A Survey of Compartmental Models in Epidemiology

Dawson Collins, Sophie Gee, Sam Goldrup,
Bryant McArthur, Cohen McArthur, Isaac Rowberry

December 5, 2022

Abstract

In this paper, we explore compartmental models of epidemiology with a variety of factors taken into account. We include five main models, all based on the simplified SIR model with parameters based on the reality of COVID-19 spread in New York City from 2020 data. As we add in factors such as death, recurring susceptibility, and vaccination, we obtain more complex mathematical explanations of the phenomenon. We use `scipy.ivp_solve` as a modified Runge-Kutta algorithm to solve the differential equations. As we will elaborate upon during our model presentations and analyses, we were able to come close to certain trends in the statistics surrounding COVID-19 in New York just by using coefficients that we found from the internet. We were able to better understand certain outcomes in isolation as we modeled the virus as factors were introduced such as vaccine, reinfection, death, and levels of contagiousness.

1 Background/Motivation

We are exploring the effects of a variety of factors on the spread of the COVID-19 epidemic over time in the population of New York City in 2020. This problem is important and especially relevant in recent years because of the wide-spread impact that COVID-19 has had on populations and the questions surrounding effectiveness of vaccination and how differing stages of susceptibility/ infection contribute to an overall model of population dynamics.

We start with a base model of SIR and use parameters of interaction and recovery that we learn from statistics provided by the CDC, and then test our models using data collected in `new_york_cases.npy`. Obviously, the vanilla SIR model with a conserved population is not as helpful as we might hope, because reality is much more intricate. When modifying the inputs of our ODEs, we consider real-life situations such as death, vaccination, and multiple stages of infection and contagiousness. Finally, for our most robust model (SIIRSVD), we combine all three introduced factors, maintaining parameters found from CDC statistics. This way, when analyzing our models, we are able to compare our equations to real-life results from the data set.

In terms of greater impact and future applications of our models, we have found certain consistencies between 2020 data from COVID-19 spread and effects with our projections derived from more

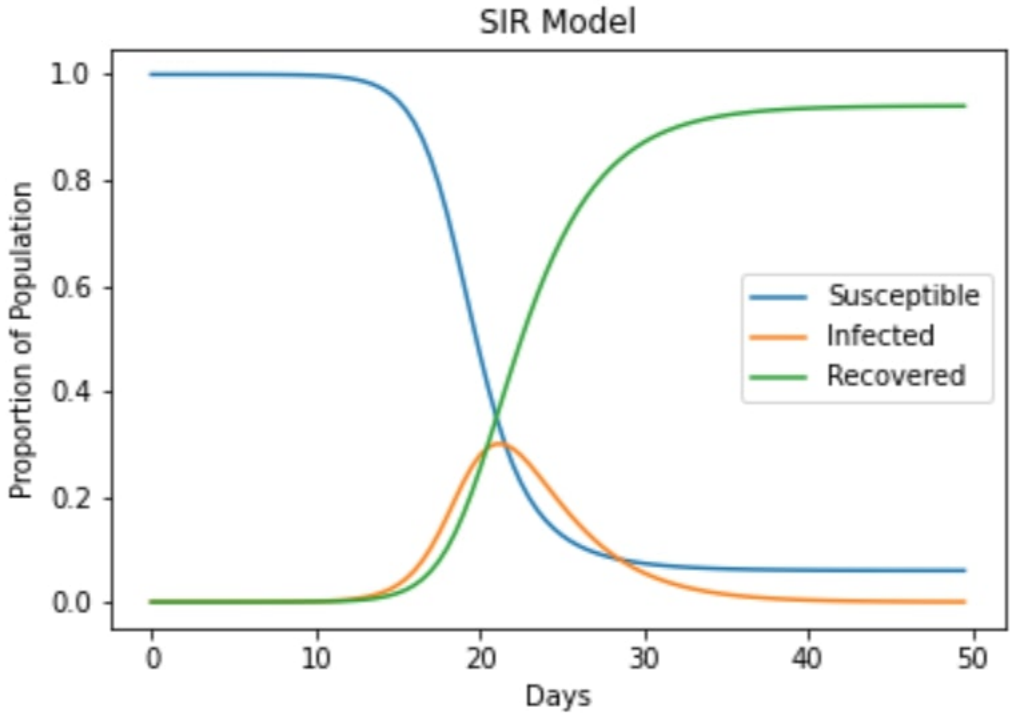
complex versions of SIR models. Because of these similarities, it is plausible to assume that our models may be used for future infections/ epidemics. With a slight change in parameters based on knowledge surrounding the level of contagiousness, death rate, rate of interaction, levels of infection, etc, we can use these models to better understand the severity of and predictions around the path a disease might take.

