

Monte Carlo Method - protocol

Simulating the spread of COVID-19 using the Monte Carlo method

Lukáš Čejka

January 5, 2022

Contents

1	Introduction	1
2	Theory	2
2.1	Coronavirus	2
2.2	Measured statistics	3
2.3	Simulation methodology	4
3	Implementation	5
3.1	Simulation setup	5
3.1.1	Constants	5
3.1.2	Random processes	5
3.2	Code examples	6
4	Results	8
5	Conclusion	10

1 Introduction

Whenever a new disease or virus emerges, mankind strives to eliminate, or at the very least, mitigate the negative effect it can have on a human being's health. However, another endeavour is sought after in parallel: minimize the spread. In order to lessen the spread of a virus, one must first understand how it spreads and the speed of reproduction - in the words of the *World Health Organization*: "*The best way to prevent and slow down transmission is to be well informed about the disease and how the virus spreads.*" [1]. The former can be done via medical analysis, however, the latter relies majorly on statistical data to determine the rate of spread. Concurrently, the collection of data requires time, even more so when it needs to be statistically significant due to the potential lives at stake. While some initial data may give an embryonic estimate, further analysis can aid in strengthening this understanding. One of many possible ways to analyse such a problem is to visualize it by performing simulations that aim to replicate the spread of a virus based on a model obtained from real-world data. Recently, Coronavirus has been on the rise with no apparent end in sight, therefore being able to simulate and model the spread of this virus is becoming a discipline of its own when strengthening knowledge around the topic.

This protocol aims to replicate the results from *Novel approach for Monte Carlo simulation of the new COVID-19 spread dynamics* [2], specifically, figures 3 and 4 on page 4 that describe daily new cases and the effective

reproductive number in a city, respectively. Similarly to the mentioned paper, MATLAB will be used to implement the simulation. First, section 2 will present a more detailed introduction to the problem and the theory behind the simulation. Then, in section 3, the implementation will be described. Last but not least, section 4 will contain results obtained from the implementation and their comparison to that of the aforementioned paper. Finally, section 5 will encapsulate the summary of this protocol and any further comments.

2 Theory

In this section, the theory behind the Coronavirus and the simulation methodology will be presented in greater detail.

2.1 Coronavirus

Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus [1]. The name "*Coronavirus*" references reference the similarity of the virus's shape under an electron microscope to that of the solar corona. The virus was discovered towards the end of 2019 in Wuhan, China [3], however, it is unclear what species the virus originated from with the many scientists suspecting bats. Moreover the means by which the virus spread from animals to humans is a widely debated topic when it comes to specifics.

Symptoms and effects Upon being infected with the virus, the majority of people will experience a mild to moderate respiratory illness - akin to pneumonia - and will recover without the need for extensive medical care. On the other hand, some people may require special treatment due a more serious extent of their illness. The symptoms experienced during the progression of the disease are only a part of the problem since many individuals report medical conditions after recovery, for example, limited lung capacity, damaged organs, etc. [4] Among the at-most risk are older individuals, as their underlying conditions (cardiovascular diseases, diabetes, etc.) tend to add to the progression of the disease. Ultimately however, there is no one age group that is safe from becoming seriously ill [1].

Virus transmission An infectious human can spread the virus via minute liquid particles produced when the person coughs, sneezes, speaks, breathes, etc. - basically, whenever air comes out of a person's mouth and/or nose. The size of one particle ranges from droplets (grains of sand) to much smaller and harder-to-detect aerosols [1]. Furthermore, once these particles are expelled from an infectious human, they can stay active on surfaces for a number of days. For example, in an indoor environment the virus can survive on non-porous materials for up to 3 days; research differs greatly and the duration depends on the type of surface material, surrounding conditions, viral load etc. - for more information see *Science Brief: SARS-CoV-2 and Surface (Fomite) Transmission for Indoor Community Environments* [5]. The means by which a human can get infected is by absorbing the virus particles via mucous membranes on the eyes, nose, or mouth - this includes inhaling them directly from surrounding air or nearby skin when the human's hand touches a surface with an active virus and then is brought near the aforementioned membranes. The fact that the virus survives on surfaces for prolonged periods of time combined with human nature to subconsciously touch our face and socialize in one way or another results in multiple and simple means of transmission.

