Analysis of COVID-19 Pandemic Using Mathematical Modeling

Hirokatsu Suzuki
Minnesota State University Moorhead, Department of Physics and Astronomy, Moorhead MN

Abstract

The latest epidemic of the COVID-19 has spread to many countries. The objective of this research was to replicate the COVID-19 model using the SIR (Susceptible-Infected-Recovered) model to the actual data in the state of Minnesota. The SIR model includes three differential equations which derives the change in populations with some parameters. A program was made to construct SIR model that had two parameters, beta and gamma. Then the least fitting method was used to determine the best beta and gamma values. The SIR model was found to be not suitable to replicate the COVID-19 pandemic with the results of high significance ratio.

Introduction

SARS-CoV-2, the virus that causes COVID-19 is not the deadliest disease in existence, but it is one of the viruses that has impacted society. The goal of this research was to simulate the COVID-19 pandemic in the state of Minnesota using the mathematical model. Specifically, the SIR model will be programmed to match the actual data. Understanding the trend and the seriousness of the disease are essential among the young generation because they not only asymptomatically carry the airborne virus, but also spread it when not implementing any non-pharmaceutical intervention.

The SIR model (Susceptible-Infected-Recovered) is an analytical model using a system of three coupled non-linear ordinary differential equations. This model is often used to model outbreaks of information in the field of epidemiology [1]. Below are the differential equations used to construct SIR model,

1. $S(t) = \dfrac{dS}{dt} = -\dfrac{\beta * S * I}{N} + \gamma * I$

2. $I(t) = \dfrac{dI}{dt} = \dfrac{\beta * S * I}{N} - \gamma * I$

3. $R(t) = \dfrac{dR}{dt} = \gamma * I$

where S is the susceptible population, I is the infected population, R is the recovered population, β is the infection rate and gamma is the recovery rate which beta has a unit of per person and gamma has a unit of per day. The SIR model is the fundamental epidemiology model used to replicate the spread of disease over the time [2].

There are many other models that are constructed based on the SIR model such as the SEIR model. The SEIR (Susceptible-Exposed-Infected-Recovered) model has almost the same structure of the SIR model with additional exposed population, E. Unlike the SIR mode, the SEIR model contains many more parameters such as death rate, quarantining population, and etc [3]. The fatality rate varies in the different age groups [4], so it is extremely difficult to modify the differential equation with this parameter. Since these parameters are not obtainable in this research, the SEIR model is not practical here.

The COVID-19 model was constructed using Python programming language. It is the most beginner friendly existing computer language which is easy to read and understand. It also allows users to apply third party packages, such as Matplotlib and Pandas, to enable more features.

Method

In the three differential equations, there are two parameters which will impact the trend of the model. β is the average infection rate per person per time, and γ is the mean recovery rate. These parameters vary in different age groups, so some assumptions were needed to simplify the case.

1. *The population remains constant. If there were changes in the population, the SIR model would become more complicated model which is not the objective of the experiment.*
2. *Since the susceptible population was not provided, the susceptible population was calculated as the population of confirmed, recovered, and death subtracted from the total population.*
3. *The death population was neglected since it was reasonably small portion of the total population.*

Data was collected from the Census Bureau and Johns Hopkins using a Python program developed by Dr. Juan Cabanela [5]. The program was edited accordingly to only extract the data from the state of Minnesota. The Minnesota data was identified by its FIPS code which is a unique value assigned to each state. The program also specified to extract dates, confirmed, deaths, active, and recovered population from the Census and JH databases. The program will take few minutes to complete the data extraction, but it was saved as a csv file which is compatible with the Pandas package.