Before diving into the modifications, we recall the simplified SIR model,

$$\begin{aligned} S'(t) &= -\beta S(t)I(t), \\ I'(t) &= \beta S(t)I(t) - \gamma I(t), \\ R'(t) &= \gamma I(t), \end{aligned}$$

where S, I, R, β, and γ represent the percentage of the population that is susceptible, the percentage of the population that is infected, the percentage of the population that is recovered, the basic reproduction number, and the recovery rate, respectively. For a full derivation of the model, see https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology#The_SIR_model.

A plot of the solution functions for $\beta = 1/2$ and $\gamma = 1/6$ over 50 days, where 5/3000000 people were infected on day 1, is shown below.

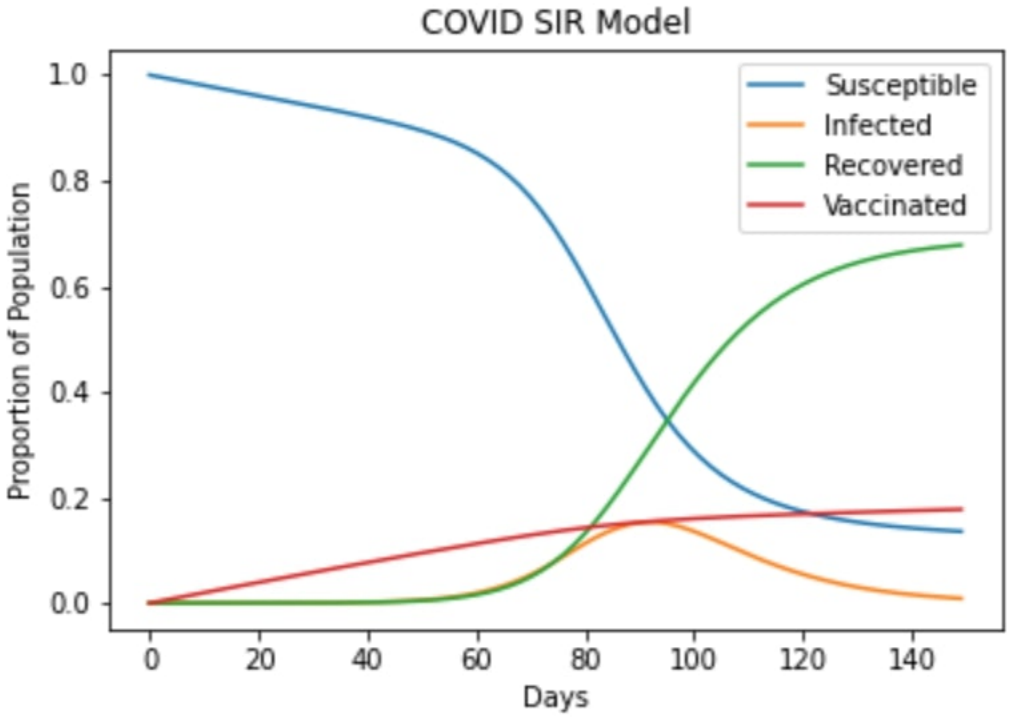


2 Modeling and Results

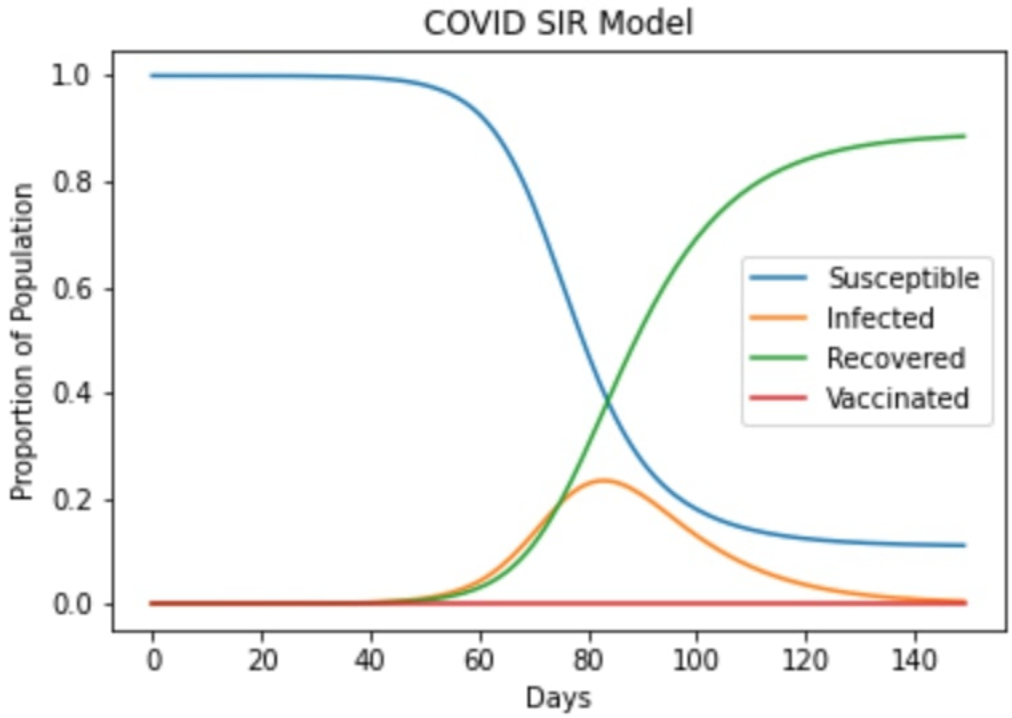
2.1 SIRV

First we add a vaccination term where the susceptible population is vaccinated at a constant rate of .2 percent of the population daily from the start. As we can see our infection curve is significantly

reduced.

