

Dynamics of Disease Spread: A Mathematical Model for Understanding the Current COVID-19 Pandemic

Abigail Saenz

Spring 2020

Introduction

Coronavirus disease (COVID-19) is an infectious disease caused by severe acute respiratory syndrome SARS-CoV-2. The outbreak was first identified in Wuhan, China in December 2019 [1]. The World Health Organization (WHO) officially declared the outbreak a pandemic on March 11, 2020 [2]. Local transmission of the disease has occurred in most countries across all six World Health Organization regions. As of June 8, 2020, COVID-19 has resulted in 6,931,000 cases worldwide. [3].

The Problem

In the absence of pharmaceutical solutions, addressing the COVID-19 outbreak will depend critically on the successful implementation of sustainable preventative measures. These measures can be optimized through a mathematical understanding of the transmission dynamics of the virus. Implementing a mathematical model is necessary for making predictions about the prognosis of the outbreak. Such a model will allow us to assess the consequences of certain behaviors, like shelter in place orders. A model will allow us to measure the success of these controls by allowing us to compare the expected outcome for a region with a strict mandate to regions with more relaxed orders. It also enables us to understand how varying the values of several parameters, such as the reproduction number R_0 or the fatality rate, alter our expectations as the pandemic continues to progress.

Why This Is Interesting

A mathematical model can help us determine what controls are optimal for containing the outbreak by measuring their success in reducing the reproductive number, R_0 . Reducing the reproductive number can be achieved by implementing behaviors like the mandated shelter in place. By reducing the expected number of individuals that an infected person will spread the virus to, the virus will spread at a comparably more manageable rate. This will reduce the strain on the healthcare system by reducing the number COVID-19 patients in hospitals. This is important because it allows healthcare workers to provide a higher quality of treatment for those infected, which may lower the fatality rate of COVID-19 if it is resource-dependent. A successful model will also allow us to make predictions about future outbreaks of infectious diseases and how to contain them.

Summary of the Proposed Solution

In order to estimate how various behaviors and parameters will effect the system in coming months, we will implement compartmental models for COVID-19.

Goals:

- Model COVID-19 as a dynamical system by developing a variant of the classic SIR

model, which divides a population into susceptible, infected, and removed compartments.

- Fit our model to data for the number of cases and deaths in the United States.
- Consider the effects of certain controls, i.e. the mandated shelter in place in California, on the reproduction number R_0 and explore the effects of varying values of R_0 .
- Explore the consequences of a functional, time-dependent R_0 and of resource-dependent and age-dependent fatality rates.
- Fit our model to data for the number of cases and deaths in the United States.

The Model

In this section, we will develop a compartmental model to study the COVID-19 pandemic based off of the standard SIR model by Kermack and McKendrick [4]. The SIR model divides a population into three compartments: susceptible, infected, and removed. We will develop a model that considers additional compartments such as exposed and dead.

The following parameters will be needed for modeling :

N : total population

$S(t)$: number of people susceptible on day t

$E(t)$: number of people exposed on day t

$I(t)$: number of people infected on day t

$R(t)$: number of people removed (recovered or dead) on day t

$D(t)$: number of people dead on day t

β : expected amount of people an infected person infects per day

D : number of days an infected person has and can spread the disease

γ : the proportion of infected recovering per day ($\gamma = 1/D$)

R_0 : the total number of people an infected person infects ($R_0 = \beta/\gamma$)

δ : 1/length of the incubation period

α : fatality rate

ρ : rate at which people die (= 1/days from infected until death)

Constructing the Compartments

The SIR model takes a population of size N and divides this population into three compartments: $S(t)$, $I(t)$, and $R(t)$. As seen in [Figure 1](#), susceptible individuals become infected at a rate of βSI and infected individuals enter the removed compartment at a rate of γI . The parameter β is the expected number of people an infected person will infect per day. The parameter γ is the proportion of infected individuals recovering per day. We will assume, for now, that individuals in the removed compartment will have lasting immunity and are incapable of re-entering the susceptible population.



Figure 1: Flowchart for the SIR Model. The arrows between compartments represent the transition from one compartment to another at a given rate.

From these transitions, we obtain the following set of differential equations that describe changes in the number of individuals in each compartment at time t :

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dI}{dt} = \beta SI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (1)$$

These equations provide directions for “transitions” between one compartment to the next. We represent these transitions using arrows, as seen in [Figure 1](#).

Constructing an Exposed Compartment

Though the SIR model enables us to understand the transitions between compartments, these three compartments alone are not very useful for modeling COVID-19. The model fails to take into consideration factors such as individuals who are exposed to the virus but are not yet infectious. To create an exposed compartment, $E(t)$, we need to consider an incubation period in which a susceptible individual is exposed to the virus but can not yet spread it. Transitions would then occur from $S \rightarrow E \rightarrow I \rightarrow R$. We will again consider transitions from one compartment to another as being represented by arrows, as seen in [Figure 1](#).

