
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.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 192 (2021) 3370–3379

Procedia Computer Science

www.elsevier.com/locate/procedia

25th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

# Cellular Automata in Covid-19 prediction

Piotr Podolski[a,*], Hung Son Nguyen[a]

[a] *Faculty of Mathematics, Informatics and Mechanics*
*University of Warsaw*
*Banacha 2, 02-097 Warszawa, Poland.*

**Abstract**

At the end of 2019 a new coronavirus emerged, turning into a world pandemic. The new coronavirus is called COVID-19. Different countries handled the pandemic differently and our main focus in this article is on Poland. For better counteracting and managing the situation a model for predicting the dynamics of the pandemic is needed. In this article we present a model for simulating future infections taking into account various preventive measures and locations in Poland. We based the model on a two-dimensional cellular automata, with spatial dependencies between regions, different population and size of simulated regions.



*Keywords:* Covid-19 prediction; cellular automata; simulation

## 1. Introduction

In recent years we faced and old challenge, which is a viral outbreak of a disease. The COVID-19 virus spread between nations rapidly, leaving its mark on health, economy and lifestyle. In the time this article was prepared, multiple countries were struggling with successive waves of this disease. Vaccination is carried out with different outcomes in different countries and it is uncertain how the effect of mutating the virus will affect the herd immunity. One of the goals of this article is to prepare a tool for forecasting the dynamics of this virus in different regions, making it possible to prevent and suppress the virus in simulated regions taking into account different factors, such as the population density, spatial dependencies and the distribution of age and sex in population. The simulated region that we focused on in this research is Poland in year 2020.

* Corresponding author.
*E-mail address:* podolski.pio@gmail.com

1877-0509 
10.1016/j.procs.2021.09.110

## 2. An overview of the virus

In this article we are proposing a Cellular Automata model, that takes into account the biology of the virus. Thus we find it necessary to give brief overview of the virus and his influence on an individual. A virus is a submicroscopic infectious agent, it comprises of two parts: genetic material and a protective protein coating. It has the ability to quickly replicate itself once it penetrates to a living cell. Disease arises from cell damage caused by the genetic rewriting procedure. One of main goals of the virus is to replicate itself, but once the host acquires antibodies to fight the infection, the virus needs to find another host.

### *2.1. Infection life-cycle*

Once the virus enters a host, it usually starts to replicate itself. When the host has no antibodies, the host becomes infectious and diseased. Based on current knowledge [1] we can show a timeline of infection, showing the life-cycle of a virus.

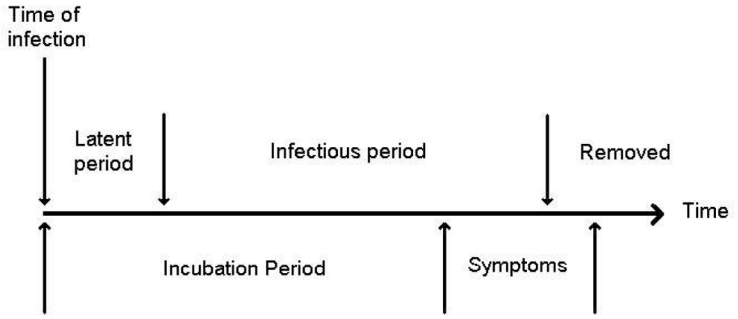



Fig. 1: The relationship between infectiousness and virus symptoms [1]

- **Latent period** It occurs in the early stages of infection, when the virus has not yet developed the ability to spread to a new host.
- **Infectious period** In this phase, the virus is contagious and can be passed on to other people through the natural spread of the virus.
- **Recovered or Removed** A patient who has acquired natural immunity or has died is no longer able to participate in the replication process. In both cases, the virus cannot spread any further.
- **Incubation period** In the early stages of infection, there may not be any signs of infection; this initial stage is called the incubation period. It is at the intersection of this period and the infectious period that viruses spread most. [1]
- **Symptomatic period** This is the stage of the infection where signs of infection are visible, for example, an infected person goes to a doctor.

