

Research of Approaches to Constructing Predictive Models of the Spread of Viral Diseases

Alexander Samarkin

*Institute of Medicine and Biology
Pskov State University
Pskov, Russia
Alexsamarkinru@gmail.com*

Iuliia Bruttan

*Institute of Engineering Sciences
Pskov State University
Pskov, Russia
bruttan@mail.ru*

Natalya Ivanova

*Institute of Medicine and Biology
Pskov State University
Pskov, Russia
medobr@pskgu.ru*

Igor Antonov

*Institute of Engineering Sciences
Pskov State University
Pskov, Russia
igorant63@yandex.ru*

Maria Bruttan

*Phystech School of Biological and
Medical Physics
Moscow Institute of Physics and
Technology (National Research
University)
Moscow, Russia
bruttan.mw@phystech.edu*

Abstract - The article is devoted to the analysis of the available mathematical models in epidemiology and the possibility of their modification. We note that the situation with the COVID-19 virus pandemic is characterized by several features not comprehensively studied in the existing models. For a rational response to existing challenges, it is necessary to have a predictive and analytical apparatus in the complex (national and regional scale) mathematical models with a planning horizon of 2 years (the expected period of mass production of vaccines). The article discusses the existing approaches to predicting the spread of the COVID-19 virus in Russia based on mathematical models of epidemics. The possibilities and limitations of the proposed approaches are considered. In the conditions of the Russian Federation, transport connectivity at the interregional and intraregional levels plays an important role, and for megalopolises - transport flows within large agglomerations and the age structure of the population. In contrast to previous pandemics and epidemics, public policy plays a significant role. The approach, which consist in building multi-agent models that combine the advantages of compartment models and models based on the Monte Carlo method (individually oriented) is proposed by the authors. It is planned to use compartment models to assess the dynamics of the process and individually-oriented models - at the level of individual territories and districts.

Keywords - prediction, predictive models, viral diseases, mathematical model.

I. INTRODUCTION

The SARS-COVID19 virus pandemic is an object of close study by both epidemiologists and specialists in data analysis and numerical modeling, since the high degree of informatization of developed countries allows to obtain open statistical information practically on the day of its registration.

An analysis of the spread of viral infections is important both from a scientific point of view and from a practical viewpoint, since the volume of sick people in the population determines the burden on the health care system, makes it possible to estimate the required number of beds in hospitals, the amount of drugs consumed, the number of ventilators and medical personnel, the nature and severity of the applied restrictive measures.

Mathematical modeling of these processes has been carried out since the 20s of the XX century, however, a modern highly mobile society, a developed healthcare system, the specificity of the COVID-19 virus as such sets the task of developing and verifying, if not completely new mathematical models in a significant way corrected model.

Thus, the work is relevant, and the proposed approaches can be used for prognostic by government agencies, the leadership of medical institutions and the business community.

Online ISSN 2256-070X

<https://doi.org/10.17770/etr2021vol2.6578>

© 2021 Alexander Samarkin, Iuliia Bruttan, Natalya Ivanova, Igor Antonov, Maria Bruttan.

Published by Rezekne Academy of Technologies.

This is an open access article under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

II. MATERIALS AND METHODS

A. Base model

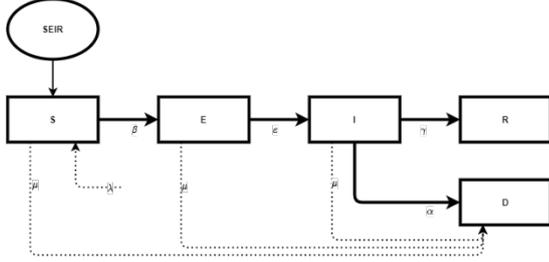


Fig. 1. Base SEIR model with populations dynamic (where S - susceptible, E - exposed, I - infectious, R - recovered), D - dead part of population, measured in a natural values).

Consider as a reference the SEIR model proposed by [1] and widely used at present [2, 3]. This is a modification that takes into account natural and additional mortality and population growth.

