PREDICTION OF COVID-19 CASES IN NEW YORK STATE USING SEIR MODEL

S. Nandi[1] and C. Vipulanandan[1], Ph.D., P.E.
[1]Texas Hurricane Center for Innovative Technology (THC-IT)
Department of Civil and Environmental Engineering
University of Houston, Houston, Texas 77204-4003
E-mail: shuvam.nandi1995@gmail.com, cvipulanandan@uh.edu Phone: (713) 743-4278

Abstract: With the ongoing pandemic due to the spread of SARS-CoV-2 virus the world has been in a state of disarray. Tens of millions of unemployment cases have been registered in the United States as well as more than a 100 million jobs lost in India due to the current stay-at-home-orders [1]. Mathematical epidemiology plays a crucial role in restricting the spread of large-scale infectious diseases such as COVID-19. Using compartment models in epidemiology, researchers can forecast the progression of diseases and understand the effect of interventions on the spread of infectious diseases, such as lock-downs and social distancing [2]. In this study we take a look at how the SEIR model can be used to predict the total COVID-19 cases in the state of New-York. For the purpose of this study Octave package was used to write the code to implement the model.

1. Introduction:

Mathematical models help to project how an infectious disease is likely to progress and predict the outcome of an epidemic. Compartment models help simplify this further. According to the theory, the population is assigned to compartments with labels – for example in the case of the SIR model we have three compartments: S (Susceptible), I (Infectious) and R (Removed). People then move through each compartment at a given rate. All of the models are dependent on a set of ordinary differential equations.

The SEIR model has four different compartments, namely, S (Susceptible), E (Exposed), I (Infectious) and R (Removed). The susceptible compartment consists of those individuals that are prone to catching the disease. The exposed compartment consists of those individuals that have acquired the disease however are not at the stage to transmit them to others. The infectious compartment consists of those who can transmit the disease. Lastly, the removed compartment consists of those who have either recovered or those who have died.

2. Objective:

The objective was to investigate the SEIR model for predicting the total positive COVID-19 cases in New York State. Data containing the cumulative number of positive cases of COVID-19 in New York State were collected from 4 March to 1 June (over a period of 90 days) and compared with the model prediction. The data was taken from [3].

3. Methods:

As stated earlier, the following compartment variables were used: S for susceptible; those who are able to contact the virus, E for exposed; those who have been infected but are not yet contagious, I for infectious; those who are capable of transmitting the virus and R for removed; those who have either recovered or are dead. The model assumes that the net change in the population is zero. In other words, the birth rate and death rate are assumed to have negligible effect on the entire population.
Since the birth rate and death rate are assumed to be equal, $S+E+I+R = N$; where N is the total population.

The rate coefficients or functions between the different states are β (transmission rate; rate from S to E in day-1), σ (incubation rate; rate from E to I in day-1) and γ (recovery rate; rate from I to R in day-1). This is clearly depicted in Figure 1.

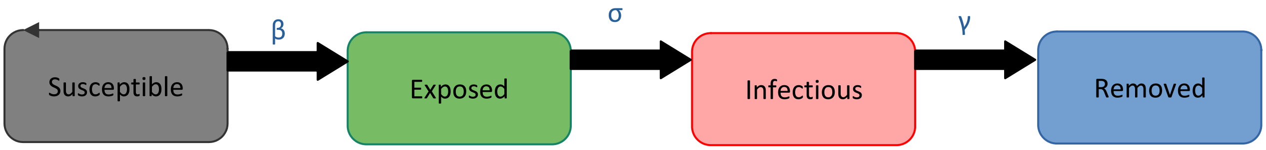

Figure 1: SEIR model compartments and rate coefficients

This model has four governing equations:
Equation for S: $\frac{dS}{dt} = -\beta S \frac{I}{N}$
Equation for E: $\frac{dE}{dt} = \beta S \frac{I}{N} - \sigma E$
Equation for I: $\frac{dI}{dt} = \sigma E - \gamma I$
Equation for R: $\frac{dR}{dt} = \gamma I$

The above system of non-linear ordinary differential equations is solved by numerical integration in Octave. The initial conditions required for all variables are S_0, E_0, I_0 and R_0. In this case, the values of S_0, E_0, I_0 and R_0 were assumed to be 1880400, 600, 6 and 0 respectively. These values were set according to the data collected from the internet for positive cases of COVID-19 in New York State on March 4, 2020. S_0 value was assumed to be equal to the total population of New-York. I_0 value was taken as 6. This corresponds to the number of positive cases of COVID-19 in New York State as measured on March 4. Since there were no reports of any recovered or deceased cases on March 4, the value of R_0 was taken as 0. Since, no information on the number of exposed populations were available a value of 600 was assumed.

The transmission rate, β is given by the expression:

$$\beta = \frac{R_O}{n_{id}} = R_O \gamma$$

R_o is termed as the reproduction number and denotes the number of people infected by each infectious person. It varies with time. It starts off large but decreases with time due to public health actions such as quarantine and social distancing. n_{id} is the number of days spent in the infectious state. As seen from the above expression, the rate from infectious state to removed state is expressed as $\gamma = \frac{1}{n_{id}}$. Initially the value of R_o was assumed to be a constant value in order to fit the curve. In Figure 2 we can see how the curve changes for different values of R_o. As seen from the figure for a fixed value of R_o we are unable to get a good fit for the curve.