Sample case Let us consider a sample case where a susceptible human being comes into contact with an infectious human being. First, the susceptible subject is moving around and travelling, or commuting to work, school, shops, etc. In the midst of these commutes the subject can come into contact with an infectious human being. From this point, the subject may or may not develop the disease depending on a number of factors, such as duration of exposure, viral load, immunity strength, etc. Assuming that the subject will eventually developed the disease, it will go through a so-called *incubation period* which lasts until the subject develops symptoms. The most infectious

period begins roughly 1-3 days before symptom onset and generally ends 7 days after symptoms appear. Following the infectious period, the subject may experience some symptoms up to 21 days since the exposure - not counting long-lasting post-illness medical conditions. This description is shown in figure 1 [2].

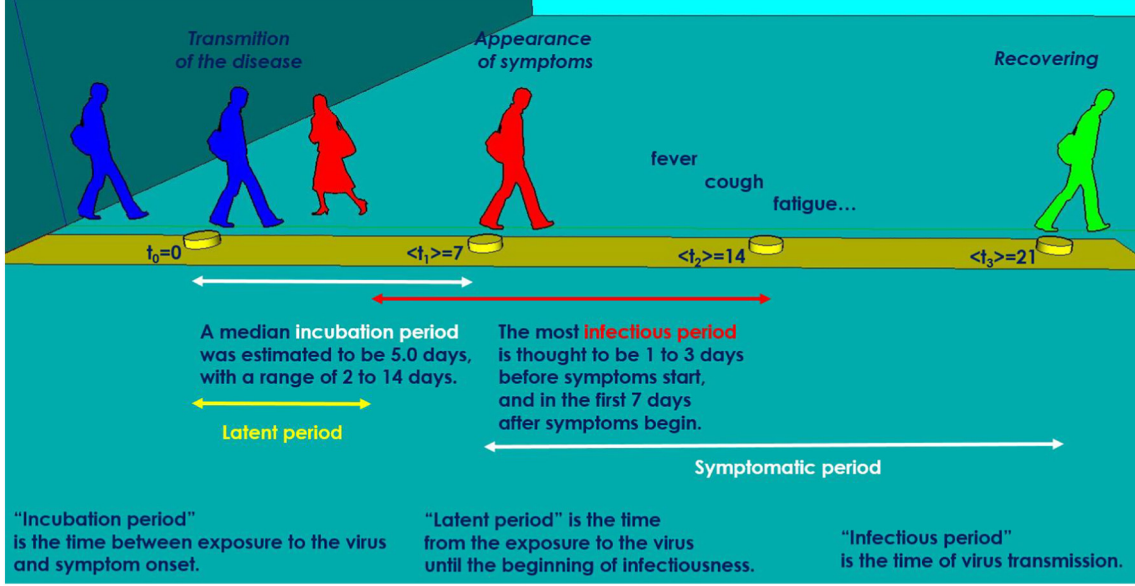


Figure 1: Abstract design showing various time intervals of the virus transmission from an infectious human to another human. The time moments represent mean values. Source: *Novel approach for Monte Carlo simulation of the new COVID-19 spread dynamics* [2]

2.2 Measured statistics

The aim of this protocol is reproduce the following statistics from the referenced paper [2]: Daily New Cases (DNC) and the effective reproductive number R_e .

DNC DNC is simply the number of newly infected human beings on a given day.

