, it is necessary to evaluate the order of \(k\) autoregression. The easiest way to estimate \(k\) is to calculate the partial autocorrelation function (PACF) of the simulated time series and to determine the number of PACF members that are significantly different from zero [15].

The quality of approximation with the ADL model may be determined by two indicators - the degree of consent of the accounting data and the model number of cases and the magnitude of the phase shift between the number of accounting data and the model time line. The consent of the amplitude of model and in-kind data was assessed by the \(R^2\) regression equation determination ratio [2]. To determine the phase shift of the model and the natural series of morbidity, it is possible to calculate the cross-correlation function (CCF) of the model time series and the number of in-kind data.

Calculations:

The coronavirus epidemic model as a first-order phase transition was tested on data of the disease's development in 55 administrative divisions of the United States. According to the data from 01.22.2020 to 04.12.2020, the \(F(x)\) function was built to distribute the number of cases in one day and the reverse function of \(F(x)\).

For states with different populations and different levels of morbidity, critical numbers are very close when the exponential phase of the epidemic begins (Table 1).

Similar data was obtained for all countries from the WHO registry (Fig. 3).

The peak function of \(F(x)\) accounts for the critical number of cases = 30. Thus, this means that the epidemic is probably going
Fig. 1. The potential function $F(x)$ and the first-order phase transition for patients throughout the United States from 01.01.2020 to 04.12.2020 (total 553,152 cases).

Fig. 2. The dynamics of coronavirus development in different US states. Data is aligned with the start date of the exponential phase of the epidemic in the state, taken as $t = 0$ (1 - North Dakota, 2 - California, 3 - Illinois, 4 - Wyoming).

Fig. 3. The potential function $F(x)$ and the first-order phase transition for patients in all countries from the WHO registry from 01.01.2020 to 04.08.2020.

to an exponential phase when the number of infected is between 20 to 30 people in a country.

It is possible on next step to calculate the change in time of incidence within large territories or the planet as a whole, using the second order phase transition model. Fig. 4 shows a change in the time of the order parameter square for 211 countries in the WHO registry.

A similar pattern is true for a large country such as the United States. Fig. 5 shows the change, over time, in the relative number of states and other administrative units in the U.S. where the coronavirus was registered (data taken from the Johns Hopkins University site).

Fig. 6 shows the dynamics of the relative coronavirus incidence in the United States from 01.23.2020 to 04.08.2020.

As the epidemic develops in a single country, the following phases may be identified:

- lag-phase, when a consistently low number of cases is recorded over a long period (1 - 2 people per country);
- exponential phase, when the relative number of cases (the number of cases per inhabitant) increases exponentially over time;
- stabilization and fading phase, when the relative number of cases does not change over time and later begins to decrease.

As shown in Fig. 6, the lag-phase of the disease lasted from January 23 (the date on which the first patient was recorded, according to WHO) until March 6, after which the disease went into an exponential phase, which can be approximated by the equation $h(t) = 0.2136t - 9390$ (where $h(t)$ is the estimated relative number of cases).

The ADL model may be used to more accurately approximate incidence data. To estimate the order of the autocorrelation member in Eq. (2), a partial autocorrelation function (PACF) of a number
of statistical data differentials and exponential equations \( \ln p(t) - h(t) \) was calculated. This transformation brings the data of a different time range to a stationary form. Fig. 7 shows PACF for the U.S. time series from 03.06.2020 to 04.08.2020.

It has shown in Fig. 7, that the order k of the autocorrelation function in [2] equals 2 and, therefore, it is possible to write the equation of the morbidity dynamics:

\[
\ln p(t) = a_0 + a_1 \ln p(t-1) + a_2 \ln p(t-2) + bt + e(t) \quad (3)
\]

Equation ratios [3] can be interpreted as the susceptibility of magnitude \( \ln p(t) \) to change values \( \ln p(t-1) \), \( \ln p(t-2) \) and \( t \):

\[
X_1 = \frac{\partial \ln p(t)}{\partial \ln p(t-1)} = a_1; \quad X_2 = \frac{\partial \ln p(t)}{\partial \ln p(t-2)} = a_2; \quad X_t = \frac{\partial \ln p(t)}{\partial t} = b \quad (4)
\]

If the susceptibilities [4] are positive, then with the growth of \( \ln p(t-1) \), \( \ln p(t-2) \), relative contamination rates \( b(t) \) also grow. With negative values of [4], relative infection rates fall.