The population at each time is considered to consist of compartments, namely S (susceptible), E (exposed), I (infectious), R (recovered), and D (dead). S is the part of the population that can be infected, E is the carriers of infection in the incubation period, I is infected, R is cured. Deaths due to illness and natural attrition from all groups (D) are also included. The probability of transition from one compartment to another due to infection is given by coefficients (β , ϵ , γ , α), and the transition rules are represented by continuous lines. The natural loss of population is assumed to be the same and is given a coefficient μ , and the birth rate is λ . Note that the total number of living people in the population is $N=S+E+I+R$.

The model presented is deterministic and is described by the following system of differential equations (1):

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta}{N}SI - \mu S + \lambda N \\ \frac{dE}{dt} = \frac{\beta}{N}SI - \epsilon E - \mu E \\ \frac{dI}{dt} = \epsilon E - \gamma I - \alpha I - \mu I \\ \frac{dR}{dt} = \gamma I - \mu R \\ \frac{dD}{dt} = \alpha I \end{cases} \quad (1)$$

The first equation describes the probability of infection of immunocompromised individuals depending on the frequency of contact with the infected (SI) at the transmission rate β . Infected individuals become carriers of E infection, which in turn, become infected with I with a coefficient ϵ . Those infected can recover and become R, or die (D) with coefficients γ , α respectively.

A natural decrease with a μ coefficient is derived from all compounds except D and R for obvious reasons. The birth rate is proportional to the number of living population with a coefficient λ . We will then use these designations without further comment.

Everywhere S, E, I, R, D are functions of time (in days), and their initial values are given with the index 0 ($S(0)$, $E(0)$, $I(0)$, $R(0)$, $D(0)$). Several relative parameters are also

traditionally introduced, namely: R_0 - basic reproduction rate of virus, n_{id} - incubation period, $\beta = R_0/n_{id}$ - viral propagation rate, γ - withdrawal rate from infected state (see detailed description, e.g. [4-6]).

Option b) takes into account natural fertility and mortality, which is relevant for long-term epidemics with a cyclical pattern of about a year or more. The considered systems of differential equations are represented in the form of models Simulink (see <https://github.com/Alex-Samarkin/Rezekne2021COVID>).

B. MATLAB realization of the base model

MATLAB and Simulink are widely used for simulation modelling due to the high prototyping speed, an exceptionally convenient visual editor, stable and well-documented numerical modeling algorithms. Then applied: Runge-Kutte45 solver, variable integration step, relative error $1e-6$. The consolidated block diagram of the model is presented in the Fig. 2. There is omitted documentation and info blocks, same as block of visual representation of results.

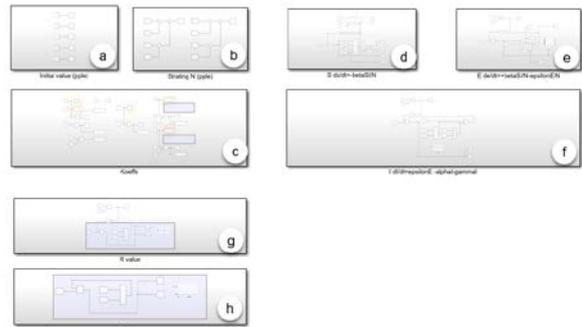


Fig. 2. Realisation of a base model SEIRD (where marked blocks is: a - initial value setup, b - calculator of population size, c - setup of coefficients, D-H - SEIRD calculators).

The following is the most important coefficient calculation block for further modelling.

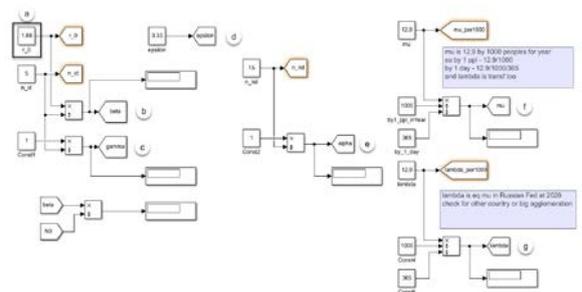


Fig. 3. Calculation of coefficients block (where a - transmission rate coefficient, b,c,d,e,f,g - β , γ , δ , ϵ , μ , ν).