	A	B	C	D	E	F
1	Index	dates	Confirmed	Deaths	Active	Recovered
2	0	3/22/20	2	0	2	0
3	1	3/23/20	2	0	2	0
4	2	3/24/20	3	0	3	0
5	3	3/25/20	3	0	3	0
6	4	3/26/20	4	0	4	0
7	5	3/27/20	4	0	4	0
8	6	3/28/20	4	0	4	0
9	7	3/29/20	4	0	4	0

Figure.1 A screenshot of data extracted from the Python program. The program extracted the data from the March 22nd, 2020 to the present date.

In the program, few variables such as the total population of the state of Minnesota and time range used as the initial conditions in the SIR model. Then each population was imported from the data file. Since the susceptible population was not provided, it was recalculated by subtracting confirmed, recovered, and death population from the total population. There were two functions created in order to calculate the SIR equations and plot them. The interactive feature was added to the plot later, and the advantage of adding the interactive feature was to be able to change the beta and gamma values while updating the plot simultaneously. While programming the plots, there were many combinations of beta and gamma that produced similar SIR model. In order to avoid this issue, least square fitting method from SciPy package was used to analyze the best values of beta and gamma that produced best SIR model to match the actual data.

The program includes three separate files for active, susceptible and recovered population. In each file, a single population was plotted from the actual data as well as the corresponding SIR model. For example, active population was plotted along with the infected population within the SIR model in figure 2. Beta and gamma were adjusted so the SIR model matched the actual data plots.
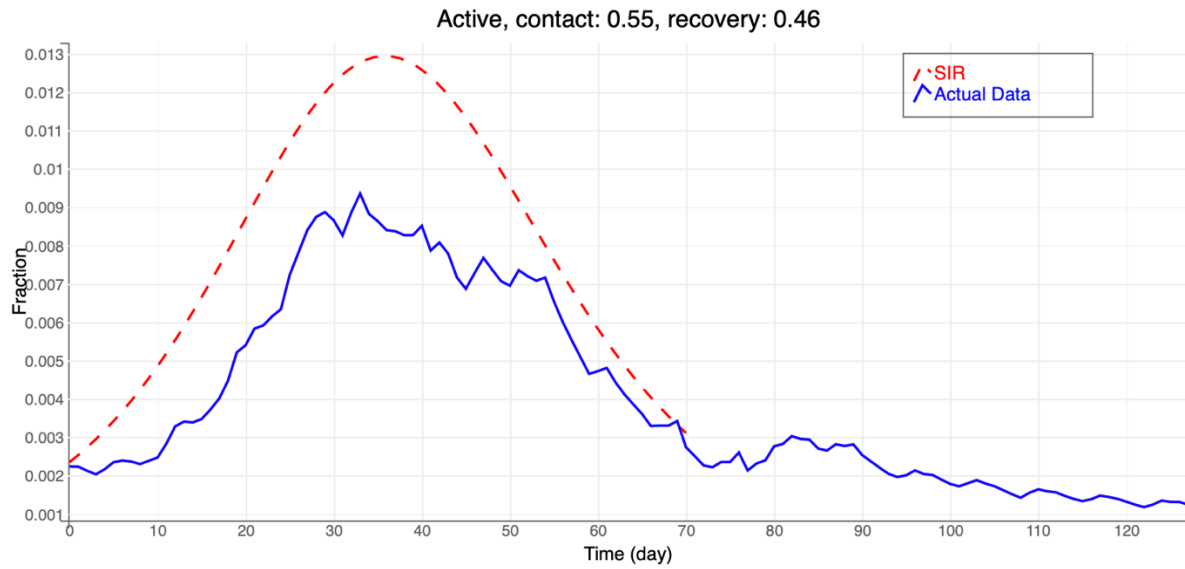
Result