Effective reproductive number The effective reproductive number is one of two numbers associated with determining disease spread speed in epidemic theory. It is defined as *the average number of secondary cases per infectious case in a population made up of both susceptible and non-susceptible hosts* [6]. If $R_e > 1$, then the number of cases will increase - start of epidemic. If $R_e = 1$, then the disease is endemic, which in the instance of Coronavirus means that it is not spreading. If $R_e < 1$ then there will a decline in the number of cases [6]. While there is no one single formula defining this number, the authors of the reference paper [2] used the definition listed in equation (1)

$$R_e = 1 + \ln \left(\frac{I_{t_i}}{I_{t_{i-1}}} \right)^{\frac{1}{c}} \quad (1)$$

where I_{t_i} represents the number of infectious cases at time t_i and c can be considered a constant that differs slightly from 1, however, in practice it does not have a significant impact on the result, therefore for the purpose of this protocol it was set as 1.

2.3 Simulation methodology

The simulation of the spread of Coronavirus in this protocol is built on a **SIR**-based model which is often used in mathematical modelling of infectious diseases as it divides a specific population of humans (e.g. a city, country, or just a group of humans) into labelled compartments:

- **Susceptible** - humans that are susceptible to the infection.
- **Infectious** - humans that are currently infectious and able to spread the disease.
- **Recovered/Removed** - humans that have either recovered or died (removed from simulation) and therefore cannot be infected and cannot infect again.

When it comes to simulating the behaviour described in section 2.1, it is paramount to include only processes that have an impact on the results. The aim of this protocol is to reproduce the Daily New Cases (DNC) and effective reproductive number R_e , as was stated in the introduction. For both of the monitored statistics, there are certain aspects of the sample case that need not be accounted for, e.g., dividing time intervals related to the health status of the human subject. The only health status needed for the purpose of simulating the spread of the virus is when the human is infectious, as it is during this period that the subject can infect susceptible humans. Therefore the only remaining aspects that ought be an integral part of the simulation are:

- **Daily movement** - 1st part of the sample case. The human subject travels and commutes (random process).
- **Physical distance** - 2nd part of the sample case. The human subject is exposed to an infectious human being (random process).
- **Infectious period** - 3rd part of the sample case. The human subject is sick for a specific time period (random process).

Daily movement During their daily lives, human beings majorly follow a predictable routine on an individual level. However, when trying to generalize the movement for an average human, there is simply no other way than to create some sort of random process. In the case of daily movement, there are two random processes to consider: direction and length of step. The former will follow a uniform distribution on a unit circle and the latter a Gaussian distribution. Thus, for the purpose of this simulation, every human subject will move in a random direction with a random step length. However, it needs to be mentioned that the reference paper [2] only specifies the random distribution of the step length, i.e., it does not mention what distribution the direction of daily movement should follow.

Physical distance During their daily movement, human beings can come into contact with other human beings, some of whom may be infectious (this simulation does not consider quarantining human beings). As mentioned above, in real life multiple variables contribute to the successful transmission of a virus. The simulation of each combination was not considered in the reference paper [2], therefore, they will be aggregated into physical distance. Let there be two human beings. One of whom is infectious and the other susceptible. Thus, the probability of the virus being transmitted from the infectious human to the susceptible human is solely based on the physical distance between the two humans. As in the reference paper [2], this probability will be based on the Exponential distribution - see equation (2).

$$f(x, \lambda) = \lambda e^{-\lambda x} \quad (2)$$

where $x > 0$ represents distance and λ is the rate parameter.

Specifically, the probability of infection can be obtained by integrating Exponential distribution's PDF from the distance between two individuals d to infinity:

$$\int_d^{+\infty} \lambda e^{-\lambda x} dx = e^{-\lambda d} \quad (3)$$

Infectious time Following the exposure of a susceptible human subject to the virus and a subsequent transmission of the virus, the subject becomes infectious. Since every human being is different when it comes to immunity strength and other factors that dictate how long the subject will be infectious for, it is assumed that the duration of the subject’s infectiousness is a random variable from the Gamma distribution. For the purpose of keeping to discrete time units, the value obtained from the continuous Gamma distribution is rounded to up to the nearest integer. See equation (4) for the PDF of the Gamma distribution

$$f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (4)$$

where $x > 0$ is a random variable, α is the shape parameter and β is the rate parameter.