C. Criticism of the base model

There are several assumptions behind the integrated models: the compartments are homogeneous, the external effects have an instantaneous effect on the whole compartment, which obviously does not work under real-world conditions.

The stochastic nature of the processes (e.g., the likelihood of infection) is ignored.

It ignores not only the geographical extent of compacts, but also their length over time. Here, time refers to both the time of the simulation, for example, the year, and the time of the simulated processes: the time of the infected in the incubation period, the time of the disease. It is obvious, for example, that the probability of recovering or dying increases at the end of the disease term (n_{nd}).

Population responses and government action to halt the epidemic, and the socio-economic impact of the epidemic and the response to it, have been ignored, creating a system with multiple feedback loops. This makes it difficult to identify model parameters.

The way the model is described also has a major impact on the modelling procedures. For example, a system-based description of common computers allows the use of efficient computer mathematics systems. A description in the form of algorithms of transition from one complex to another - modeling in one programming language, multi-agent and individualized models may require implementation in hardware and software.

D. Ways to overcome the shortcomings of the standard model

One of the main drawbacks of the model is that its results are deterministic. It arises because of the deterministic nature of model parameters and deterministic rules of transition from one state to another. It can be shown ([7, pp. 384-388]) that transitions from one compartment to another in a stochastic interpretation of this process are subject to the law of exponential distribution. The same approach can be used in the analysis of time-long compounds (such as the E, I base model).

The SEIR model is well developed in the modelling of most viral infections [8, 9], so the authors have chosen an evolutionary modernization. The main idea is to introduce additional compartments into the model (for example, by geographical principle or by age groups). It also seems logical to model the process of the disease, as a transition from a compound with a period of illness i to a complex with a period of illness $i+1$.

Models based on the time distribution of patient groups are known, as the Erlang-SEIR model (Fig. 4).

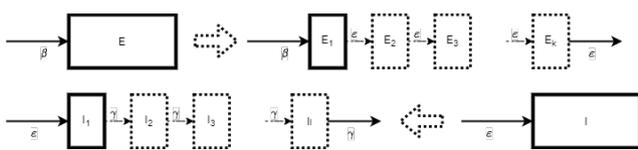


Fig. 4. The Erlang scheme in the incubation E and infected I models.

From the authors' point of view (without detracting from the value of the approach itself), introducing such complexity on relatively small samples does not have a significant effect due to the effect of large numbers,

nevertheless the reader can find their implementation in ErlangSEIR1 - ErlangSEIR3_4 repository.

However, even the more complicated model [10] is actually deterministic. As the basis of her work are coefficients, the authors modelled their fluctuations over time, taking into account both cyclical temporary processes (daily, weekly, monthly, quarterly and annual) and purely random [11].

The determining parameter of the model is the transport factor, which measures the average number of new infected persons in a population made by one infected person per day.

At a value above 1, the epidemic increases, otherwise it fades. The coefficient is integral to an entire population and may vary significantly from one population group to another and is currently calculated in retrospect from the results of day-to-day infection statistics.

The diagram on a Fig. 5 shows the correction block for the reference value of the transport coefficient. The amendments are additive.

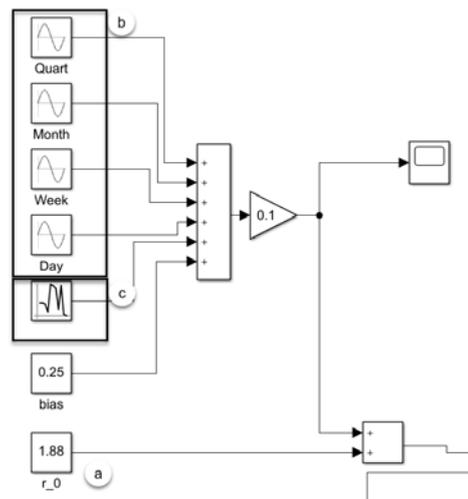


Fig. 5. Block diagram for generating a transport coefficient variable where a) is a base value, b) is a periodic oscillation, c) is a purely random component.