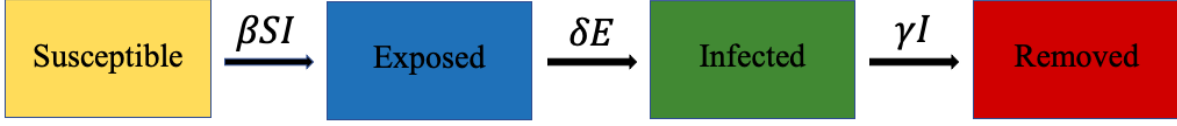


Figure 2: Flowchart for the SEIR Model. The arrows between compartments represent the transition from one compartment to another at a given rate.

Individuals are still being removed from the susceptible population at a rate of βSI , so they will enter the exposed compartment at this rate. Similarly, individuals in the infected population are still recovering at a rate of γI . The only new transition is from compartments $E \rightarrow I$. We will assume for now that the probability that an individual moves from the exposed compartment to the infected compartment is 1, meaning that everyone who is exposed to the virus will become infected following an incubation period. The population is all individuals in the exposed compartment. The rate is $1/(\text{the number of days that the incubation period will last})$, which we will denote δ . Thus, we have that individuals who are exposed become infectious at a rate of δE , as seen in [Figure 2](#).

From these transitions we obtain the following equations for modeling the changes between compartments:

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dE}{dt} = \beta SI - \delta E \\ \frac{dI}{dt} = \delta E - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \quad (2)$$

Still missing from this model is the possibility of individuals in the infected compartment dying and being completely removed from the population. For this reason, we need to update the model such that individuals in the infected compartment enter one of two possible compartments: dead and recovered.

Constructing the Dead Compartment

Using the same method used for deriving the exposed compartment, we find that individuals move from $I \rightarrow D$ at a rate of $\rho \alpha I$, as seen in [Figure 3](#). The probability of individuals entering the dead compartment is equal to the fatality rate, α . The rate at which these individuals die is given by the parameter ρ . The probability of infected individuals entering the recovered compartment is given by $(1 - \alpha)$. The rate at which these individuals recover

is again given by the parameter γ . Thus, we find that individuals move from $I \rightarrow R$ at a rate of $\gamma(1 - \alpha)I$, as seen in [Figure 3](#).

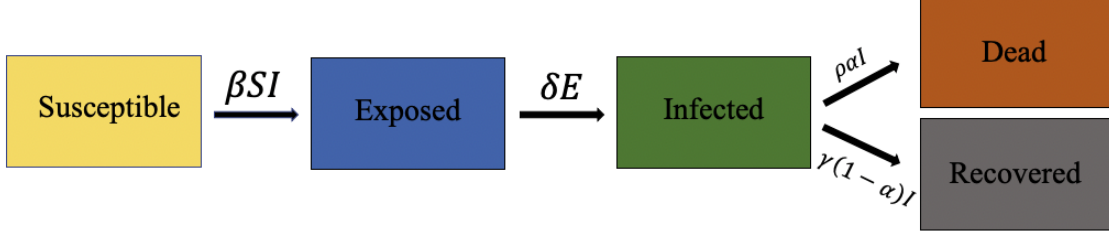


Figure 3: Flowchart for the SEIRD Model. The arrows between compartments represent the transition from one compartment to another at a given rate.

From these transitions we obtain the following equations for modeling the changes between compartments:

$$\begin{cases} \frac{dS}{dt} = -\beta SI \\ \frac{dE}{dt} = \beta SI - \delta E \\ \frac{dI}{dt} = \delta E - (1 - \alpha)\gamma I - \alpha \rho I \\ \frac{dR}{dt} = (1 - \alpha)\gamma I \\ \frac{dD}{dt} = \alpha \rho I \end{cases} \quad (3)$$

Using this SEIRD model, we will consider a population of 100 individuals with 99 susceptibles, 1 initial exposed individual, 0 infected individuals, 0 recovered individuals, and 0 deceased individuals. We fixed parameters $\gamma = 1/4$, $\delta = 1/5$, $\alpha = 1/5$, and $\rho = 1/9$. For various values of the reproduction number R_0 , the expected number of people an infected person will infect per day, we obtain the results shown in [Figure 4](#).

We initialized that population $N(t) = 1$, so that the percent of individuals is given on the y-axis. For a high value of the parameter $R_0 = 5$, the virus spreads rapidly and the number of susceptible approaches zero by Day 40. The proportion of individuals exposed and infected obtain their maximums near the 20 day mark. Reducing the reproduction number to $R_0 = 3.5$ results in the population of susceptibles approaching 0 but never reaching it by Day 100. The reduced R_0 also flattens the curves for both exposed and infected individuals, reducing the value of their maximum rate of growth and delaying their respective peaks to occur between 20 and 30 days. Reducing the reproductive number to $R_0 = 2$ results in roughly 20% of the susceptible population never becoming exposed to the virus by Day 100.

It also significantly flattens the curves for both the susceptible and infected populations and delays their respective peaks to occur between the 40 and 50 days.