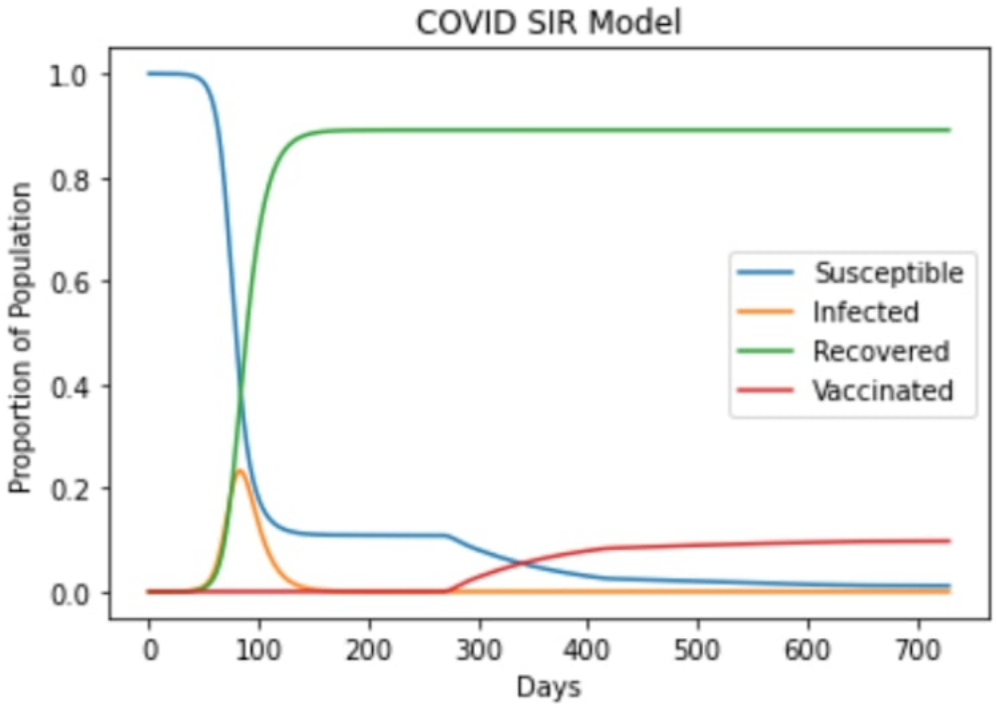


Unfortunately we didn't have the vaccine right from the start. If we introduce the vaccine around day 100 at a slightly higher rate of .4 percent of the population daily we have the following model:



In reality, we weren't able to introduce the vaccine in New York until almost day 300. At which point we had a huge push for vaccinations, some days reached a vaccination total of over 1% of

the population. Later, that rate slowed down as more people became vaccinated. Using these parameters our model now looks like this:



The reason this graph looks inaccurate is because there are still several other elements at play that we haven't implemented into our model. For example, we did not only vaccinate individuals who had never been infected before, "Recovered" individuals could in fact become susceptible again. When we introduce the SIRS model it will enable us to vaccinate a part of the susceptible population who has already been infected by COVID-19.