Similarly (the article does not show) the random variation of other model coefficients is considered.

The pandemic is accompanied by governmental action, and more or less significant effects of the containment situation. These effects may occur in different ways, but the main interest is their influence on the transport coefficient.

A typical response of different States to a pandemic (prior to mass vaccination) is the imposition of restrictive measures such as:

- wearing masks;
- restriction of movement;
- introduction of social distance standards.

It is easy to see they are aimed at lowering the transport coefficient and are multiplicative. The multiplicative corrections modelling unit is presented in the Fig. 6.

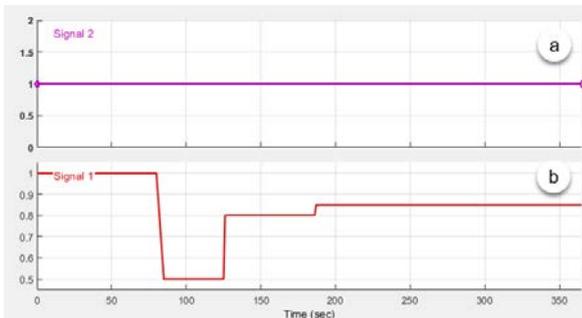


Fig. 6. Multiplicative adjusters of the transport coefficient. Here: a) - no lockdown, b) - three-step restrictive measures.

Figure 6. b shows three stages of lockdown (Pskov region in 2020, reconstructed by authors):

- 80 days – no reaction or little effective action;
- 80-130 – hard lock-down with strict control;
- 130-180 – soft lock-down (some restrictions are lifted);
- 180-365 – ineffective lock-down (restrictions are formally observed, with little or no implementation by private businesses and the public).

As seen below, a step-by-step implementation of the lockdown may cause a second wave of the pandemic (but is not necessarily the cause of the wave).

At the regional level, in Pskov oblast it is possible not to consider the geographical length factor. At the Russian Federation level, however, this may be a determining factor in the relatively mild course of a pandemic.

Of particular interest is the age structure of the region’s population. According to the authors, different age groups may have different contacts with the rest of the population, experience the disease differently and show different mortality rates. Based on data from Rosstat, the population of the Russian Federation and the Pskov region has an established age structure. We will present it in three age groups: children under the age of 15, the working-age population and pensioners over the age of 55.

We note that the first group (children) are almost asymptomatic but have the highest transport activity. Thus, children have a high transmission rate, but a low mortality rate. Note also this group is poorly covered by testing procedures. The able-bodied population has typical statistical indicators - base transport coefficient about 2, with moderate mortality (excluding persons at risk). As with children, this group has weekly and daily rhythms because of the need to go to work. The non-working part of population has the highest mortality, but this is combined with low transport activity.

TABLE 1 AGED GROUPS

	Proportion of aged groups (%)		
	Age 0-15	Age 15-55	Age >55
Value	20%	53%	27

An additional Simulink block library (SEIR_Lib file) was developed to implement the three-year model. The consolidated block diagram of the model is presented in Fig. 7.

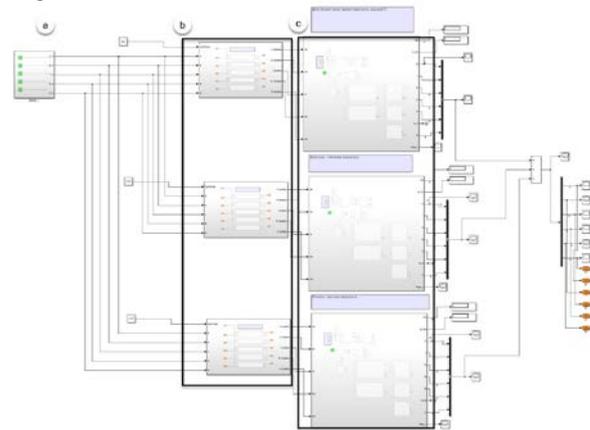


Fig. 7. Model of epidemics by age structure of population. Here: a) - general population structure, b) - age separation blocks, c) - age-specific agents - private SEIR models.