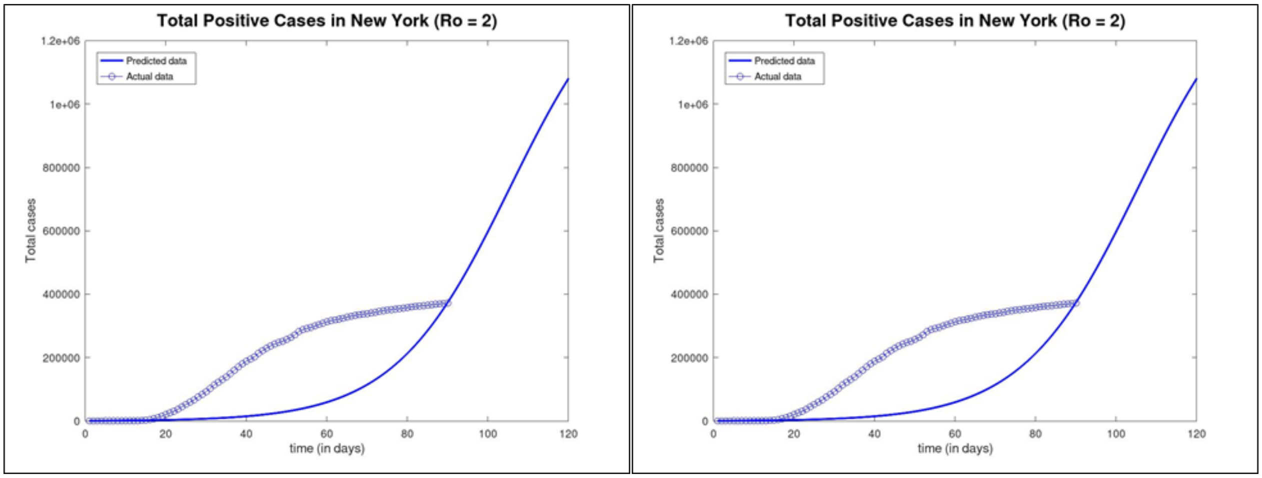

Figure 2: Total cases vs time curve for (a) $R_o = 2$ and (b) $R_o = 5$

Finally, R_o was represented as a logistics function as expressed below:

$$R_o = \frac{R_o, start - R_o, end}{1 + e^{(-k(-t+t_o))}} + R_o, end$$

$R_{o,\ start}$ and $R_{o,\ end}$ are the values of R_o on the first and last day. t_o is the t-value (day) of the inflection point or the date of the steepest decline in R_o. k controls how quickly the value of R_o declines. From trial and error, it was found that the curve fit best when $R_{o,\ start}$, $R_{o,\ end}$, t_o and k were equal to 5.65, 0.55, 25 and 0.1 respectively.

In Figure 3(a) it can be seen that the curve fit quite nicely for the above assumed values. Figure 3(b) depicts how the population in each compartment changes with time. It can be seen from the plot that as time goes by the number of susceptible decreases to a minimum value whereas the number of removed increases to a maximum value after a certain number of days have passed. The exposed and infectious group first increase to a maximum value then decreases from there on.

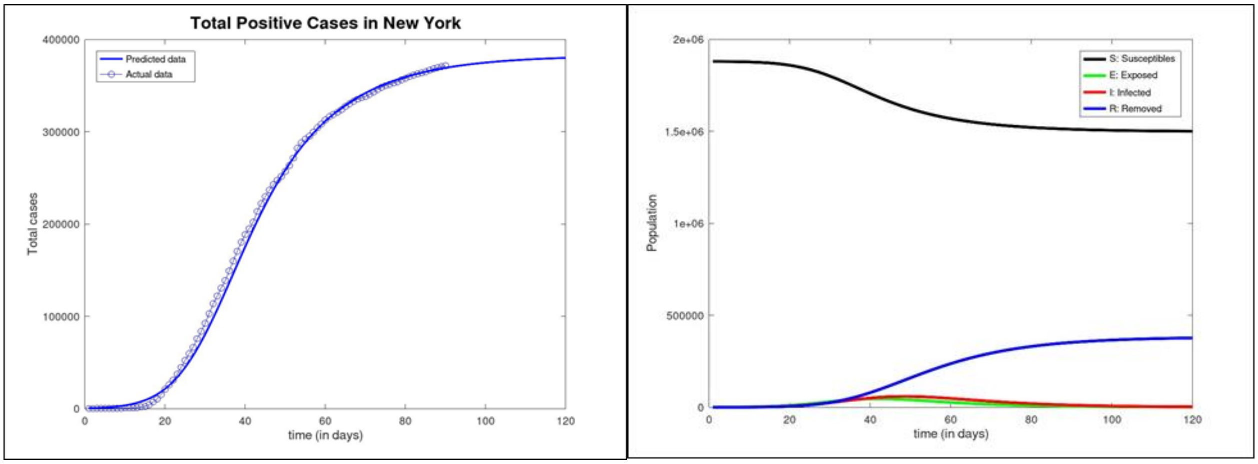

Figure 3: (a) Total cases vs time curve when Ro is taken as a logistics function, and (b) corresponding plot showing the changes in the different compartments over time

4. Conclusion:

This study showcases how compartment models such as the SEIR model could be used to predict the outcome of an epidemic such as COVID-19.

5. Acknowledgements:

This study was supported by the Texas Hurricane Center for Innovative Technology (THC-IT), University of Houston, Houston.

6. References:

1. news.uschicago.edu/story/how-covid-19-has-changed-global-economy
2. theconversation.com/how-to-model-a-pandemic-134187
3. https://www.statista.com/statistics/1109721/new-york-state-covid-cumulative-cases-us/