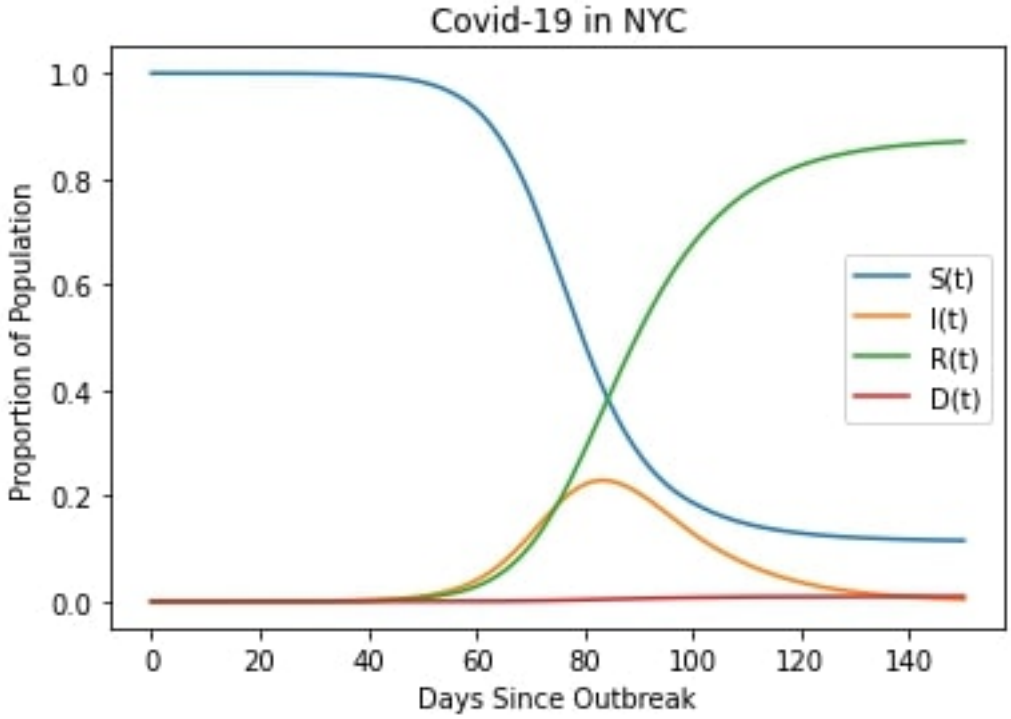
2.2 SIRD

For the SIRD model, we introduce a term $D(t)$ which represents the percentage of the original population that is deceased. The resulting model is given by

$$\begin{aligned} S'(t) &= -\beta S(t)I(t), \\ I'(t) &= \beta S(t)I(t) - \gamma I(t) - \mu I(t), \\ R'(t) &= \gamma I(t), \\ D'(t) &= \mu I(t), \end{aligned}$$

where μ represents the death rate. Following the format of Problem 2 from the V4 SIR lab, we assume that $\beta = 2.5/10$, $\gamma = 1/10$, and that there are approximately 8.399 million people in New York city. We solve the system for 150 days, starting on March 20th, 2020, where the number of cases on the first day is 52. According to https://coronavirus.jhu.edu/data/mortality, the death

rate for the U.S. is about 1.1%. Thus, we set $\mu = 0.011/10$. We have



Although it is difficult to tell from this plot, the percentage of the original population that is now deceased is about 0.0096, which translates to roughly 80443 people. Unfortunately for our model (but fortunately for the residents of NYC), the file `daily_data_NewYork.csv`, obtained from https://www.kaggle.com/datasets/ruchi798/a-tale-of-two-cities, lists the actual number of deaths in our time period as 18619.

2.3 SIIR

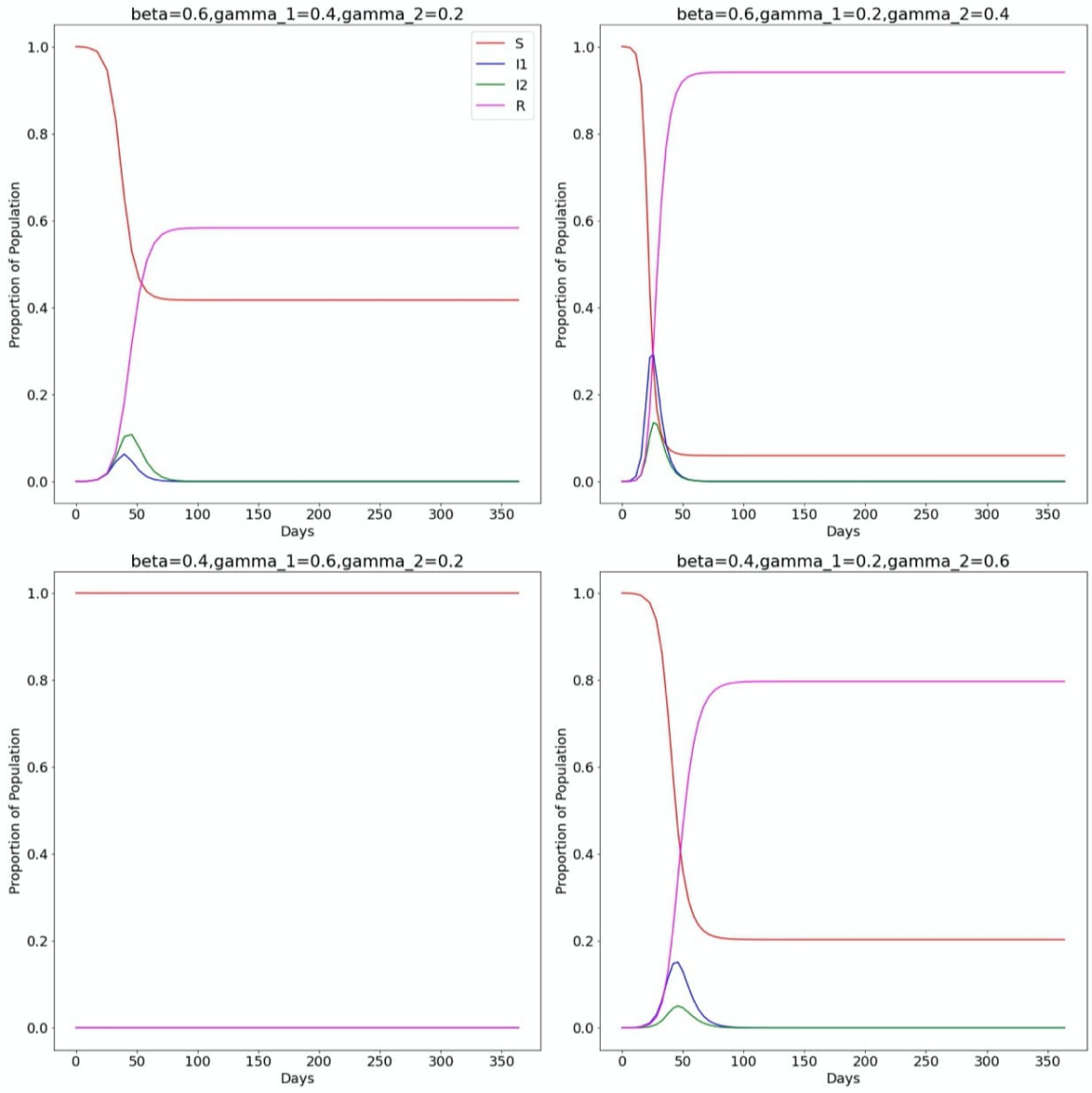
The SIIR model explores an environment where the "Infected" compartment is divided into two phases. Typically the first phase is thought of as a contagious phase and the second is no longer contagious. This model is a step in the right direction for modeling COVID-19 before the vaccine was developed.