Figure.2 Active population with the best fit values of beta (0.55) and gamma (0.46).
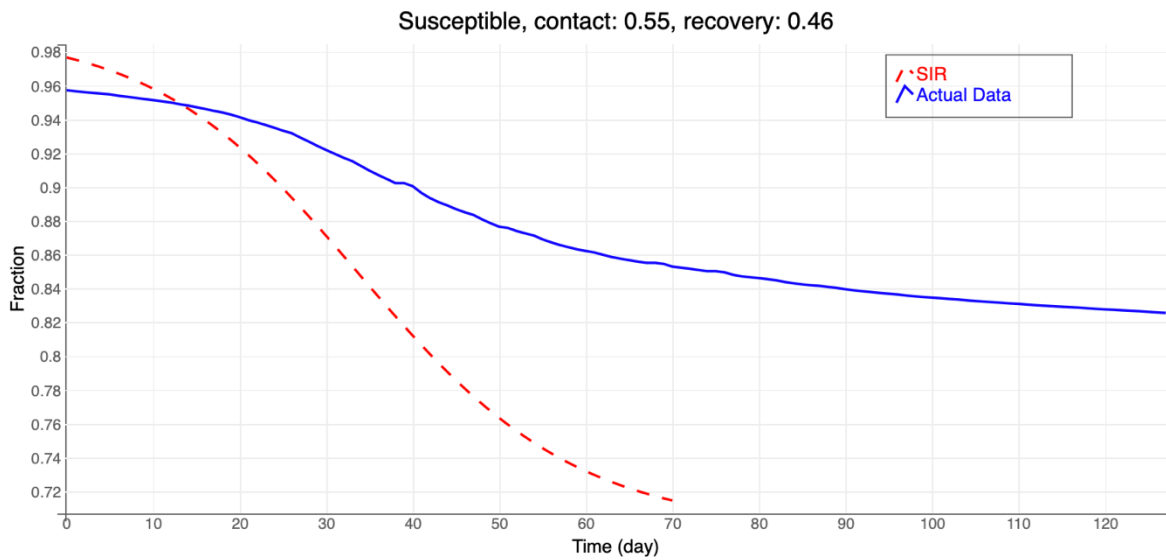

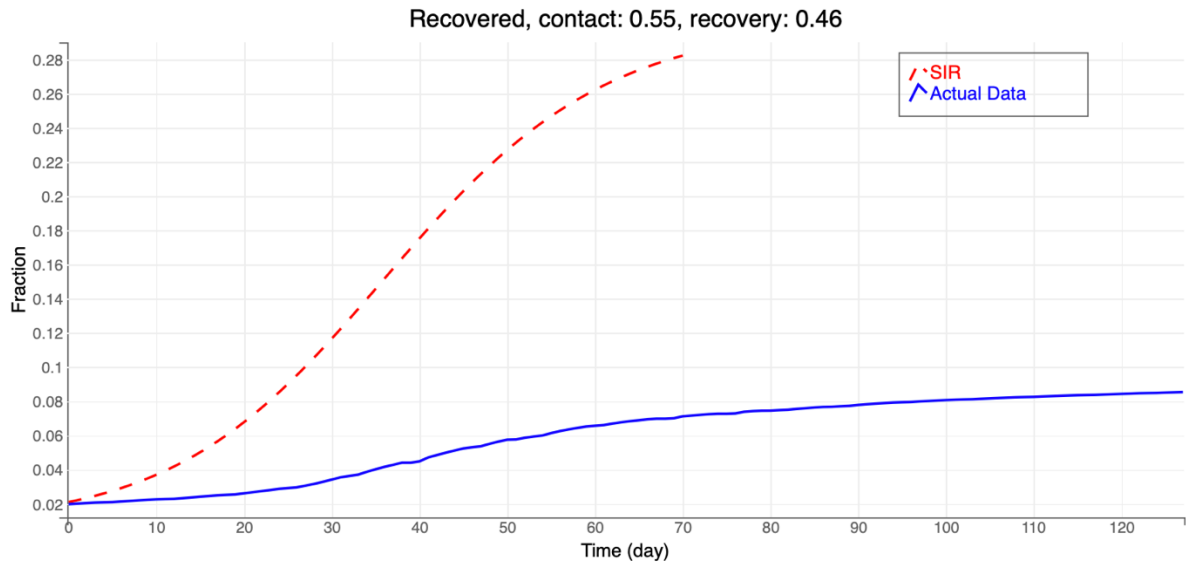Figure.3 Susceptible population with best fitted values of beta (0.55) and gamma (0.46) from figure 2.