III. RESULTS AND DISCUSSION

The results of the simulation can be divided into groups – verifying correctness of the calculations (not mentioned here) obtained in the calculation of deterministic models, by models considering random factors and rhythms of the transport coefficient, Agency models that consider the age structure of the population.

The results are built by MATLAB using Scope blocks part of Simulink.

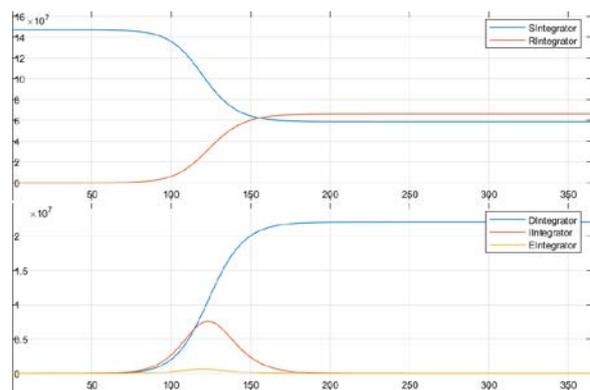


Fig. 8. Classical SEIR model, fully determined by coefficients.

The results on Fig. 8 reflect the classic course of the epidemic – the increasing wave of infections, the formation of natural immunity and the end of the epidemic [13].

By many sources and as we have seen the model does not explain the emergence of a second wave of pandemics [14]. The figures also correlate with the dates of the first

wave of the pandemic in the Russian Federation and Pskov region, but are several times higher than the statistics.

The authors consider that: 1) The geographical extent of the State is significant, as in fact the first wave of the pandemic affected 2-3 major cities; 2) The population testing strategy affects only populations at risk, for which a potentially dangerous course of disease is possible [16]. The multiplicative effect of these factors can explain the deviation of the calculated results from the observed ones [15].

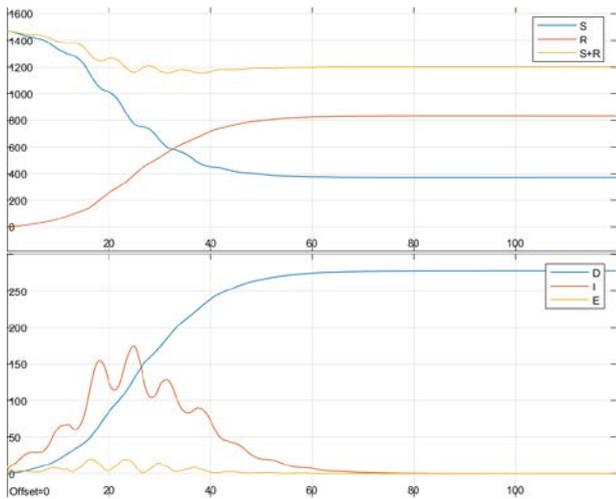
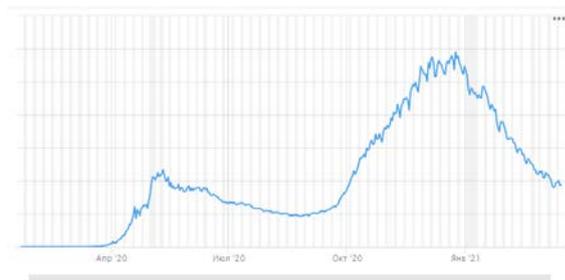


Fig. 9. Erlang-SEIR model considering fluctuations in transport coefficient values (120 days of a pandemic are simulated)

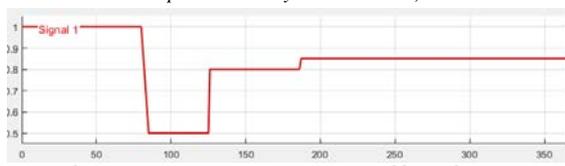
Sinusoidal rhythms that change the transport coefficient cause characteristic sawdust oscillations on the curves of the infected and the recovered. This result allows to simulate similar fluctuations, measured in the count of cured and infected.