Unfortunately, data was not found separating infection into two different states. In place of data matching, we conduct some interesting analyses. We conduct bifurcation analyses with an exploration into quarantines as well as perturbation analysis.

We first experiment with bifurcations in the SIIR model. The simulations for this model assume an initial value of 1000 infected symptomatic (contagious) individuals in the population of New Yorkers on day 1 of the 2020 pandemic. With two infection compartments, there are two infection coefficients. γ_1 is the person per day rate of transfer from the infected and contagious state to the infected and noncontagious state, and γ_2 is the rate of transfer to the recovered state.

This model assumes that it is impossible to contact the virus for a year, which may be a strong assumption depending on the number of people who contract COVID-19 twice in the same year.

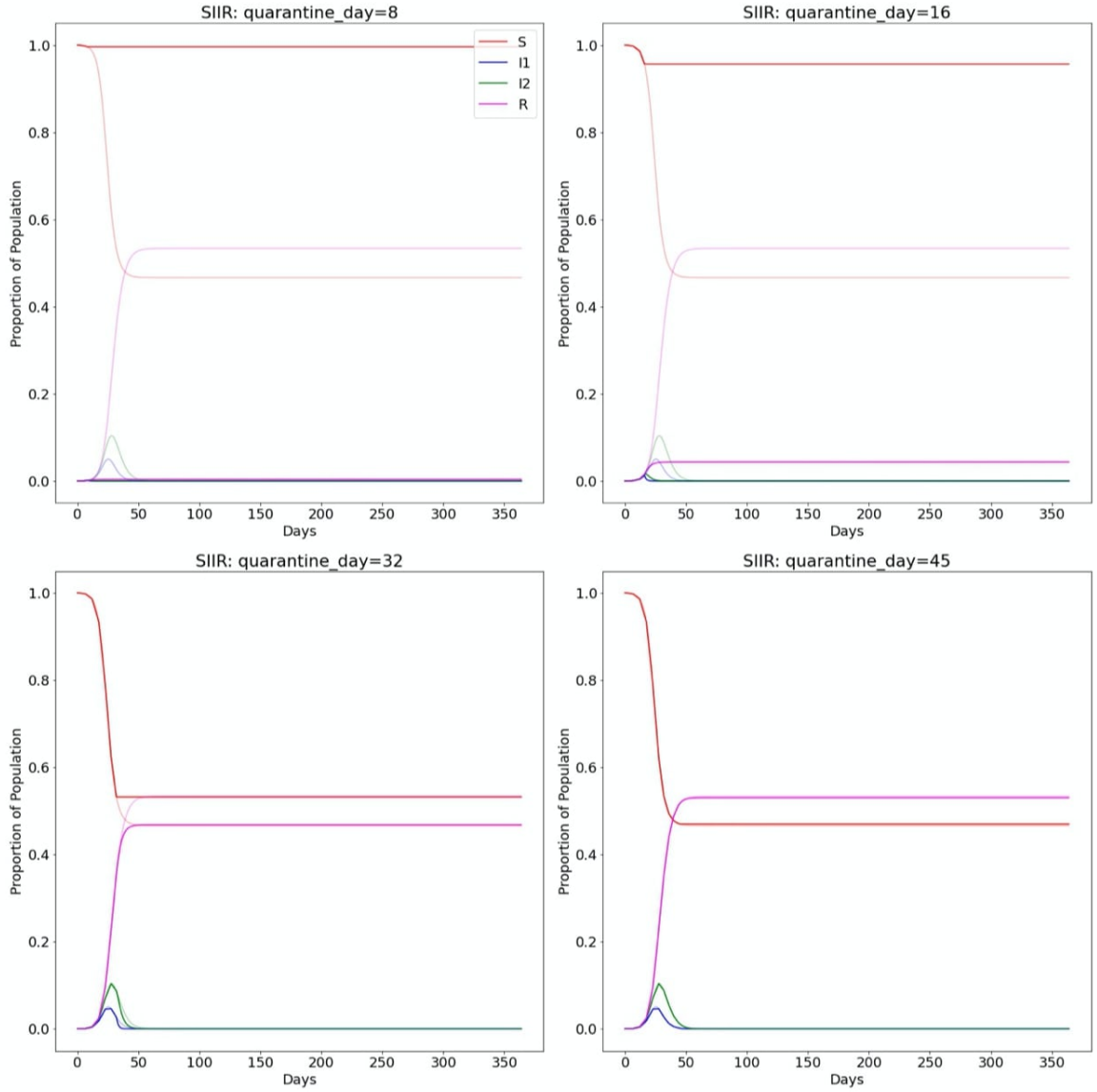
However, in all of the simulations, a steady state is reached well before the first quarter of the year.