Figure.4 Recovered population with best fitted values of beta (0.55) and gamma (0.46) from figure 2.
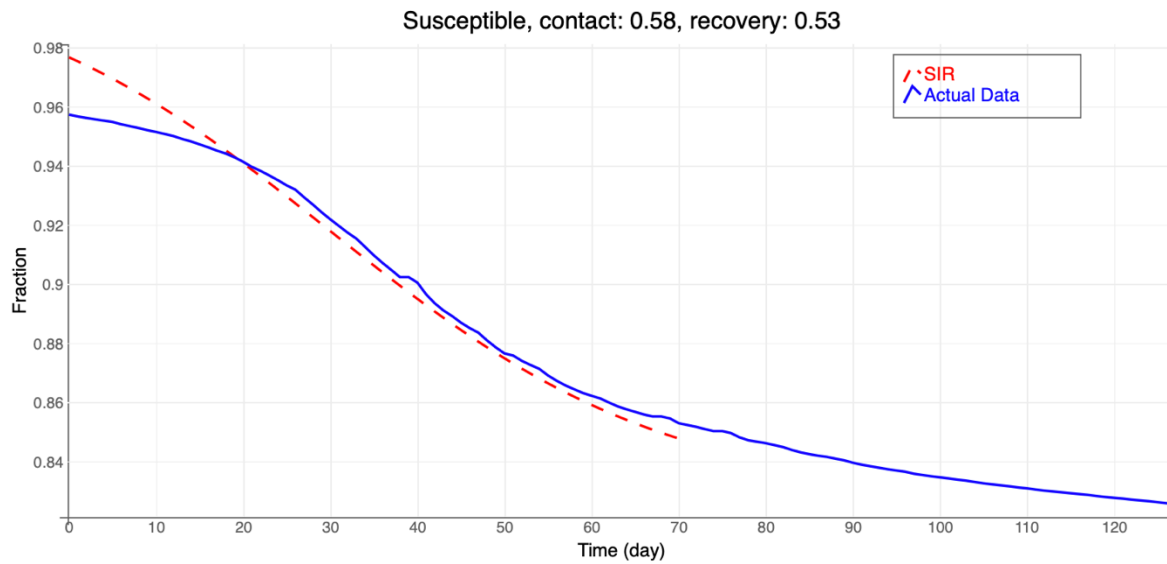

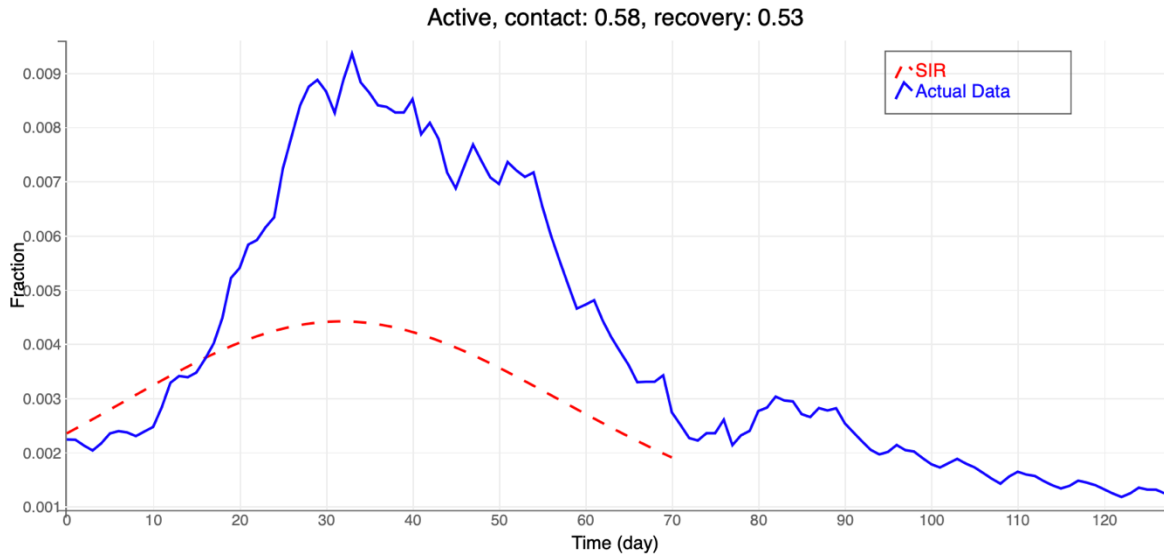Figure.5 Susceptible population with the best fit values of beta (0.58) and gamma (0.53).