3 Implementation

In this section, the simulation setup will be described along with examples of code.

3.1 Simulation setup

The simulation setup of the SIR model can be divided into two parts: constants and random processes.

3.1.1 Constants

The constants of the simulation revolve mainly around the city. In the reference paper [2], the authors specify that their model aims to simulate the spread of the Coronavirus in a city with a given population density. However, there are additional constants - indirectly mentioned in the paper - that need to be accounted for, e.g., the number of days the simulation should run, whether humans are allowed to escape the city bounds and the distribution of humans’ starting positions within the city. Table 1 specifies the simulation configuration for this protocol.

Table 1: Properties of the simulation configuration and their values, along with a column specifying whether the property is mentioned in the reference paper [2].

Property name	Property value [units]	Mentioned in reference paper
City dimensions	1000 × 1000 [m]	Indirectly
City population	2000 [humans]	Indirectly
Population density	2000 [humans/km ²]	Yes
Initial infectious population	0.5 [%]	Yes
Length of simulation	100 [days]	Indirectly
Humans constrained in city	Yes	No

In summary, the modelled city in the shape of a 1000 × 1000 m (1 km²) square will house a population of 2000 people of which 10 (0.5%) are infectious since the genesis of the simulation. By default, humans are constricted to the city bounds and the spread of the virus will be simulated over the course of 100 days. For reference, larger cities (e.g., Paris and Mumbai) have a population density of up to 20 000 people per km² [7][8].

3.1.2 Random processes

In this part all random processes explained in subsection 2.3 will be specified - specifically the parameters of their respective distributions that were selected by the authors of the reference paper [2] as they present characteristics consistent with their respective properties when it comes to medical observations.

Daily movement The daily movement of a human being in one day of the simulation utilizes two random variables: angle (direction) of movement and the length of the step. The angle follows a uniform distribution on a unit circle, in other words, all directions have the same probability of being selected, in theory. The step length (length of movement) follows the Gaussian (Normal) distribution (allowing negative distances) with a mean step equal to $1/48$ of the city dimensions and a standard deviation equal to $1/144$ of the city dimensions. These values were chosen differently than specified in the reference paper as they best recreate the results in the reference paper [2]. If the values specified in the reference paper [2] are followed exactly, then a human being would either have a mean step equal to 2 m a day or 250 m a day in a 1000×1000 m city - depending on the interpretation (see the 1st bullet point in section 2. *Simulation methodology* of the reference paper [2]).

In the implementation accompanied by this protocol, the new location of a human is generated with respect the human's current position using Polar coordinates and then converted to Cartesian coordinates. If humans are constricted to the city bounds and a coordinate violates these bounds, then the coordinate is set to be the value of the corresponding overstepped bounds.

Physical distance The probability of infection given a distance between and infectious human being and a susceptible human being follows the Exponential distribution. There are two parameters listed in equation (2): λ and x . The mean value is equal to $1/\lambda$ which in this simulation is equal to 8 m. The distance between humans x is assumed positive. For example, given two humans - one infectious and one susceptible - that are 2 m apart, the probability of the susceptible human being infected is equal to (according to equation (3))

$$e^{-\frac{2}{8}} = 0.7788 \quad (5)$$

in other words, 78%.

Infectious period The infectious period of a human being is a random process following the Gamma distribution, with a mean value of $\mu = \alpha/\beta$ and the variance $\sigma^2 = \alpha/\beta^2$. Similarly to the reference paper [2], $\alpha = 6$ (dimensionless) and $\beta = 2/3 \text{ days}^{-1}$ making the mean infectious period 9 days and the standard deviation equal to 4.9 days.