### *2.2. Basic models*

Most of the work in modeling infectious disease epidemics is mathematically inspired and based on differential equations and SIR model.

SIR modeling relies on the assumption of constant population and neglects the spatial effects. Constant population means, that there are no births and migrations. These models often fail to consider individual contact or interaction processes and assume that populations are homogeneously mixed and do not include variable susceptibility.

The simplest aggregate model, commonly known as the SIR model, a population of size $N$ is divided into three states: susceptible ($S$), infective ($I$), and removed or recovered ($R$).

The following discrete time process describes the system dynamics: each infected person can infect any susceptible person (independently) with probability $\beta$ or can recover with probability $\gamma$.

Let $S(t)$, $I(t)$ and $R(t)$ denote the number of people who are susceptible, infected and recovered states at time $t$, respectively [4]. Let

$$s(t) = \frac{S(t)}{N}; \quad i(t) = \frac{I(t)}{N}; \quad r(t) = \frac{R(t)}{N}$$

The "complete mixing" assumption means that each individual is in contact with everyone in the population, it can be shown that the following system of differential equations (known as the SIR model) describes the dynamics:

$$\frac{ds(t)}{dt} = -\beta \cdot s(t)i(t); \quad \frac{du(t)}{dt} = \beta \cdot s(t) \cdot i(t) - \gamma \cdot i(t); \quad \frac{dr(t)}{dt} = \gamma \cdot i(t),$$

## 3. An overview of Cellular Automata

Cellular Automata is a dynamic system characterized by discretization of time and space. Usually it consists of a regular grid of cells connected to their neighbors and each cell has finite space of states. The grid can be of any dimension, in this article we assume it to be two dimensional.

### *3.1. The cell*

The basic component of Cellular Automata is the Cell. Each cell is a Finite State Automaton that evolves according to some update rules. Each state of cell depends from its previous states and the states if its neighbors. Classically cells form into a lattice, being placed side by side, but there are no limitations and other techniques are allowed. Cells do not need to identical and each cell can be different from others.

### *3.2. Update rules*

Cellular Automaton state is defined by a set of the current states of all its cells. The states of each cell are governed by a predefined update rules. The transition between current state and the next state for each cell is a function of the current state of that cell and current states of its neighbors.

### *3.3. Interaction with neighbors*

As stated above, the state of each cell depends from its neighbors. There are different possibilities of defining the neighborhood the a cell. The most common ones are 4-connected Von Neumann neighborhood and 8-connected Moore neighborhood.

## 4. Data

Data used for the development of this model consists of generally available data and data provided by different governmental institutions and includes:

- Data on infectivity parameters and disease course parameters (virus biology);
- Data defining the population;
- Data on the vicinity of regions / geographic areas;
- Data identifying the initial (current) state of the epidemic;
- Daily infection history with division to districts.

- Additional data influencing social contacts, like closing schools, shops;
- Additional data influencing the infectivity and course of the disease like wearing a mask, etc..

## 5. Model

The proposed model assumes the structure of a graph of mutually connected Cellular Automata, where the nodes represent Cellular Automata that model some region and the edges represent geographical connections between those regions. The edges facilitate the population flows between nodes.

Each Cellular Automata that models some region is characterized by the size, population and additional parameters adjusted to the region that it simulates. The way the parameters are adjusted will be discussed in section on training the model.

The modeled area is usually a two-dimensional grid of size $n \times n$. It consists of cells that are linked to other cells. Depending on the simulation approach, we allow connections to the nearest 8 neighbors or in the form of 8 neighbors for a radius of size 1 or more neighbors for larger neighborhood radius.

### *5.1. Disease progression*

In proposed model we try to reflect the biology of the virus and the way the host changes his states of infection.