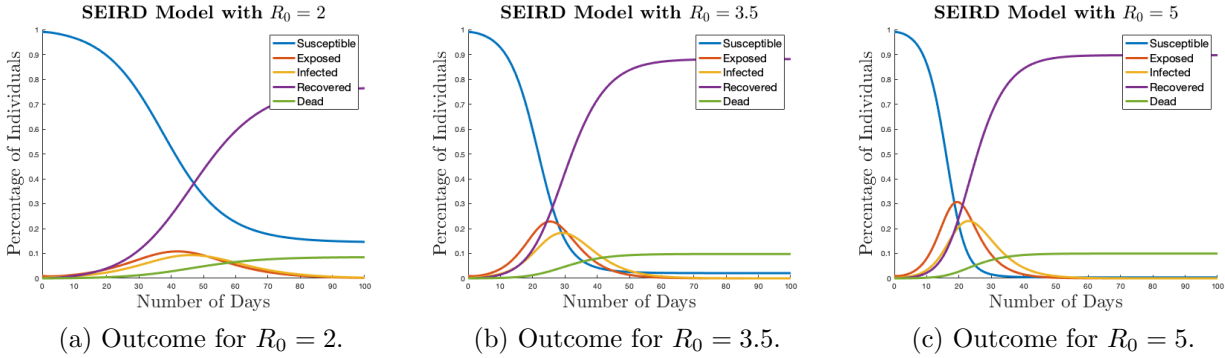


Figure 4: SEIRD model for various values of the reproduction number R_0 . As the value of the reproduction number varies, the dynamics between the compartments change such that for lower values of R_0 , less of the population is becoming exposed/infected with the virus.

These results, as shown in Figure 5ab, suggest that reducing the value of R_0 flattens the curve for the both the exposed and infected compartments. This can help contain the outbreak by reducing the number of people becoming exposed and infected to the virus at a single time. Mitigating this spread minimizes the strain on the health care professionals by providing them with fewer patients at a time. This may allow for better treatment of those infected, which may in turn reduce the fatality rate. Figure 5c also suggests that for a critical value of R_0 , a portion of the population will never become infected. I will further discuss the implications of a time-dependent R_0 in a future section.

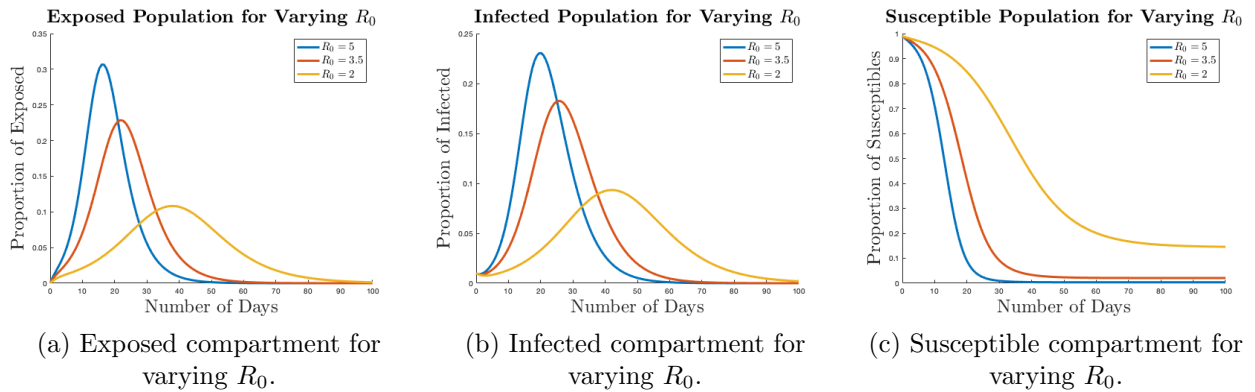


Figure 5: Modeling the changes in the exposed, infected, and susceptible compartments for various values of the reproductive number. Reducing the reproduction number is successful in critically lowering their maximum values. A lower reproduction number also slows the rate at which the susceptible compartment continues to decay, effectively increasing the minimum value obtained such that a portion of the population never acquires the virus.

Time-Dependent R_0

In this section, we will explore the consequences of implementing a time-dependent reproduction number in our model.

Effects of a Constant Reproduction Number on the SEIRD Model

As shown previously, the dynamics of our system are very sensitive to initial conditions. A reduction in the reproductive number, R_0 , has a large impact on the system and can help mitigate the amount of people moving from $S \rightarrow E$ and consequently, mitigate the amount of people transitioning from $E \rightarrow I$.

The imposition of a mandated shelter in place can greatly reduce the value of R_0 . We will consider the consequences of this reduction by simulating a lockdown on days $L = 50$ and $L = 40$ across a period of 100 days and comparing these to the results obtained in Figure 4, where no lockdown was imposed. For $t < L$, we will fix the reproduction number $R_0 = 5$. For $t \geq L$, we will fix the reproduction number to be an arbitrary value $R_0 = 0.9$. The choice of R_0 was random, with our only motivation being to choose a value far less than $R_0 = 5$ to investigate how smaller values of the reproduction number will effect the behavior of the system.

We will consider a constant population $N(t) = 1,000,000$. Initially, there will be 999,999 susceptible individuals, 1 exposed individual, 0 infected individuals, 0 recovered individuals, and 0 deceased individuals. We will fix the same parameters from Figure 5, where $\gamma = 1/4$, $\delta = 1/5$, $\alpha = 1/5$, and $\rho = 1/9$. Figure 6 shows the effects of implementing a lockdown for various days, L , of the outbreak.