3.2 Code examples

The simulation was implemented purely in MATLAB as it already houses many functionalities required for this project along with simple visual output of results. Even though MATLAB benefits heavily from vectorization, the simulation class structure was designed with an object-oriented approach using handle classes to assure ease of readability and serve as practice when it comes to software development. The class structure can be seen in figure 2. It can be seen that the main driving class is `simulation`, which houses the `city` class where the simulation will take place, the `configuration` class where the simulation configuration is stored and finally, the `data` class where all data related to the simulation is stored, processed and outputted.

An example of the process of determining newly infected humans on a given day can be seen in code blocks 1, 2, 3, 4, 5.

Listing 1: Function for simulating one day, where `obj` is `city`.

```

1 function simulate_day(obj, day)
2     obj.population.update_health_status(day);
3     obj.population.update_humans(day);
4
5     for human = obj.population.humans_by_status.all
6         human.move_to_random_position()
7     end
8 end

```

Listing 2: Function for updating the health status of the population on a given day, where `obj` is `population`.

```
1 function update_health_status(obj, day)
2     obj.update_infectious(day);
3     obj.update_recovered(day);
4 end
```

Listing 3: Function for updating infectious humans, i.e., deciding which susceptible humans will every infectious humans infect, on a given day, where `obj` is `population`.

```
1 function update_infectious(obj, day)
2     for human = obj.humans_by_status.infectious
3         human.infect_susceptible_humans(day, obj.humans_by_status.susceptible);
4     end
5 end
```

Listing 4: Function for infecting susceptible humans by an infectious human on a given day, where `obj` is `human` - specifically, an instance of `human` whose health status is infectious.

```
1 function infect_susceptible_humans(obj, day, susceptible_humans)
2     for susceptible_human = susceptible_humans
3         infection_prob = obj.get_infection_prob_of_human(susceptible_human);
4
5         if rand() <= infection_prob
6             susceptible_human.set_health_status('infectious', day=day);
7         end
8     end
9 end
```

Listing 5: Function for getting the probability of infection between an infectious human and a susceptible human, where `obj` is `human` - specifically, an instance of `human` whose health status is infectious.

```
1 function prob = get_infection_prob_of_human(obj, human)
2     distance = obj.get_eucledian_distance_from_human(human);
3     prob = round(exp(-distance/obj.infection.mean_distance), 2);
4 end
```

Alternatively, the means by which a successful infection is decided can be done using the logic in listing 6:

Listing 6: Alternative means of infection that better reflects the use of the exponential distribution.

```
1 % Create the exponential distribution (initially)
2 pdcrucial = makedist('Exponential','mu',dist);
3 ...
4 % Generate a random number (inside the loop of physical distance checking)
5 distr = random(pdcrucial);
6 ...
7 % Check if infection occurred
8 if d < distr
9     susceptible_human.set_health_status('infectious', day=day);
10 end
```