Figure.6 Active population with best fitted values of beta (0.58) and gamma (0.53) from figure 5.
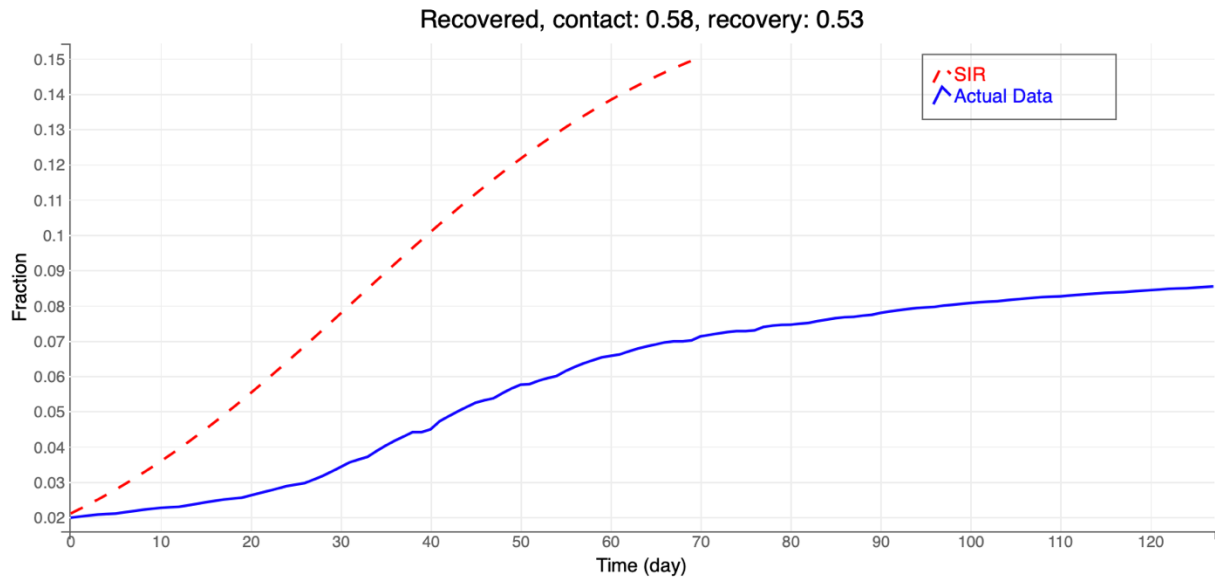