Looking at the graphs above, we see some interesting bifurcations from altering our interaction coefficients. We see that the infection coefficient must be larger than the first recovery coefficient for there to be a steady state population that is not entirely susceptible people. Otherwise, recovered people begin to comprise the steady state. The steady state seems to be entirely affected by the relative magnitudes of the parameters

It is worthwhile to see if a change in a parameter at a certain point in time can cause changes in the steady state. To explore this, we reduce the infection coefficient ten fold at different days in the year long simulation. This is analogous to a quarantine. This means that instead of 1 spreading per day by the infected and contagious New Yorkers, there is 1 every ten days. The γ_1 and γ_2 coefficients are fixed at 0.7 and 0.3, respectively, reflecting the modeler's assumption that recovery

is faster in the first phase than in the first phase.

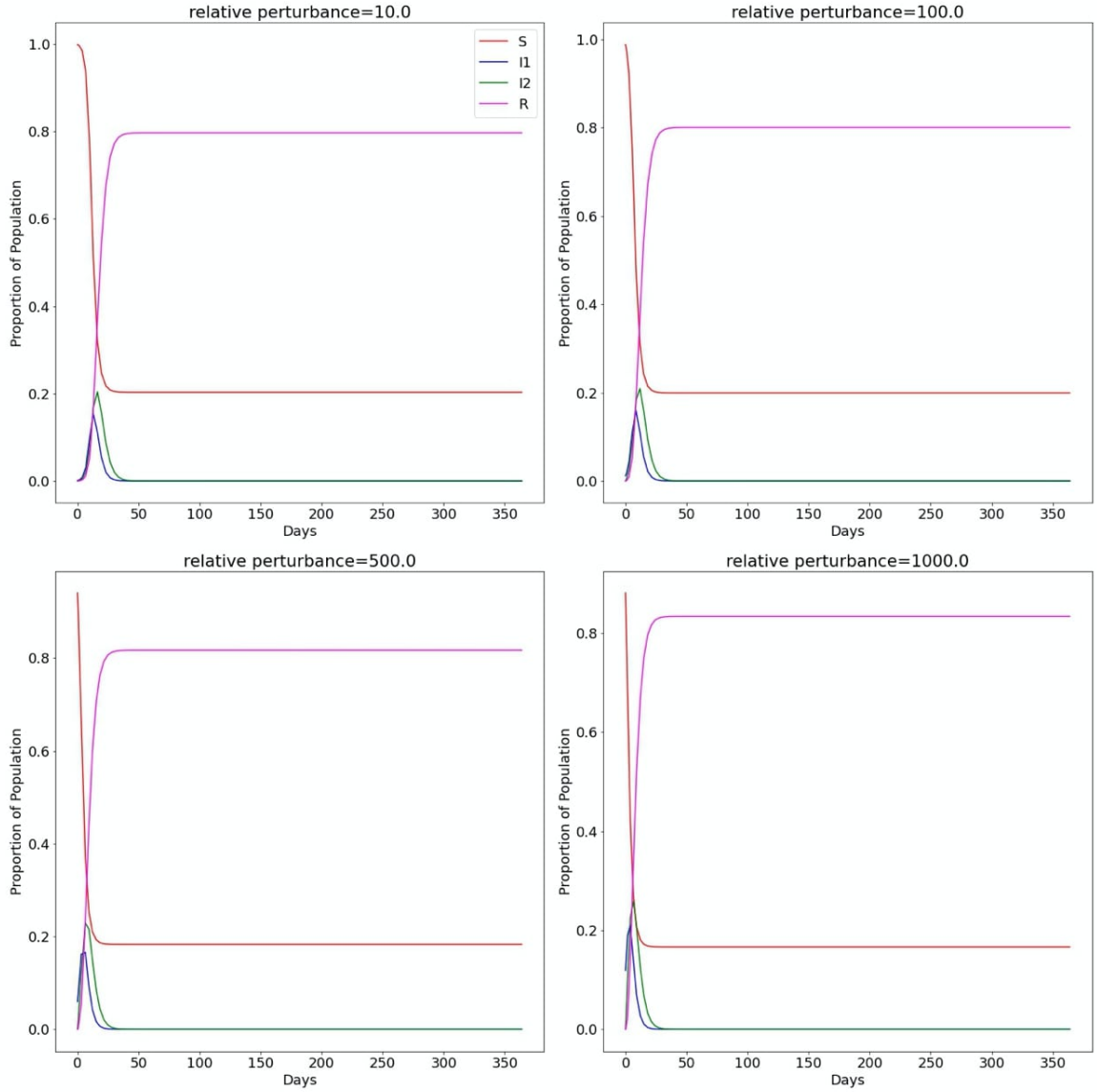


The solid colored lines are the results with a quarantine, and the faded lines are a counterfactual where no quarantine was imposed.