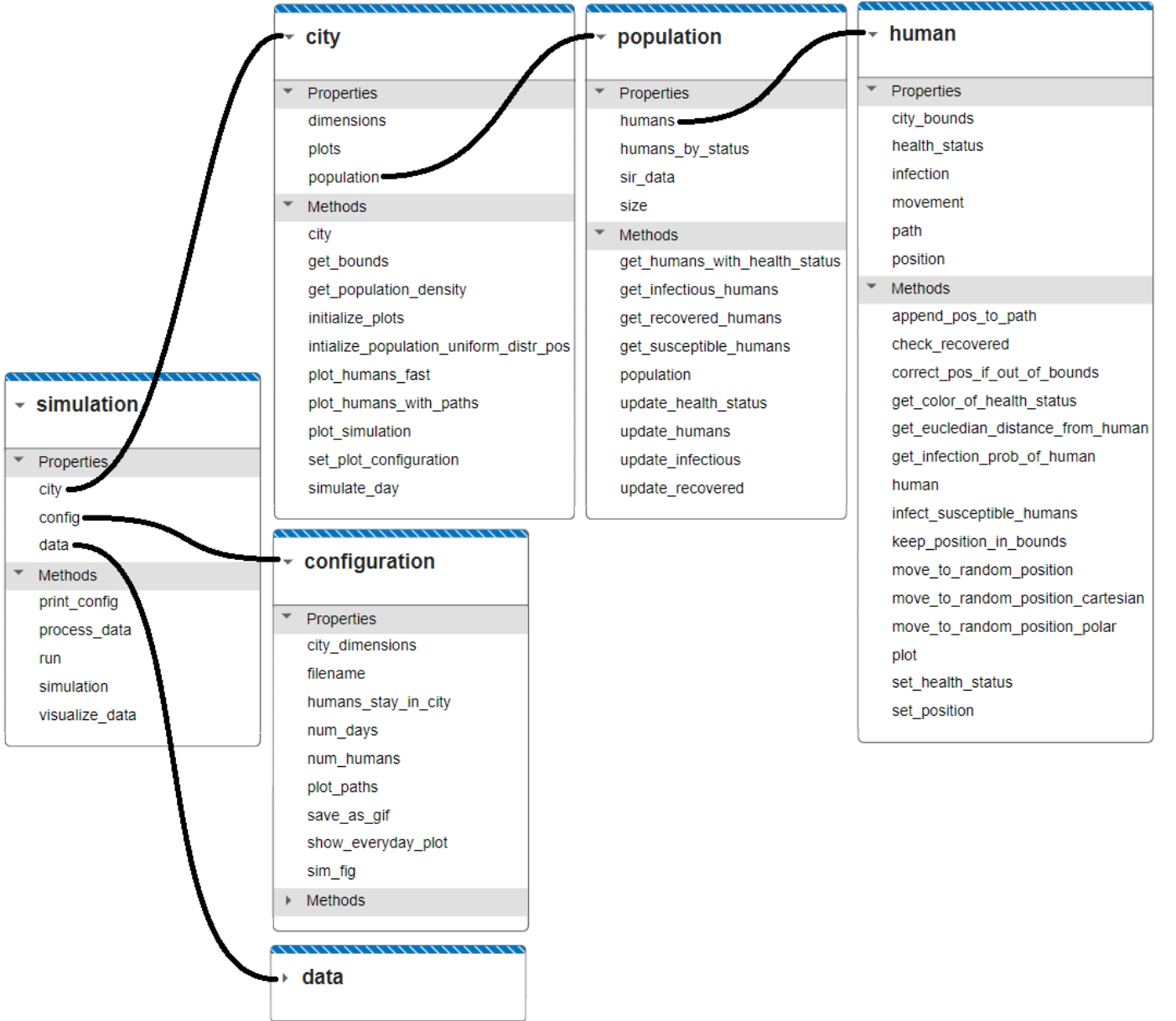


Figure 2: Class diagram for the simulation, where black lines represent connections between class properties and classes. For example, the `simulation` class contains properties: `city`, `config` and `data`, all of which represent instances of classes `city`, `configuration` and `data`, respectively.

4 Results

Results obtained from the simulation can be seen in figure 3 and the results from the reference paper [2] can be found in figure 4.

When comparing DNC results, it can be seen that the trend is comparable with both simulations exhibiting a sharp rise in cases from the beginning until 15-20 days into the simulation. Moreover, the gradual lessening of daily new cases eventually becomes zero about 50-60 days into the simulation - with minor deviations.