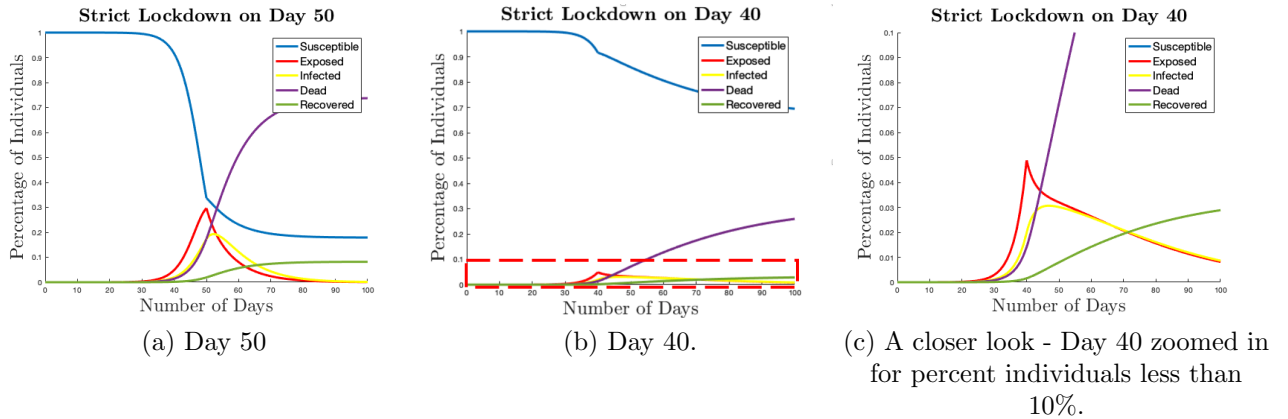


Figure 6: Lockdown imposed that reduces the value of R_0 for $t = 50$ and $t = 40$. Figure 6c provides a closer look at the consequences of imposing a strict lockdown on day 40 by focusing on values less than 10% of the population.

For a lockdown imposed on day 50, the infected population reaches a maximum of 20%. Similarly, the exposed population reaches a maximum of 30% of the population being affected. At the end of the 100 day timespan, 20% of the population remains susceptible to the virus. These results are reflected in [Figure 6a](#).

For a lockdown imposed on day 40, we see a reduction in the maximum number of exposed and infected individuals. The infected population reaches a maximum with only 3% being infected. Similarly, the exposed populations peaks with only 5% of the population being affected. By the end of the 100 day timespan, 70% of the population is susceptible and have not acquired the virus. These results are reflected in [Figures 6bc](#).

These results suggest that imposing the lockdown on an earlier date effectively contains the spread of the virus by reducing the reproductive number. However, in reality R_0 doesn't jump from one value to another. It continuously changes and may increase or decrease several times, i.e. if social distancing measures are relaxed and then tightened again.

Effects of a Logistic Reproduction Number on the SEIRD Model

In order to model the effects of a non-constant reproduction number, we have chosen to model the impact of social distancing measures using the following logistic function:

$$R_0(t) = \frac{R_{0_{\text{start}}} - R_{0_{\text{end}}}}{1 + e^{-k(-t-t_0)}} + R_{0_{\text{end}}} \quad (4)$$

Here, the parameters for this function are given by:

$R_{0_{\text{start}}}$: The value of R_0 on the first day of the outbreak.

$R_{0_{\text{end}}}$: The value of R_0 on the final day of the outbreak.

t_0 : The date of steepest decline in R_0 .

k : The rate of decline for R_0 .

We expect the date of steepest decline, t_0 , to be the main lockdown date.

We model R_0 for various values of the parameter k , as shown in [Figure 7](#). The figure is consistent with our assumption that R_0 declines more rapidly for larger values of the parameter k .

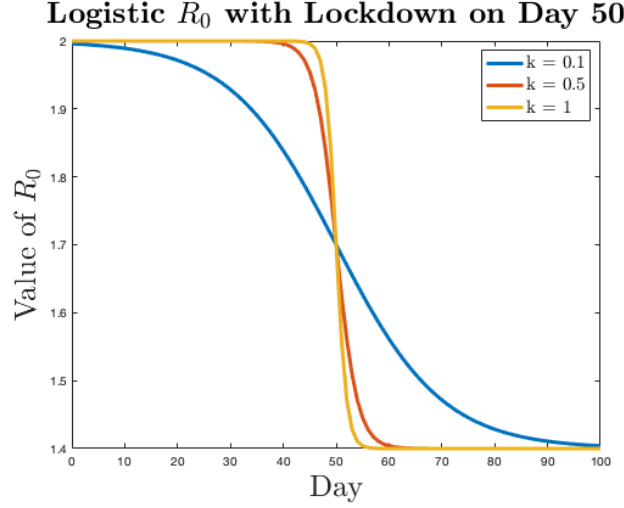
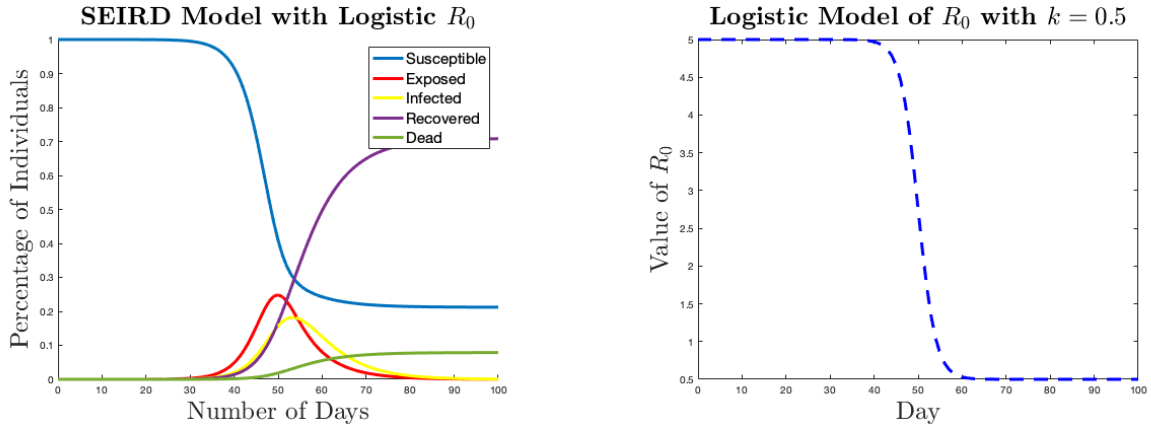