Figure.7 Recovered population with best fitted values of beta (0.58) and gamma (0.53) from figure 5.
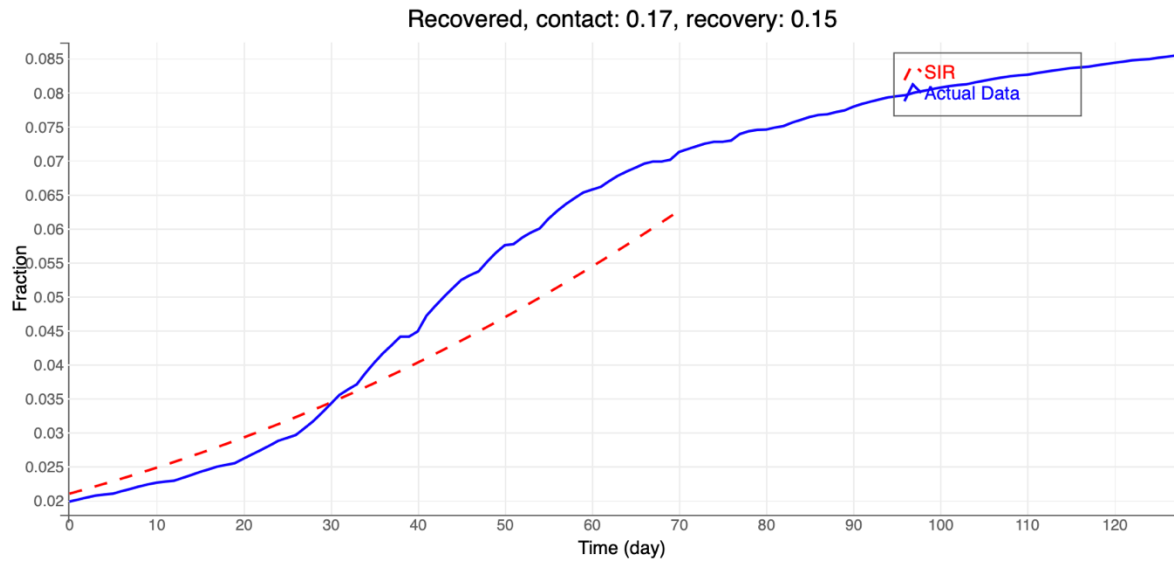
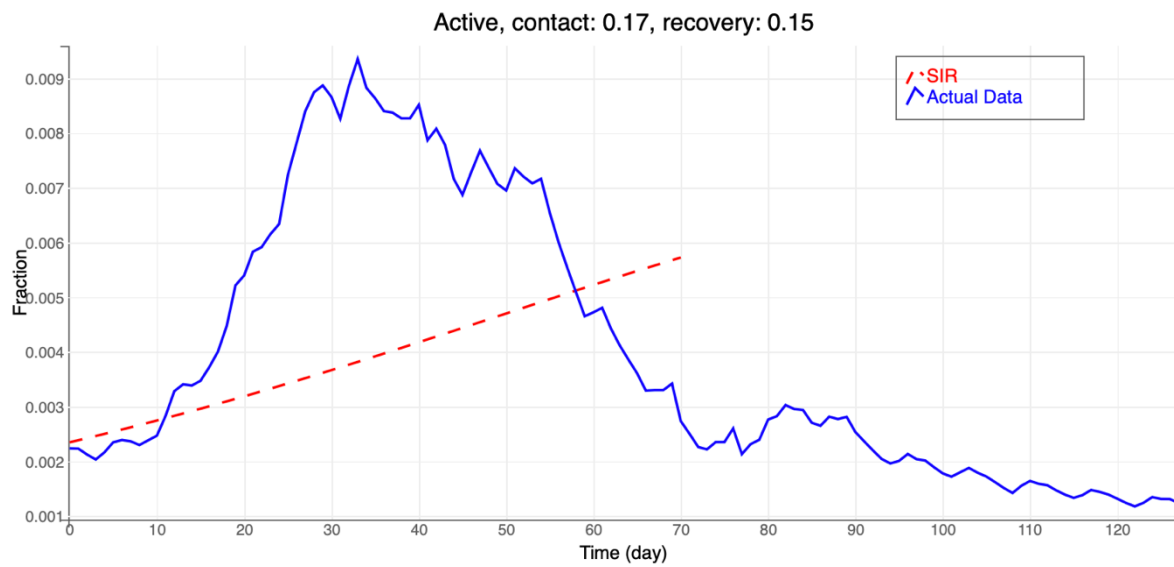Figure.8 Recovered population with the best fit values of beta (0.17) and gamma (0.15).