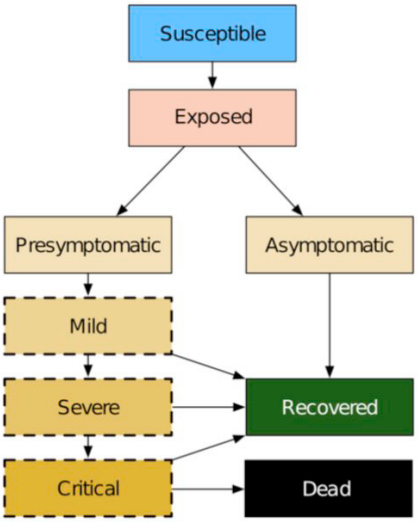



Fig. 2: Possible states and transitions of an agent. Yellow shading indicates that an individual is infectious, states with dashed border are considered symptomatic. [5]

As shown on Fig. 2 a person or an Agent, we will call them interchangeably, at the beginning is Susceptible. After a contact with an infected host he becomes exposed and if gets infected, he changes his state to Presymptomatic or Asymptomatic. Both of the states indicate that an agent is infectious, but the timeline of being infectious or the virus is still in incubation phase is described in the Fig. 1.

Each of the states has some defined duration range and probability of changing to other state. The durations and probabilities are taken based on current literature [6], [7], [8].

An agent at the end becomes Dead, which means he is removed from the simulation or Recovered. In both cases he is unable to be infected again. In our research we do not take into consideration the loss of immunity of an agent.

### 5.2. Simulated world and Agents

In selected modeling approaches, it is possible to model only one region with cellular automata, and it is also possible to model the world as an interconnected regions with population flows between these regions.

In our case of Poland, we created a structure where the root region is the whole country, the level deeper are voivodeships, currently 16, and for each voivodeship we divide them into districts, with 380 total for Poland. Districts are organized into a graph, where districts are the nodes of this graph and edges are signalizing the geographical neighborhood between districts.

The modeled world consists of Cellular Automatons for each region, where the simulated region is a district. Each district consists of cells, that represent basic units in the Cellular Automata. Each cell can be of state empty or occupied by an agent.

An agent is a logical unit that is used for tracking the progression of the disease. Each agent can either simulate one person or a small sub-population.

The structural characterization of a person consists of attributes (also dependent on a specific model):

- Age
- Sex;
- Current state of the patient from the states set $\{S, L, I, R, D\}$;
- Is wearing a mask.

For the simulated region, each Cellular Automata can be described with following properties.

- Total population size;
- Population size for each of the states from set $\{S, L, I, R, D\}$;
- The gender distribution;
- The age distribution.
- Spatial distribution of population wearing a mask.

Creating the world in the following experiments consists in collecting data on:

- Population size for each district;
- Population gender distribution;
- Population age distribution;
- Distribution of children attending schools broken down by age;
- Number of children attending kindergartens;
- Number of students in higher education institutions;
- Built-up area in each district;
- Total area of each district.

In addition, the model parameters that characterize virus biology, i.e.:

- Total infection time since infection;
- Virus incubation time;
- Virus contagiousness period (how long an individual infects others);
- Duration of mild symptoms;
- Duration of severe symptoms;
- Duration of critical symptoms;
- Probability of getting infected without mask;
- Probability of contracting infection while wearing the mask;
- Likelihood of symptoms occurring;
- Probability of severe symptoms;

- The probability of a critical symptoms;
- Probability of death.

### 5.3. Running the simulation

Before running the simulation for a selected starting date and time duration (number of days) we need to initialize the model. Once the world model is defined, and populations of agents are initialized based on the data provided, the system is ready to run the simulation.

Once the world is initialized, we randomly select the region order and for each of them we simulate single day inside each of them. For each simulated region we run the logic for each agent and move them inside that region. Once the simulated day is over, we move some percentage of agents between neighboring regions, to simulate the flow of people.

Each agent changes his state once infected. The state changes once a day and the logic of that change is described below.