When comparing results of the effective reproductive number, we can see that there are error bars present in figure 4b, however, they appear to be missing in figure 3b. This is due to the fact that the authors in the paper do not clearly explain what these error bars represent - reproductive efforts for this indicator failed. Nevertheless, data obtained from the simulation seems to correspond to the trend obtained from the reference paper [2] with

some minor differences. The main difference stems in the minimum and maximum values, which in the case of the reference paper [2] almost seem to have been calculated using a different formula than what the one found in equation (1). The main indicator for this reasoning is that the initial values for R_e reach a maximum of 3, which would mean that the number of cases should have tripled in the first 5-6 days, which according to figure 4a does not seem to be the case. For reference, in order achieve the same initial R_e , the mean infection distance needs to be tripled in the simulation. Nonetheless, the overall rising and declining characteristics of the effective reproductive number are pertinent between the two simulations.

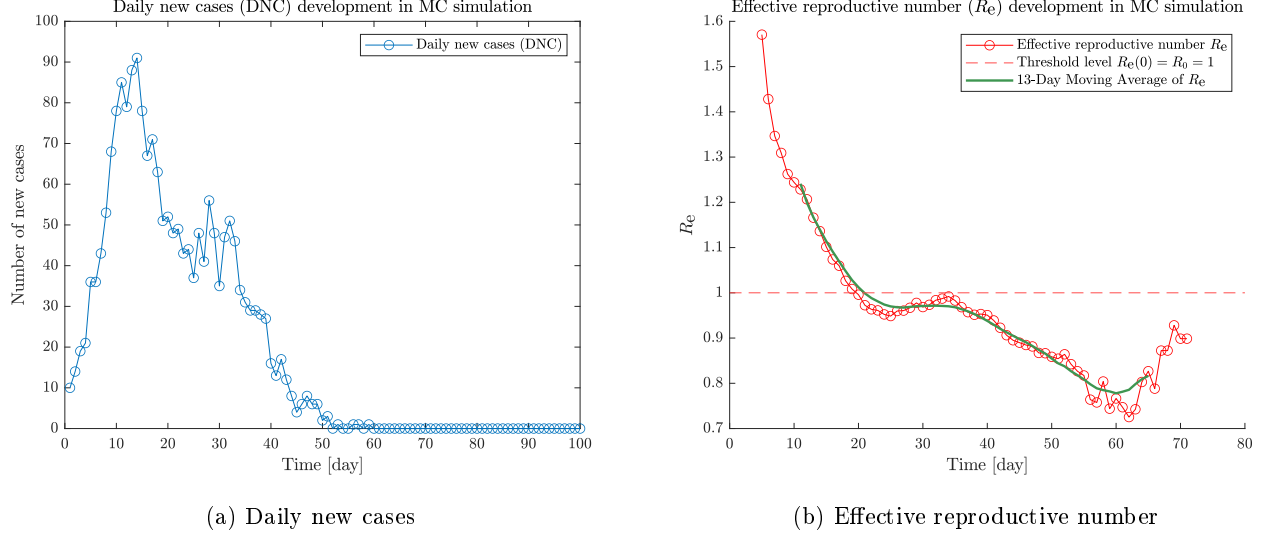


Figure 3: Graphs depicting the daily new cases and the effective reproductive number obtained from the simulation.