Imposing quarantines in the first 10 days of a disease spread has a large impact, as evidenced by the divergence of the faded lines from the more strongly colored ones. Between day 10 and 32, the marginal impact of imposing a quarantine one day later decreases because the gap narrows. By day 45, imposing a quarantine seems to have no detectable impact on the steady state. 45 is a small enough time window for some inefficient governments to fail to take action fast enough. In addition, our bifurcation analysis suggests that it is also worthwhile to consider the interaction coefficients before considering to impose a quarantine. One major misgiving the modeller has about this quarantine analysis is that eventually a quarantine must be lifted, and post quarantine dynamics need to be analyzed. However, since the vaccine took around a year to become available, and since many cities like New York had the opportunity to impose aggressive policies against the COVID-19 spread, it is not an entirely unreasonable assumption that the quarantine lasts for the remainder of

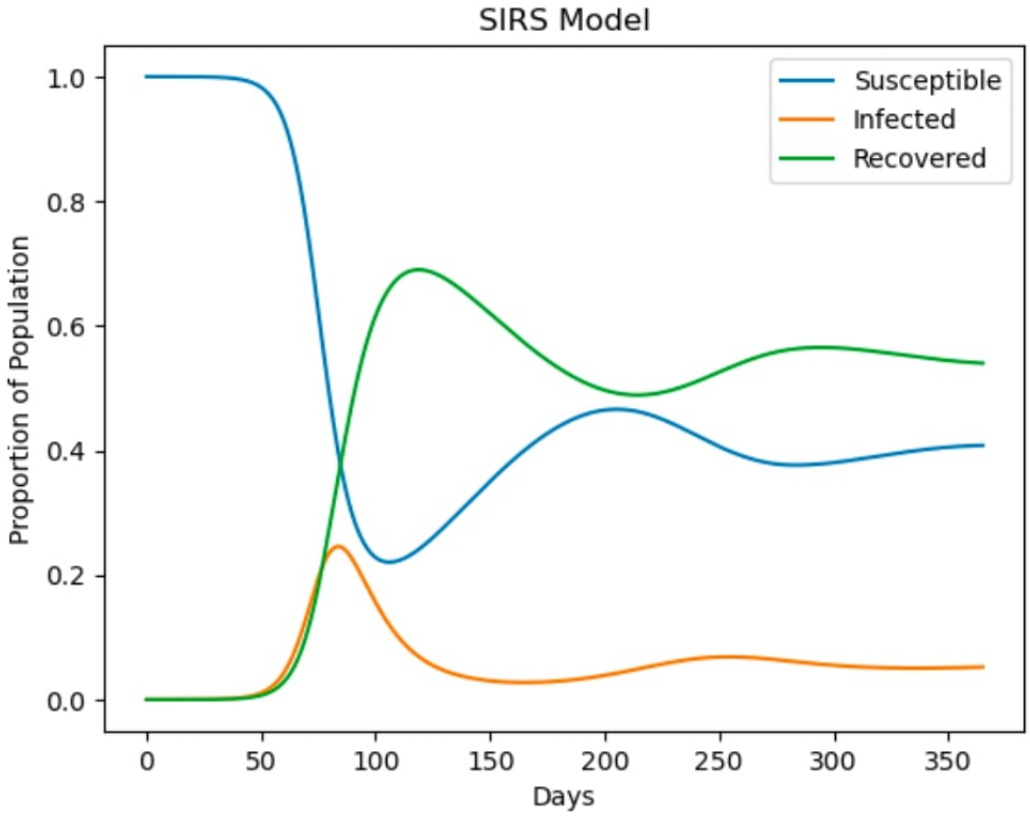
the year.

The model appears to be robust to large changes in the initial number of infected people. Simulating on coefficients of $(\beta, \gamma_1, \gamma_2) = (1, 0.5, 0.3)$, we see that the steady state shifts once the initial value of infected people is scaled up by 500.