- The state of a cell changes from susceptible $S$ to latent $L$, once it gets into contact with an infected cell in its specific neighborhood. A cell acquires disease from infected neighbors with the probability determined by the infectivity parameter $\beta$. The cell remains latent for the number of time (update) steps as defined by the delay parameter $\lambda$.
- The state of the cell changes from latent $L$ to infectious $I$ after it transfers into state $L$ for given $\lambda$. In this model, for the sake of simplicity, we assume that any cell exposed to the pathogen will become infectious. In the state $I$, cells are capable of transmitting infection to neighboring cells. For example, if for disease $D$, $\lambda = 2$ units, then after two steps the cell will enter the contagious state $I$.
- After a period of time defined by the infection period $\theta$, the cell status changes from infectious $I$ to recovered or deleted $R$. Once cells have reached the state $R$, they become no longer capable of transmitting infection.
- From the state $R$, the cell returns to susceptible $S$ or remains in the state $R$, the latter meaning its complete immunity. The enabled "treatment mode" determines the transition from state $R$ to $S$ or vice versa.

### 5.4. Training the model

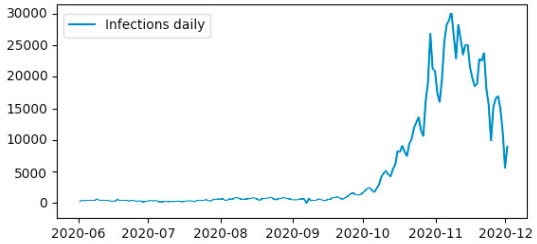
The goal of training is to select the best parameters for the model in order to minimize the cost function (1). Based on the experiments, the loss function MAE, defined as L1, is selected (1). Due to its properties and the specificity of cellular automata, it is more resistant to the *outliers*.

The specificity of a Cellular Automata relies on the fact, that this is a model that performs a simulation, where at different steps it could achieve infection counts of a high amplitude, with *outliers* from the mean of infection counts over a 7 day window. We are interested in minimizing the difference between predicted infected people count in simulated epidemic and real world data.

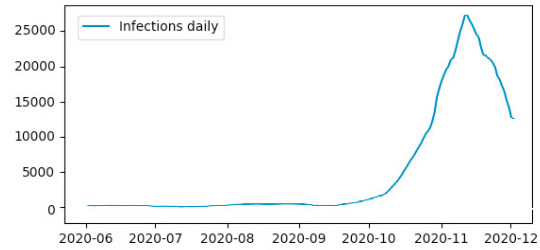
$$\mathrm{MAE} \stackrel{def}{=} \frac{1}{n} \sum_{t=1}^{n} |y_i - y_i^p| \tag{1}$$

The parameters selected for the Cellular Automata for each region are:

1. The size of the grid per simulated agent;
2. Radius counted in the cells of the unit's travel range (this parameter simulates the unit's travel range);
3. The number of movements of the unit during the simulated day (this parameter simulates the mobility of a society);
4. Infection range;
5. Probability of being moved (will the agent move or stay home).




(a) Daily infections used to training data without averaging




(b) Daily infections used to training data using averaging over 7 day window

Fig. 3: Comparison of training data with and without averaging

These parameters influence the size of the grid for each Cellular Automata, the level of concentration, mobility of agents and potential range of infection (which could be number of randomly met agents).

#### *5.4.1. Selection of parameters*

The process of selecting parameters for proposed model is carried out in two ways. Both are based on searching the space of possible parameters, the parameters are selected to minimize the cost function (1).

The selection of parameters is carried out for each district separately, because each district is characterized by a different size, population and the course of the pandemic.

The selection of parameters is carried out on the basis of two methods:

1. Grid Search - searching among all possible combinations of parameters;
2. Randomized Search - randomly searching selected $k$ parameters by randomizing them from the space of all possible combinations;

#### *5.4.2. Training phase*

Training takes place on the basis of 7-day averaged data in order to minimize the decline in the incidence on weekends and on Mondays.

The parameter selection phase consists in downloading the start date of the simulation, then 12 days are selected (it can be parameterized for any number of days) before the simulation start date and then the curve of the simulation for these 12 days is compared with the real data. Next for each set of possible parameters the minimum MAE score is score is selected and the optimal parameters are saved for each district.