The calculations showed that, not only for the U.S., but also for other countries, the order k of autoregression for model [3] may be adopted equal to 2. Fig. 8 reflects the epidemic’s dynamics in the U.S. and the model’s calculations [3] with the Statistica 10 package.

The calculated data for the parameters of the ADL model for the exponential phase of the disease in the United States are presented in Table 2.

The values of \( a_1 > 1 \) can be characterized as a parameter proportional to the relative number of latent and manifest cases at the moment \( t-1 \). As can be seen from Table 2, the determination factor \( R^2 \approx 0.95 \) for the ADL model is higher than the \( R^2 \approx 0.85 \) for a simple exponential model (Fig. 6). Temporary series of data on relative morbidity and model calculations by [3] is synphase. This is seen on the cross-correlation function (Fig. 9).

![Fig. 7. PACF time series for USA (1 PACF; 2 - St. Err.).](image)

![Fig. 8. Data for USA (1), the model using data to 04.04.2020 (2) and the forecast (3) according to the model (3) from 04.09.2020 to 04.17.2020.](image)

![Fig. 9. The cross-correlation function (CCF) of statistics data and model series for United States. (1 - CCF; 2 - St. Err.).](image)

![Fig. 10. Data for Russia (1), the model according to the data until 04.04.2020 (2) and the forecast (3) according to the model from 04.09.2020 to 04.17.2020.](image)

| Table 2. Parameters of the ADL-model of the epidemic in the United States. |
|------------------|----------------|----------------|----------------|----------------|
| Parameters of model | value | Std.Err. | \( t(32) \) | p-level |
| \( a_0 \) | 0.429 | 0.756 | 0.568 | 0.574 |
| \( b \) | -0.010 | 0.011 | 0.962 | 0.343 |
| \( a_2 \) | -0.096 | 0.185 | 0.519 | 0.607 |
| \( a_1 \) | 1.099 | 0.176 | 6.236 | 0.000 |
| \( R^2 \) | 0.996 | 2.662 | 1.0 |

St. err. - standard error of the coefficient, \( t \) - value of \( t \)-Student criterion for the coefficient; p - confidence level.

In a number of countries, the disease intensity begins to decrease, and a monotonic exponential model cannot describe this process. Fig. 10 shows data on the incidence of coronavirus in Russia.

Table 3 shows the calculations of the ADL model’s parameters model to describe the epidemic in Russia.

As shown in Table 3 and Fig. 10, the ADL model describes very well the development dynamics of the disease in Russia. Similar calculations of the ADL-model coefficients were performed for a number of countries and are summarized in Table 4.

Table 3. Parameters of the ADL-model of the epidemic’s development in Russia.

| Parameters of model | value | Std.Err. | \( t(24) \) | p-level |
|------------------|-------|-----------|----------------|----------------|
| \( a_0 \) | -10.516 | 2.805 | -3.760 | 0.000 |
| \( b \) | 0.115 | 0.032 | 3.620 | 0.001 |
| \( a_2 \) | -0.369 | 0.172 | -2.146 | 0.042 |
| \( a_1 \) | 0.761 | 0.169 | 4.489 | 0.000 |
| \( R^2 \) | 0.987 | 709 | 200 |

Using the values of the parameters of the model [3] for individual countries, we can give a short-term prognosis of incidence. Indicators of the quality of the forecast can be defined as the difference between the data \( \ln p(t) \) of the official statistics of morbidity and forecast calculations (Table 5).