Figure.9 Active population with best fitted values of beta (0.17) and gamma (0.15) from figure 8.
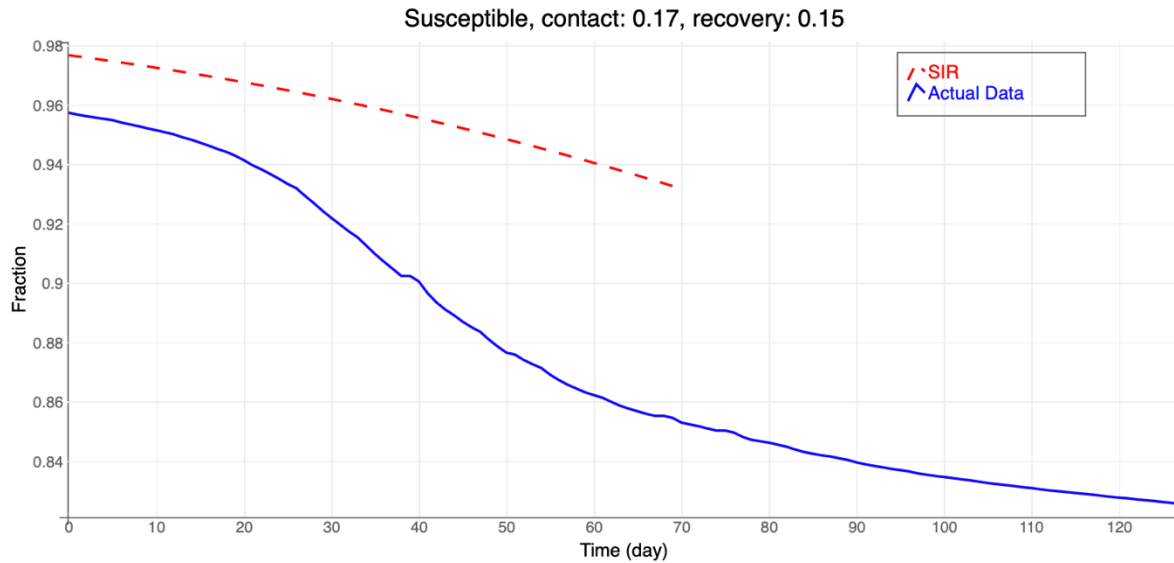
Susceptible, contact: 0.17, recovery: 0.15

Figure.10 Susceptible population with best fitted values of beta (0.17) and gamma (0.15) from figure 8.

The program was able to best fit the SIR model with the actual data to some extent. As shown above in figure 2 to 4, the values of beta and gamma were all different. In figure 2, the best fit values of beta and gamma were $55.0 \pm 0.8\%$ and $46.0 \pm 0.6\%$ respectively for the active population. In figure 5, the best fit values of beta and gamma were $58.0 \pm 2.0\%$ and $53.0 \pm 2.0\%$ respectively for the susceptible population. In figure 8, the best fit values were $17.0 \pm 0.7\%$ and $15.0 \pm 0.8\%$ respectively for the recovered population.

Conclusion

Overall, the objective of the research was reached, however, the SIR model is not the best model to simulate the COVID-19 pandemic. In order to access the facticity of the result, significance ratio was determined. According to the article, the COVID-19 virus takes between 2 to 14 days to recover [10]. In addition, an average infection rates a person can infect is between 2 to 20 days [11].