## 6. Experiments

Data showed in figures are training data with real daily infection counts for Poland in year 2020. In Figure 3 we show the need for data preparation for the model. We are using 7-day average for smoothing the curve of infections, to remove weekend drops in tested positive cases of COVID-19.

### *6.1. Runs without and with adjusted parameters*

In this section we would like to show the influence of adjusting parameters, as described in section 5.5, on the model predictions. As can be observed in Figures 4 and Figure 5, there is a significant difference between the results for the model with trained parameters and the one without training. With adjusted parameters, the model reflected the real observed dynamics of the virus for the Masovian voivodeship. The results for voivodeships are aggregated results from simulated districtrics.

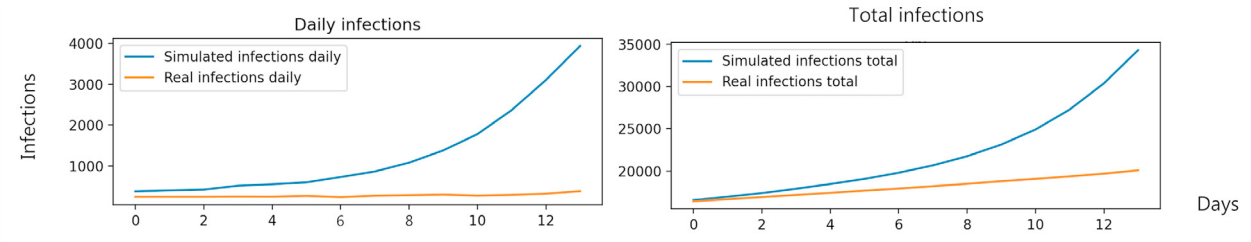



Fig. 4: Predicted COVID-19 daily infections for 14 days simulation, starting from 1-10-2020 without adjusting model parameters for Masovian voivodeship

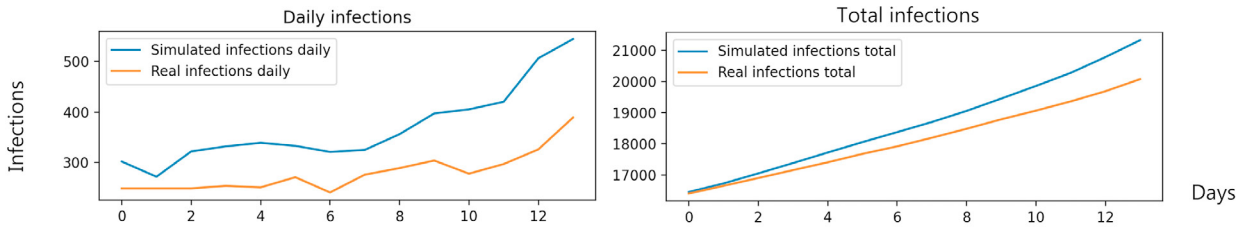



Fig. 5: Predicted COVID-19 daily infections for 14 days simulation, starting from 1-10-20 with adjusted model parameters for Masovian voivodeship

### 6.2. *Results for trained model for Poland*

In this section we show cumulative results of predicted COVID cases for Poland over different periods of simulations. The start date was chosen to show the beginning of the second wave of infections in Poland and to verify if the model was able to predict that effect. As can be observed the model has correctly reproduced the basic epidemic dynamics of the disease and was able to predict the increase in infections. For 14 and 20 days simulation (Figure 6 and Figure 7 respectively) it reflected the real pandemic and for 30 days (Figure 8) it showed properly what would have happen after day 20 if no preventive measures were taken.