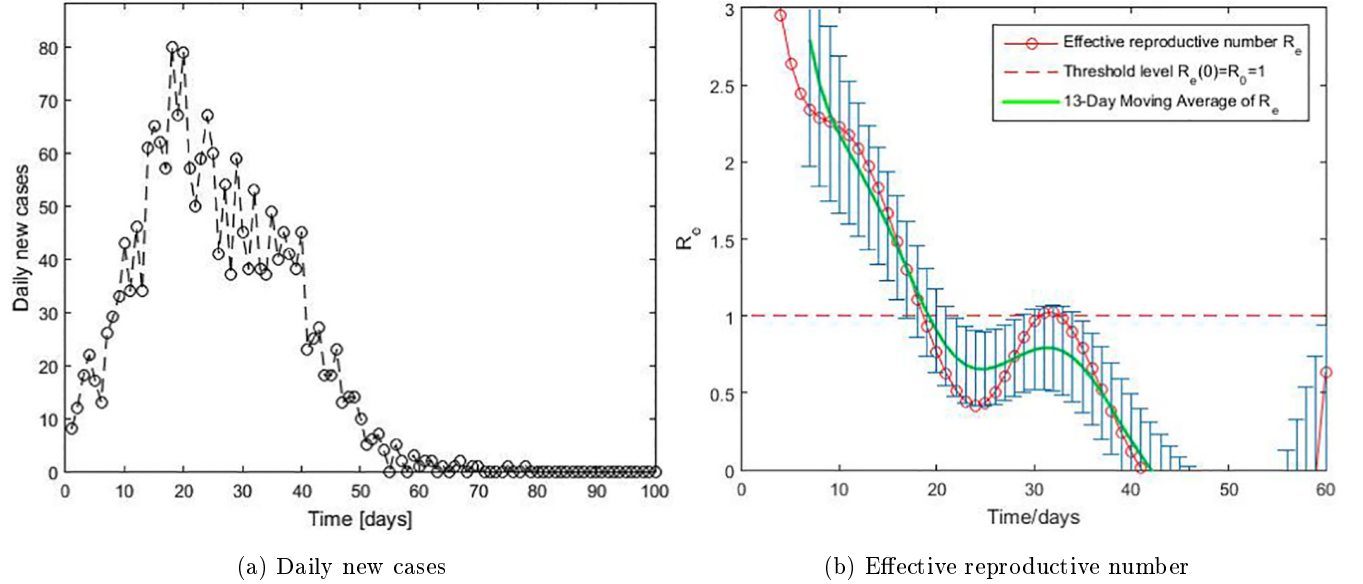


Figure 4: Graphs depicting the daily new cases and the effective reproductive number obtained from the reference paper [2].

5 Conclusion

In summary, the simulation of the spread of Coronavirus was implemented in MATLAB based on the reference paper [2] for the purpose of this assignment project for the Monte Carlo Method subject. The main objective of this work was to replicate results, specifically the number of daily new cases and the effective reproductive number from the reference paper [2] (see figure 4). While the results overall seem correlatable, there are discrepancies which the author was not able to remove. The differences could have originated, for example, from the fact that some variables, parameters and means of calculating statistical data mentioned in the reference paper [2] were open to interpretation. Another potential source of mismatching R_e data is that the authors mention using an unspecified method of data filtering.

References

- [1] WHO. (2021) Coronavirus disease (covid-19). [Online]. Available: https://www.who.int/health-topics/coronavirus#tab=tab_1
- [2] S. Maltezos and A. Georgakopoulou, “Novel approach for monte carlo simulation of the new covid-19 spread dynamics,” *Infection, Genetics and Evolution*, vol. 92, 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1567134821001933>
- [3] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, and Z.-L. Shi, “A pneumonia outbreak associated with a new coronavirus of probable bat origin,” *Nature*, vol. 588, no. 7836, pp. E6–E6, 2020-12-03. [Online]. Available: <http://www.nature.com/articles/s41586-020-2012-7>
- [4] MayoClinic. (2021) Covid-19 (coronavirus): Long-term effects. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-long-term-effects/art-20490351>
- [5] CDC. (2021) Science brief: Sars-cov-2 and surface (fomite) transmission for indoor community environments. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/surface-transmission.html>
- [6] H. Barratt, M. Kirwan, and S. Shantikumar. (2020) Epidemic theory (effective & basic reproduction numbers, epidemic thresholds) & techniques for analysis of infectious disease data (construction & use of epidemic curves, generation numbers, exceptional reporting & identification of significant clusters). [Online]. Available: <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/epidemic-theory>
- [7] (2021) Comparateur de territoire. [Online]. Available: <https://www.insee.fr/fr/statistiques/1405599?geo=DEP-75>
- [8] About mmr. [Online]. Available: <https://mmrda.maharashtra.gov.in/about-mmr>