Figure 7: Logistic Model of R_0 for various values of the rate of decline, k . Lower values of k correspond to an accelerated rate of decay in R_0 at Day 50.

We now implement the logistic R_0 into the model previously used in Figure 6. We specified initial and final values of R_0 to be 5 and 0.5, respectively. We fixed the parameter $k = 0.5$ and our lockdown day $t_0 = 50$. The results are shown in Figure 8.



(a) SEIRD model with logistic R_0 , where the main lockdown date is $t_0 = 50$.

(b) Logistic model of R_0 for $k = 0.5$, where the main lockdown date is $t_0 = 50$.

Figure 8: SEIRD model with implemented logistic R_0 where rate of decay is $k = 0.5$ and the lockdown is imposed on Day $t_0 = 50$. R_0 was programmed to decline from the value 5 to a value of 0.5, according to equation 4.

We do the same for the case that the lockdown is imposed on $t_0 = 40$, as shown in

Figure 9.

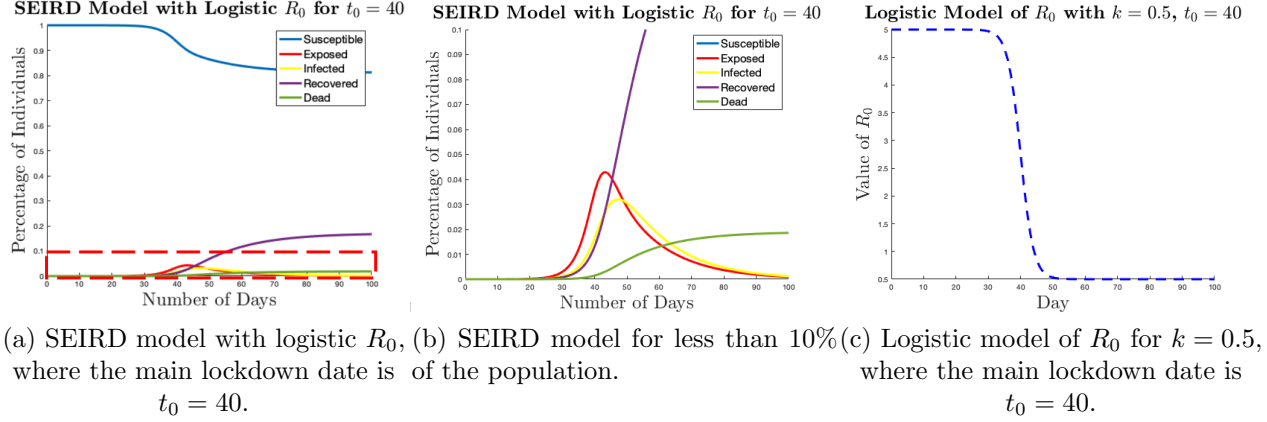


Figure 9: SEIRD model with implemented logistic R_0 where rate of decay is $k = 0.5$ and the lockdown is imposed on Day $t_0 = 40$. R_0 was programmed to decline from the value 5 to a value of 0.5, according to [equation 4](#).

In both [Figure 8](#) and [Figure 9](#), we see that by allowing R_0 to decline from 5 to 0.5 at a rate of $k = 0.5$, the curves for each compartment begin to flatten out at our specified values $t_0 = 50$ and $t_0 = 40$. It is evident that a logistic R_0 causes much smoother changes in the curves, as compared to the results obtained in [Figure 6abc](#) where we specified an instantaneous change in R_0 at days $L = 50$ and $L = 40$ that remained constant for all time.

Resource and Age-Dependent Fatality Rates

Much like the reproduction number R_0 , the fatality rate is most likely not constant. It can depend on a number of things, but we will focus on two specifically: resource-dependence and age-dependence. In this section, we will discuss the consequences of resource and age-dependent fatality rates on our SEIRD model with implemented time-dependent R_0 .

Resource-Dependence

We expect resource-dependent fatality rates to increase proportionally to an increase in the number of infected individuals in the population. This is because for a fixed amount of resources, a higher proportions of individuals will cause a shortage of resources.