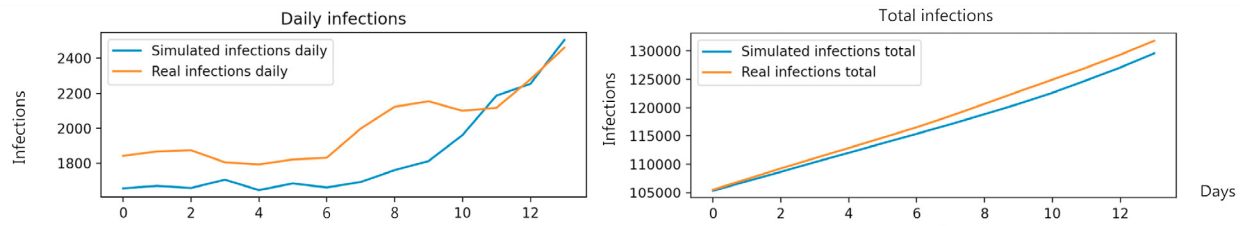



Fig. 6: Predicted COVID-19 infections for Poland, starting from day 10-10-2020, duration of the simulation is 14 days

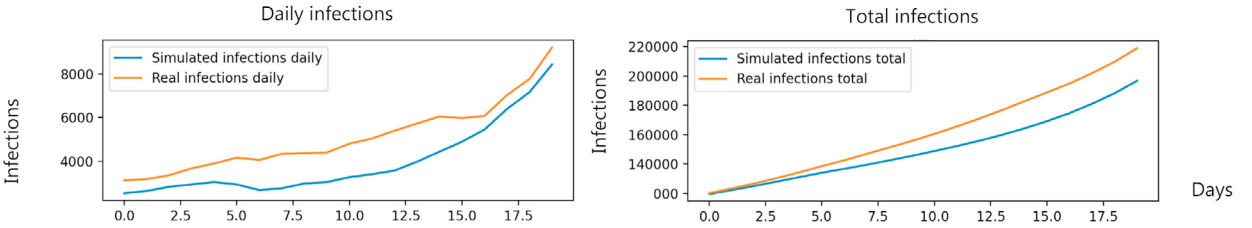



Fig. 7: Predicted COVID-19 infections for Poland, starting from day 10-10-2020, duration of the simulation is 20 days

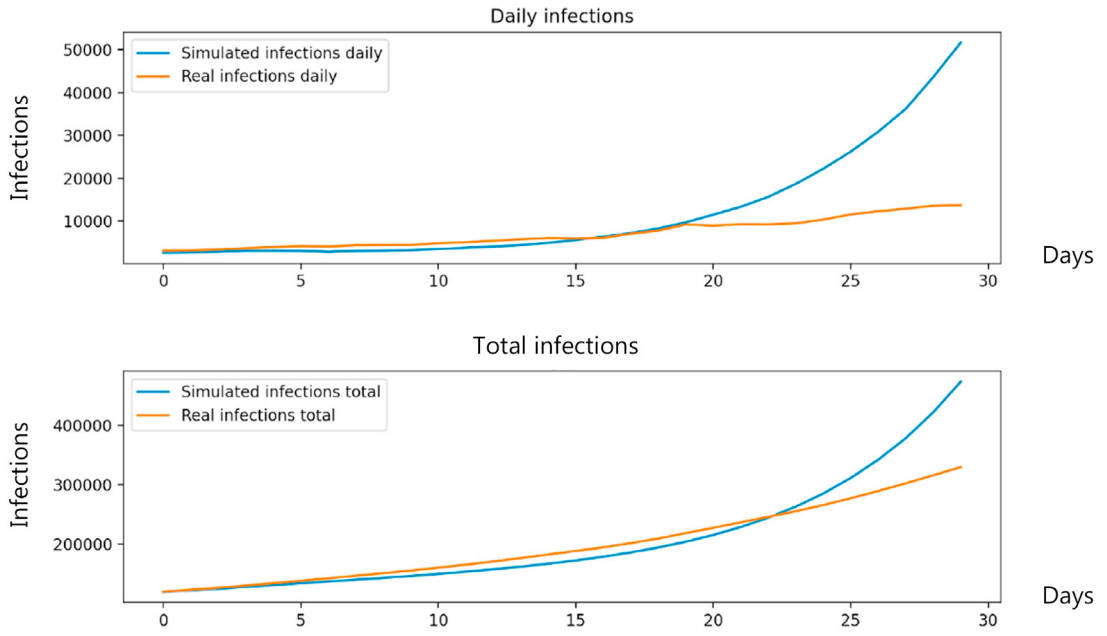



Fig. 8: Predicted COVID-19 infections for Poland, starting from day 10-10-2020, duration of the simulation is 30 days