The minus sign in Table 5 means that the incidence data is less than the predicted estimates. Such ratios observed for Korea and Russia, and it is presumed that their slight decrease in incidence compared to the model may be due to quarantine measures. The same is true for data during the week from 04.09 to 04.16 in Israel.
that is, the coronavirus spread rate in the United States is more than twice the spread rate across the planet. These differences appear to be related to the peculiarities of population movements, governmental public health measures, economics etc. These patterns of the disease spread need to take into account when planning how to fight the pandemic.

To describe the behavior of a bistable system, characteristic times can be entered - the characteristic time $\tau$ of oscillations near the state $x_1$, the characteristic time of oscillations near the state $x_2$, and the typical time $\theta$ of the system’s transfer from state $x_1$ to state $x_2$ and back. In this case, two time series of cases can be accounted: a temporary series characterizing the population near $x_1$ and a time range characterizing the population near $x_2$. These series may be considered stationary, the values of which fluctuate near the values of $x_1$ and $x_2$.

Many questions were referred immediately to the public health systems worldwide with burst of COVID 19 pandemics. Even advanced national care systems found itself unprepared to “tsunami-like” growing amount of corona virus patients. The major efforts are being made in achieving of adequate medical staff, equipment, drugs and etc. However, the basic strategy planning of every single country is depends on prognostic epidemiologic modelling while the strict monitoring of the rapidly changed parameters. The mathematical models were created by scientists in order to predict the course of the outbreak and identify governmental interventions that could be effective in decrease of epidemics progression.

Three possible approaches for such modeling may be considered. The first approach is to model the available data using nonlinear regression (for example, the agent-based model [16]. Using this method allows obtaining accurate estimates of existing data and give a short-term forecast of incidence, but it is difficult to describe the mechanisms of the pandemic.

Most coronavirus spread models use various versions of the SIR-model (for example, SEIR-models [17-19]). In this case, the problem is the size estimation of the S-group. If we accept that the entire country’s population should be attributed to this group, then, based on the SIR-models, it should be concluded that the pandemic will stop only when the majority of the population becomes ill [20,21]. Moreover, the SIR-models do not raise the question of the critical number of patients, after reaching the beginning of the exponential phase. Choosing the model of sufficiently large initial values of the cases number makes it difficult to answer this question.

Finally, a recursive bifurcation model is proposed in which the development of a disease is considered a transition from a condition with a low number of cases of the epidemic [11]. This method is similar to our proposed approach. Our model describes the infection processes as first and second order phase transitions. The development model of the coronavirus pandemic as a first-order phase transition indicates either the possibility of the population returning to a state with a low level of cases or the epidemic returning. In this case, the appearance of a second or multiple waves of the disease depends on fluctuations in the number of patients near state $x_1$, the height of the potential barrier and the influence of various external factors. The possibility of the pandemic's
later waves appearing in the model as phase transitions is easily explained, while SIR-models and derivative models do not predict such effects. In Markov models, the influence of an external field can be introduced only by force through specified changes in the coefficients of the transition matrix.

The possibility of describing the spatial dynamics of pandemic development as second order phase transitions is somewhat unexpected. Perhaps this effect may be related to the peculiarities of the movements of infected people.

Conclusion
The model is very simple in terms of the data necessary for the calculations. To verify the proposed model, only data on the current incidence rate are required. However, the determination coefficient of model R2 is very high and exceeds 0.95 for most countries. The model permits the accurate prediction of the pandemics dynamics at intervals of up to 10 days. The ADL model allows describing nonmonotonic changes in relative infection over the time, and providing to governments and health care decision makers the possibility to predict the outcomes of their decisions on public health.

Declaration of Competing Interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments
The authors are grateful to V. Shelman (Haifa, Israel) for his interest in and his useful comments on this work.