In order to model resource dependence, we need to specify an optimal fatality rate for the case that few individuals in the population are infected and there is a surplus of resources available. We also need a factor that considers the proportion of the population that is infected. We will now model our fatality rate α as a function of time according to the following equation:

$$\alpha(t) = \mu \frac{I(t)}{N} + \alpha_{opt} \quad (5)$$

where μ is a fixed scaling factor that controls how the proportion of infected individuals influence the fatality rate and α_{opt} is a fixed optimal fatality rate.

For a large proportion of infected individuals, say half of the population, we would expect μ to be larger. If we specify $\mu = 1$, we would have a fatality rate $\alpha(t) = 1 \cdot \frac{1}{2} + \alpha_{opt} = 50\% + \alpha_{opt}$.

In the event that few infectious individuals are symptomatic and choose to stay home and not occupy resources, we could specify a smaller value of $\mu = 0.5$. For the same proportion of infected individuals, we would then have a fatality rate $\alpha(t) = (0.5) \cdot \frac{1}{2} + \alpha_{opt} = 25\% + \alpha_{opt}$.

Age-Dependence

In the event that fatality rates increase in older age groups, as suspected in COVID-19, it is important to consider how the overall fatality rate varies with respect to these different groups. For populations with large proportions of older individuals, we expect the fatality rate to be much higher than it would be in a younger population.

To model this age dependence, we will model the fatality rate α_r as the sum of all products of the fatality rate for a given age group and their respective proportions in the population:

$$\alpha_r = \sum_{i=1}^{\# \text{ of groups}} \alpha(i) \cdot p(i) \quad (6)$$

Consider dividing the population into four age groups with the following fatality rates and proportions:

| Age Group | Fatality Rate | Proportion of Total Population |
|-----------|---------------|--------------------------------|
| 0-24 | 0.01 | 0.2 |
| 24-49 | 0.05 | 0.2 |
| 50-74 | 0.3 | 0.3 |
| 75+ | 0.4 | 0.3 |

Here, $\alpha_r = \sum_{i=1}^4 \alpha(i) \cdot p(i) \approx 22\%$. If we now consider a population with a larger proportion of individuals under 50:

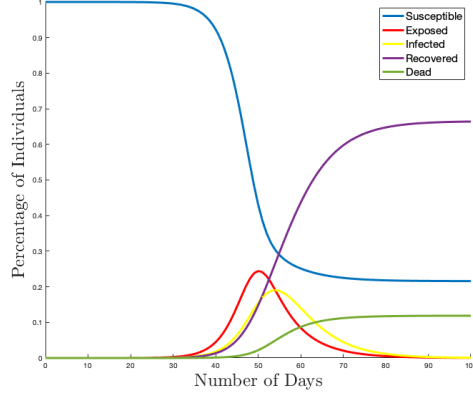
| Age Group | Fatality Rate | Proportion of Total Population |
|-----------|---------------|--------------------------------|
| 0-24 | 0.01 | 0.3 |
| 24-49 | 0.05 | 0.3 |
| 50-74 | 0.3 | 0.2 |
| 75+ | 0.2 | 0.2 |

we have that $\alpha_r = \sum_{i=1}^4 \alpha(i) \cdot p(i) \approx 16\%$.

These results suggest that populations with a larger proportion of individuals under 50 will have a lower collective mortality rate.

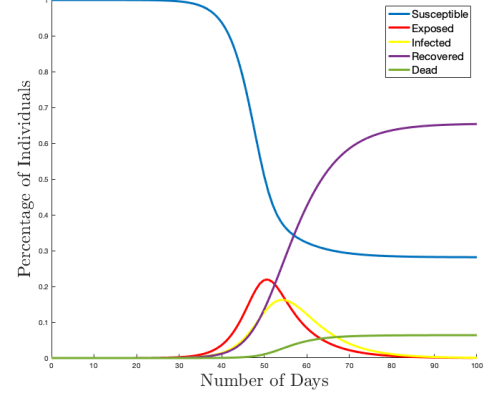
Effects of Resource-Dependent and Age-Dependent Fatality Rates on the SEIRD Model

SEIRD Model with Logistic R_0 for an Older Population



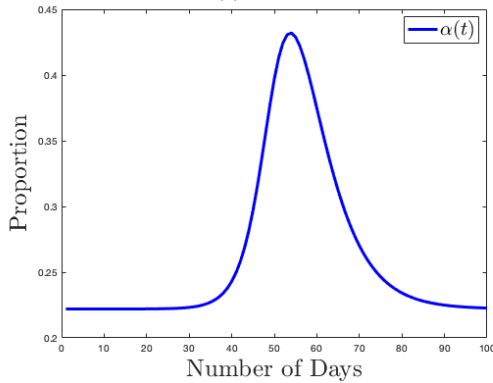
(a) SEIRD Model for an older population.