Type of population / Significance Ratio	Sig. Ratio of Beta (1 person / 2 day)	Sig. Ratio of Beta (1 person / 20 day)	Sig. Ratio of Gamma (1 person /2 days)	Sig. Ratio of Gamma (1 person / 20 days)
Active	$\dfrac{\|0.5 - 0.55\|}{0.00755} = 7$	$\dfrac{\|0.05 - 0.55\|}{0.00755} = 66$	$\dfrac{\|0.5 - 0.46\|}{0.00643} = 6$	$\dfrac{\|0.05 - 0.46\|}{0.00643} = 64$
Susceptible	$\dfrac{\|0.5 - 0.58\|}{0.01948} = 4$	$\dfrac{\|0.05 - 0.58\|}{0.01948} = 27$	$\dfrac{\|0.5 - 0.53\|}{0.01925} = 2$	$\dfrac{\|0.05 - 0.53\|}{0.01925} = 25$
Recovered	$\dfrac{\|0.5 - 0.17\|}{0.00700} = 47$	$\dfrac{\|0.05 - 0.17\|}{0.00700} = 17$	$\dfrac{\|0.5 - 0.15\|}{0.00829} = 42$	$\dfrac{\|0.05 - 0.15\|}{0.00829} = 12$

Table.1 Calculations of sig. ratio for beta and gamma in each figure.

From the result above, none of the significance ratio was below 3, meaning there is significant difference between the experimental result and the theoretical value. By interpreting the result, it is clear that there are other factors that needed to be considered and included into the three differential equations. Since the SIR model is the fundamental model of the epidemiological modeling, only the transmission rate and the recovery rate were taken into the equations in this research, therefore the SIR model does not produce accurate prediction for COVID-19 pandemic.

In preparation for further study, SIR model needs to be optimized. One of the improvements to the model is to take death population into consideration. Although the fatality rate is relatively small due to deadliness of the virus to different age groups, it may impact the recovered population [8]. On the other hand, the combinations of the best fit values are not looking realistic. One of the reasons would be when recovery rate is bigger than the transmission rate, the SIR model collapses. In fact, the theoretical transmission rate is smaller than the recovery rate.

This evidence supports that the SIR model is not suitable for modeling COVID-19 pandemic. By visually observing the shapes of both plots in figure 2 to 10, the SIR models do not match with the actual data. In addition to this, the SIR model failed to replicate the full plot from the actual data, meaning it was originally not designed to model such large-scale pandemic.

References

[1] H. Weiss, *MATerials MATemàtics 2013*, 1-17 (2013)

[2] W. O. Kermack, A. G. McKendrick, *PROCEEDINGS OF THE ROYAL SOCIETY A*, 700-721 (1927)

[3] S. He *et al.*, *Nonlinear Dynamics 101*, 1667-1680 (2020)

[4] A. Cagnacci, A. Xholli, *American Journal of Obstetrics and Gynecology 223(3)*, 453-454 (2020)

[5] https://github.com/JuanCab/COVID_DataViz

[6] https://github.com/HirokatsuSuzuki/Senior-Project

[7] https://www.statista.com/statistics/1119412/covid-19-transmission-rate-us-by-state/

[8] https://www.webmd.com/lung/covid-recovery-overview#1

[9] https://www.who.int/bulletin/volumes/99/1/20-265892/en/

[10] https://www.webmd.com/lung/covid-recovery-overview#1

[11] https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html#:~:text=Available%20data%20indicate%20that%20persons,days%20after%20symptom%20onset.

Acknowledgements

- Dr. Sumali Pandey, MSUM Biosciences Department
    - Went through the transmission process of COVID-19 virus and supported myself throughout the research.
- Dr. Juan Cabanela, MSUM Physics and Astronomy Department
    - Permitted me to use his data collection program and modified it accordingly. Also helped me to install the research environment on the computer.
- Dr. Matthew Craig, MSUM Physics and Astronomy Department
    - Had weekly meetings with myself and provided me with many helpful suggestions. Also helped me understand the SIR model programmed in the script.