## 7. Conclusions and future work

We showed the applied potential of the proposed model and how it can learn and simulate the space-dependent spread of the COVID-19 with adaptive parameters. The model proved to able to learn and reproduce the epidemic dynamics in most cases provided sufficiently accurate data were available.

In the case of districts with a low number of infections, there were difficulties in modeling the dynamics of the pandemic. The total errors for such districts also significantly affected the aggregated overall simulation result.

The model showed higher accuracy of predictions for the infection growth curve for districts with a greater number of infections.

In the future model development, the occurring inaccuracies are to be reduced by using more sophisticated machine learning methods, methods of scaling districts, all this certainly subject to the availability of more precise and comprehensive data.

For future work there is also a possibility to extend our model to different compartmental models like SEIQR, $SEI^3RD$ or $SEI^3Q^3RD$ [9].

Another possibility for future work is to develop scenarios of different preventive measures and vaccination and how they influence the epidemic.

## Acknowledgment

The research has been supported by the Polish Ministry of Science and Higher Education under project ProME "Predictive Modeling of the Covid-19 Epidemics". We would like to thank the reviewers for their valuable and constructive comments.

## References

[1] Scott, S., and Duncan, C. J. *Biology of Plagues: Evidence from Historical Populations*. Cambridge University Press, 2001.

[2] Fu, Shih and Milne, George *Modelling Epidemic Spread using Cellular Automata* ACAL2003: The First Australian Conference on Artificial Life; Canberra, Australia, 2003
[3] Andrewes, C. H. *The Natural History of Viruses*. W. W. Norton and Company, 1967.
[4] Fred Brauer, Carlos Castillo-Chavez, Zhilan Feng *Mathematical Models in Epidemiology*. Springer, 2019
[5] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry1, Romesh G. Abeysuriya, Gregory Hart, Katherine Rosenfeld, Prashanth Selvaraj, Rafael C. Núñez, Brittany Hagedorn, Lauren George, Amanda Izzo, Anna Palmer, Dominic Delport, Carrie Bennette, Bradley Wagner, Stewart Chang, Jamie A. Cohen, Jasmina Panovska-Griffiths, Michał Jastrzebski, Assaf P. Oron, Edward Wenger, Michael Famulare, and Daniel J. Klein *Covasim: an agent-based model of COVID-19 dynamics and interventions*. medRxiv, 2020
https://www.medrxiv.org/content/10.1101/2020.05.10.20097469v1
[6] Caitlin Daley, Megan Fydenkevez, Shari Ackerman-Morris *A Systematic Review of the Incubation Period of SARS-CoV-2: The Effects of Age, Biological Sex, and Location on Incubation Period* medRxiv, 2020
https://www.medrxiv.org/content/10.1101/2020.12.23.20248790v1
[7] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J Counotte, Stefanie Hossmann, Hira Imeri , Aziz Mert Ipekci , Georgia Salanti , Nicola Low *Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis* PLoS Med. 2020 Sep
https://pubmed.ncbi.nlm.nih.gov/32960881/
[8] Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group; Rosalind M Eggo *Age-dependent effects in the transmission and control of COVID-19 epidemics* Nat Med. 2020 Aug
https://pubmed.ncbi.nlm.nih.gov/32546824/
[9] Veronika Grimm, Friederike Mengel and Martin Schmidt *Extensions of the SEIR model for the analysis of tailored social distancing and tracing approaches to cope with COVID-19* Nature, 2021
https://www.nature.com/articles/s41598-021-83540-2