SEIRD Model with Logistic R_0 for a Younger Population



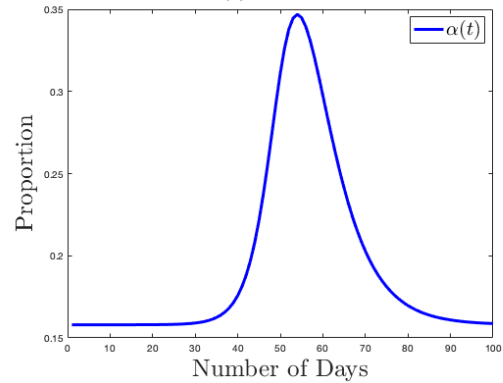
(b) SEIRD Model for a younger population

Fatality Rate $\alpha(t)$ as a Function of Time



(c) Fatality rate as a function of time for an older population.

Fatality Rate $\alpha(t)$ as a Function of Time



(d) Fatality rate as a function of time for a younger population.

Figure 10: SEIRD for combined resource-dependent and age-dependent fatality rates. Populations with higher proportions of younger individuals have a lower overall fatality rate.

In order to implement both resource-dependent and age-dependent dependent fatality rates into our compartmental model, we will use specify the value of α_r from [Equation 6](#) for the value α_{opt} from [Equation 5](#). We will use the same fatality rates and proportions from the previous example. We obtain the results seen in [Figure 10](#).

The results in [Figure 10](#) suggest that the fatality rate is dependent on the age demographics of the population, with the fatality rate being lower for populations with a higher proportion of younger individuals. The effects of this functional fatality rate are shown in [Figure 10ab](#), where there are the maximum value of the dead compartment is lower for a younger population.

Fitting the Model to the Current COVID-19 Data

Now we fit our model to data for the numbers of COVID-19 cases and deaths in the United States. This data was obtained from aggregated data sources provided by the World Health Organization, the European Centre for Disease Prevention and Control, the United States CDC. For a full list of sources and the aggregated data, see [Reference 6](#).

We begin by defining an array C of cumulative cases where the index i represents the i -th day since the pandemic started. That is, if t_1, t_2, \dots, t_n mark the end of the days 1, 2, ..., n , then the array $C = (C_1, C_2, \dots, C_n)$, where C_i is the total number of cases at the end of the i -th day. Similarly, the array $P = (P_1, P_2, \dots, P_n)$ gives a value C_i of individuals who passed away on the i -th day.

In our SEIRD model, our modeled number of cumulative cases is given by the following equation:

$$C_{mod}(t) = I(t) + R(t) + D(t). \quad (7)$$

We do not include the exposed compartment, $E(t)$, because testing in the United States is dependent on the individual being symptomatic. Consequently, the data should not reflect these cases and we would not get a very accurate fit should we include exposed individuals. We define $C^* = (C_{mod}(t_1), C_{mod}(t_2), \dots, C_{mod}(t_n))$ as an array of cumulative cases predicted by the model at times t_1, t_2, \dots, t_n . We define $P^* = (D(t_1), D(t_2), \dots, D(t_n))$.

We also consider the equation

$$diff(C) = (C_2 - C_1, C_3 - C_2, C_n - C_{n-1}), \quad (8)$$

which gives us an approximation to $\frac{dC}{dt}$, the function of people currently infected. From this we obtain three points which we will attempt to minimize the error in our model: total cumulative cases, total deaths, and people that are currently infected.

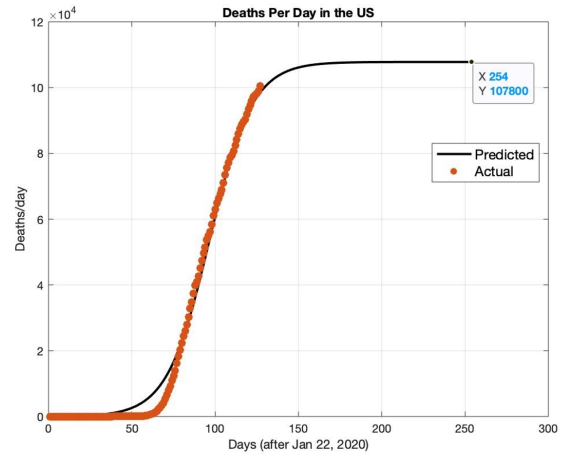
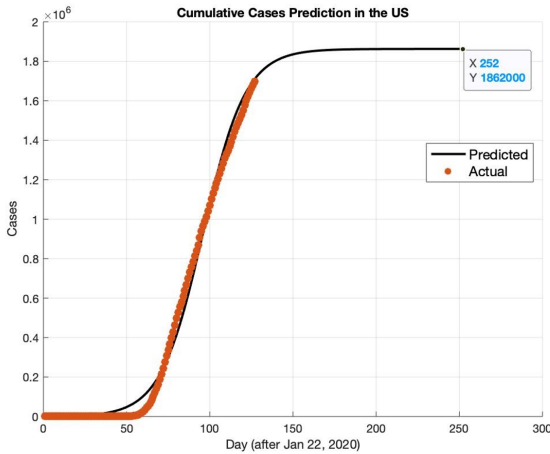
In order to match our model to available data, we defined the objective function

$$F = \|C - C^*\|_2 + \|P - P^*\|_2 + \|diff(C) - diff(C^*)\|_2. \quad (9)$$

Using MATLAB's `fminsearch` routine, we obtained the following for our best fitting parameters:

$$\begin{aligned} \beta &= 1.47 \\ \gamma &= 1.30 \\ \delta &= 1.23 \\ \alpha &= 0.0536 \\ \rho &= 1.41 \\ S_0 &= 8.44e6 \end{aligned}$$

Our results for these parameters are shown in Figure 11, where Figure 6a shows our results for cumulative cases in the United States and Figure 6b shows our prediction for deaths per day in the United States.