However, these fluctuations do not result in a continuous fluctuation in the number of infected persons and therefore do not explain the waves of the pandemic.

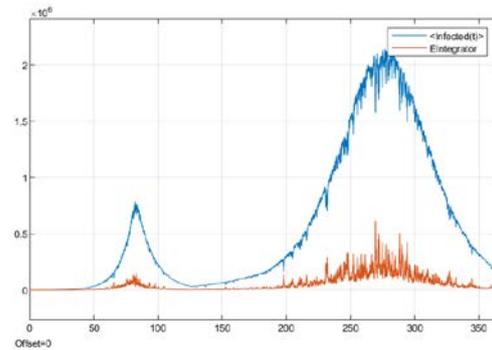
- Yandex DataLens · Public Заражений за день по странам в динамике



a) Count of Infected by day (from <https://datalens.yandex/Россия>)



b) state restriction rate (reconstructed by authors)



c) Count of infected by day - modeled data (from ErlangSEIRv3_2)

Fig. 10. Comparison of real-world data and modelled effect of restriction

The model demonstrates the emergence of a second pandemic due to the relaxation of restrictive measures by the State. The timing is relative good, but there is still a significant difference in the number of reported cases. The reason for the discrepancy is explained above.

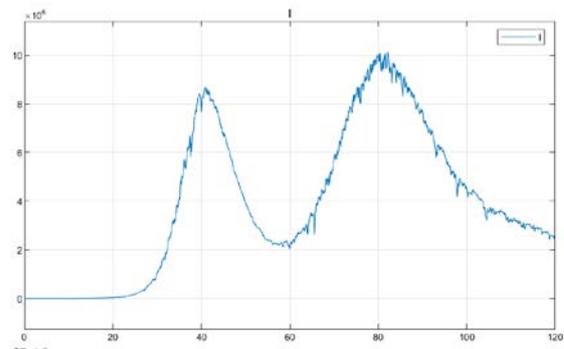


Fig. 11. The bimodal distribution of those infected by pandemic days is modelled considering the behaviour of three age groups, fluctuations in transport coefficient values and excluding restrictive measures (120 days of simulation).

IV. CONCLUSIONS

The presented paper demonstrates a wide range of results that can be obtained using several modified SEER models. The main directions of the modification are:

- The transition from constant coefficients in the system of differential equations (1) to certain functions, such as the sum of harmonic oscillations with periods corresponding to the characteristic rhythms of the population - from day to year);
- Introduction of elements of randomness (additive noise to coefficient values;
- Consideration of State and social control effects (with a multiplicative amendment taking into account the influence of lockdown on the transport factor, see also [17]);
- The formation of agent models in which the selected clusters interact with each other, and

the behavior of the cluster continues to be described by SEER models (the example is given of the division of the population into three clusters by age groups, [18, 19]).

In the continuation of the present work, geographical connectivity between the region and the State must be considered. The influence of mass vaccination processes should also be considered. The authors also propose to add economics-related models to the purely medical aspects.