References
[1] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
[2] https://www.github.com/CSSEGISandData/COVID-19 Data Repository by Johns Hopkins CSSE.
[3] Hethcote HW. The Mathematics of Infectious Diseases. SIAM Rev 2000;42(4):599–653 Iss.
[4] Hethcote HW, van den Driessche P. Two SIS epidemiologic models with delays. J Math Biol. 2000 Jan;40(1):3–26.
[5] Brauer F, Van den Driessche P, Wu J. Mathematical Epidemiology. Berlin-Heidelberg: Springer; 2008.
[6] Bonyah E, Khan MA, Okosun KO. Islam S. A theoretical model for Zika virus transmission. PLoS ONE 2017;12(10):e0185540.
[7] Bolzoni L, Bonacini E, Soresina G, Groppi V. Time-optimal control strategies in SIR epidemic models. Math Biosci Math Biosci 2017;292:86–96.
[8] Kernack VO, McKendrick AG. A Contribution to the Mathematical Theory of Epidemics // Proceedings of the Royal Society. 1927. Vol. 115. No. A771. P. 700–721.
[9] Dantze C, Gadelha Pereira J, Garcia Gonçalves LM, Piscitelli P, Miani A. Forecasting Covid19 outbreak Progression in Italian Regions: A model based on neural network training from Chinese data/ medRxiv preprintdoi:https://doi.org/10.1101/2020.04.09.20059055.
[10] Lyra W, de Nascimento J-D, Belinha J, de Almeida L, Chrispim PPM, de Andrade I. COVID-19 pandemics modeling with SEIR(=CAQHI), social distancing, and age stratification. The effect of vertical confinement and release in Brazil// medRxiv preprintdoi: https://doi.org/10.1101/2020.04.09.20060655.
[11] Shen J. A Recursive Bifurcation Model for Predicting the Peak of COVID-19 Virus Spread in United States and Germany// medRxiv preprintdoi: https://doi.org/10.1101/2020.04.09.20059329.
[12] Landau LD. On the theory of phase transitions // JETP. 1937. V.7. pp. 19–32 (Russian).
[13] Landau LD, Lifshits EM. Statistical physics. Oxford, New York: Pergamon Press; 1980. p. 567.
[14] Stock JH, Watson MW. Introduction to Econometrics. New York: Addison-Wesley, 2011, pp. 832.
[15] Box GEP, Jenkins GM. Time series analysis: forecasting and control. San Francisco: Holden-Day. 1970. p. 784.
[16] Tuomisto JT, Yrjäla J, Kolehmainen M, Bonolofor J, Pekkanen J, Tikkanen T. An agent-based epidemic model REINA for COVID-19 to identify destructive Policies// medRxiv preprintdoi:https://doi.org/10.1101/2020.04.09.20047498.
[17] Clifford SJ, Klepac P, Zandvoort KV, Quilty BJ, Eggert D, Flashe S. Interventions targeting air travellers early in the pandemic may delay local outbreaks of sars-cov-2. medRxiv. 2020: (doi: https://doi.org/10.1101/2020.02.01.20022426).
[18] Shen M, Peng Z, Guo Y, Xiao Y, Zhang L. Lockdown may partially halt the spread of 2019 novel coronavirus in Hubei province, China. medRxiv. 2220; (doi: https://doi.org/10.1101/2020.02.11.20022236).
[19] Tang B, Braigazzi NL, Li Q, Tang S, Xiao Y, Wu J. An updated estimation of the risk of transmission of the novel coronavirus (2019-ncov). Infect Dis Model 2020;11(4):248–55.
[20] Allahi M, Porteop P, Carlb M, and Gibert D. Prediction of the time evolution of the COVID-19 disease in Guadelupe with a stochastic evolutionary model // medRxiv preprintdoi:https://doi.org/10.1101/2020.02.11.2002236.
[21] Edirweera D, de Silva N, Malavigne N, de Silva J. An epidemiological model to aid decision-making for COVID-19 control in Sri Lanka// medRxiv preprint doi:https://doi.org/10.1101/2020.04.11.20061481.