(a) Cumulative cases prediction in the United States. (b) Deaths per day prediction in the United States.

Figure 11: The results of fitting our SEIRD model to the number of actual COVID-19 cases and deaths in the United States. These results were achieved using parameters of best fit : $\beta = 1.47, \gamma = 1.30, \delta = 1.23, \alpha = 0.0536, \rho = 1.41$, and $S_0 = 8.44e6$.

The relative error for the prediction of cumulative cases is given by $\frac{\|C-C^*\|_2}{\|C\|_2} = 6.5\%$. The relative error for the prediction of deaths per day is given by $\frac{\|P-P^*\|_2}{\|P\|_2} = 6.8\%$.

Our SEIRD prediction curves are shown in [Figure 12](#).

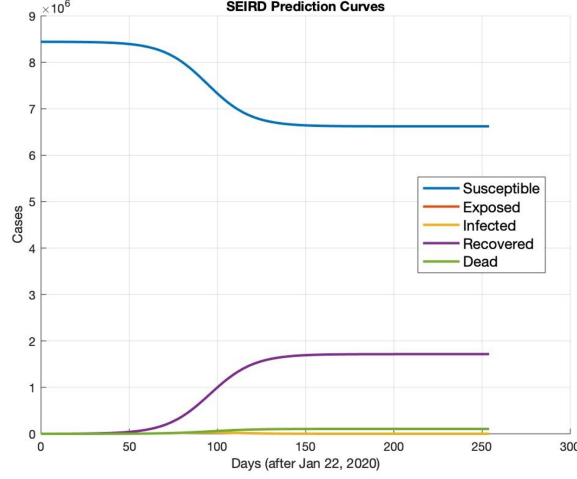


Figure 12: Prediction curves for our SERID model using parameters of best fit : $\beta = 1.47$, $\gamma = 1.30$, $\delta = 1.23$, $\alpha = 0.0536$, $\rho = 1.41$, and $S_0 = 8.44e6$.

Finally, our prediction for cases per day in the United States is shown in [Figure 13](#).

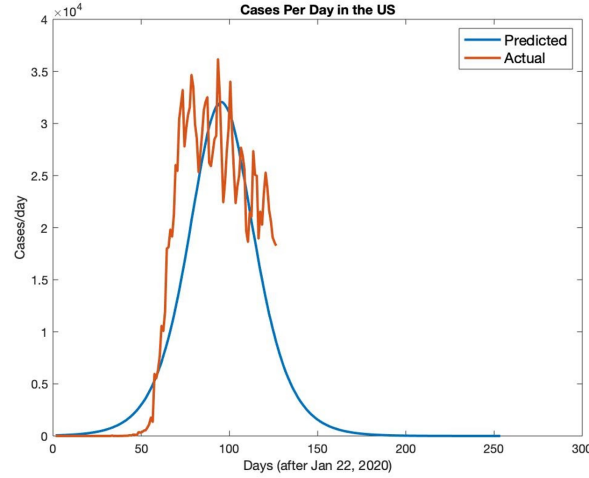


Figure 13: Our predicted cases per day in the United States using parameters of best fit : $\beta = 1.47$, $\gamma = 1.30$, $\delta = 1.23$, $\alpha = 0.0536$, $\rho = 1.41$, and $S_0 = 8.44e6$.

Concluding Remarks

The results obtained when fitting our model to data reflected our assumptions that the spread of COVID-19 is dependent on initial conditions and very sensitive to parameters like the reproductive number R_0 . Though we did not get to use our functional R_0 and resource-dependent and age-dependent fatality rates in our final model, I choose to still include them in the report because they reflect results that I would like to further investigate in the future.

Our results were reflective of the data for COVID-19 in the United States, but with non-constant R_0 and fatality rates, I would consider investigating occurrences of COVID-19 in various regions of the United States rather than collectively. I believe that this will reflect different measures being executed across different states and counties, and pursuing this further would allow us to assess the success of these different measures (i.e. the shelter in place being lifted earlier in some states than others).

Further research should investigate the consequences non-constant parameters on the model for COVID-19.

References

- 1** World Health Organization, 2020. Coronavirus disease 2019 (COVID-19) : situation report, 1.
- 2** World Health Organization. WHO characterizes COVID-19 as a pandemic. Published 11 March 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>
- 3** World Health Organization, 2020. Coronavirus disease 2019 (COVID-19) : situation report, 108.
- 4** Weisstein, Eric W. "Kermack-McKendrick Model." From MathWorld—A Wolfram Web Resource. <https://mathworld.wolfram.com/Kermack-McKendrickModel.html>
- 5** Froese, Henri. "Infectious Disease Modelling: Understanding the models that are used to model Coronavirus." Published 6 April 2020. <https://towardsdatascience.com/infectious-disease-modelling-part-i-understanding-sir-28d60e29fdcf>
- 6** "Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE." <https://github.com/CSSEGISandData/COVID-19>