V. REFERENCES

- [1] William Ogilvy Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. Lond. A*, vol. 115, no. 772, pp. 700–721, 1927, doi: 10.1098/rspa.1927.0118.
- [2] H. W. Hethcote, "Three Basic Epidemiological Models," in *Biomathematics, Applied Mathematical Ecology*, S. A. Levin, S. A. Levin, T. G. Hallam, and L. J. Gross, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 119–144.
- [3] H. W. Hethcote, "The Mathematics of Infectious Diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2000, doi: 10.1137/S0036144500371907.
- [4] C. Bulut and Y. Kato, "Epidemiology of COVID-19," *Turkish journal of medical sciences*, vol. 50, SI-1, pp. 563–570, 2020, doi: 10.3906/sag-2004-172.
- [5] H. W. Hethcote, J. W. van Ark, and J. M. Karon, "A simulation model of AIDS in San Francisco: II. Simulations, therapy, and sensitivity analysis," *Mathematical biosciences*, vol. 106, no. 2, pp. 223–247, 1991, doi: 10.1016/0025-5564(91)90078-w.
- [6] H. W. Hethcote, J. W. van Ark, and I. M. Longini, "A simulation model of AIDS in San Francisco: I. Model formulation and parameter estimation," *Mathematical biosciences*, vol. 106, no. 2, pp. 203–222, 1991, doi: 10.1016/0025-5564(91)90077-v.
- [7] R. M. Anderson, *The Population dynamics of infectious diseases: Theory and applications / edited by Roy M. Anderson*. London: Chapman and Hall, 1982.
- [8] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical Models in Epidemiology*. New York, NY: Springer New York, 2019.
- [9] H. Hethcote, M. Zhen, and L. Shengbing, "Effects of quarantine in six endemic models for infectious diseases," *Mathematical biosciences*, vol. 180, pp. 141–160, 2002, doi: 10.1016/s0025-5564(02)00111-6.
- [10] A. Kumar, K. Goel, and Nilam, "A deterministic time-delayed SIR epidemic model: mathematical modeling and analysis," *Theory in biosciences = Theorie in den Biowissenschaften*, vol. 139, no. 1, pp. 67–76, 2020, doi: 10.1007/s12064-019-00300-7.
- [11] H. Loeffler-Wirth, M. Schmidt, and H. Binder, "Covid-19 Transmission Trajectories-Monitoring the Pandemic in the Worldwide Context," *Viruses*, vol. 12, no. 7, 2020, doi: 10.3390/v12070777.
- [12] M. Melis and R. Littera, "Undetected infectives in the Covid-19 pandemic," *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases*, 2021, doi: 10.1016/j.ijid.2021.01.010.
- [13] D. Rafiq, A. Batool, and M. A. Bazaz, "Three months of COVID-19: A systematic review and meta-analysis," *Reviews in medical virology*, vol. 30, no. 4, e2113, 2020, doi: 10.1002/rmv.2113.
- [14] F. Rojas, L. Maurin, R. Dünner, and K. Pichara, "Classifying CMB time-ordered data through deep neural networks," *Monthly Notices of the Royal Astronomical Society*, vol. 494, no. 3, pp. 3741–3749, 2020, doi: 10.1093/mnras/staa1009.
- [15] Simon A. Levin, "Applied Mathematical Ecology by Simon A. Levin (auth.)", 1986.
- [16] P. J. Turk *et al.*, "Modeling COVID-19 Latent Prevalence to Assess a Public Health Intervention at a State and Regional Scale: Retrospective Cohort Study," *JMIR public health and surveillance*, vol. 6, no. 2, e19353, 2020, doi: 10.2196/19353.
- [17] H. M. Yang, L. P. Lombardi Junior, F. F. M. Castro, and A. C. Yang, "Mathematical model describing CoViD-19 in São Paulo, Brazil - evaluating isolation as control mechanism and forecasting epidemiological scenarios of release," *Epidemiology and infection*, vol. 148, e155, 2020, doi: 10.1017/S0950268820001600.
- [18] C. Wolfram, "An Agent-Based Model of COVID-19," *ComplexSystems*, vol. 29, no. 1, pp. 87–105, 2020, doi: 10.25088/ComplexSystems.29.1.87.
- [19] R. Yaari, A. Huppert, and I. Dattner, "A statistical methodology for data-driven partitioning of infectious disease incidence into age-groups," Jul. 2019. [Online]. Available: <https://arxiv.org/pdf/1907.03441>