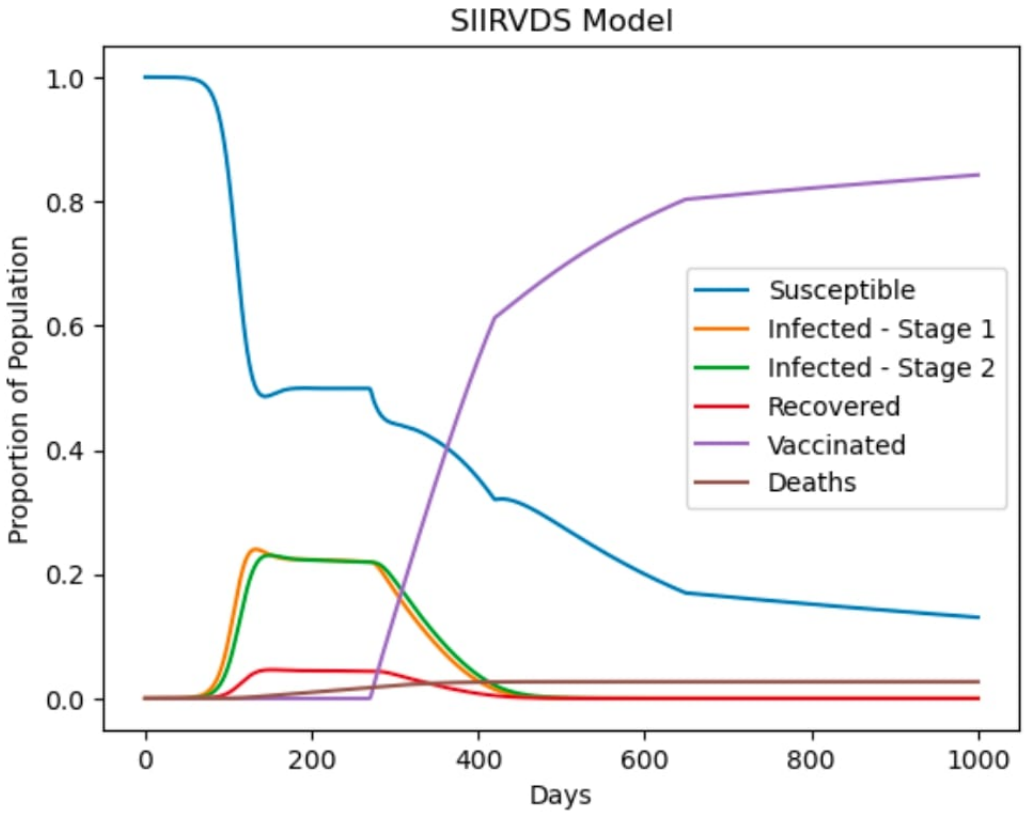
This suggests that the model is quite robust to changes in the initial number of infected New Yorkers. This means the model can scale to a good degree to other cities of similar size and varying initial levels of infected individuals.

2.4 SIRS

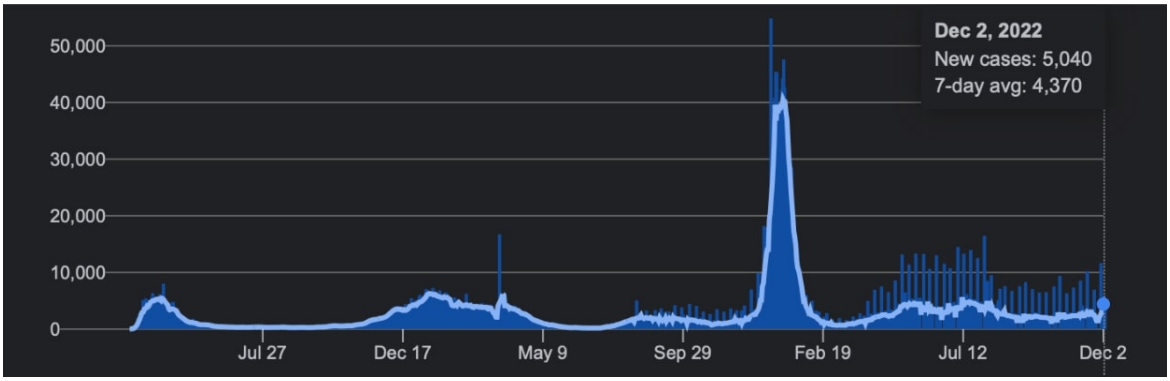


The SIRS model accounts for the fact that not all diseases give you 100% immunity for the rest of your life. People have the ability to lose their immunity, moving them back to the population that is susceptible. We can see from this graph that each respective line has the behavior of a damped oscillator. Eventually the number of people that are either susceptible or recovered overwhelms the rate the virus can spread and it is eventually eradicated from the population. This graph was using an infection rate of 2.5 people per 10 days, a recovery rate of 1 person per 10 days, and a reinfection rate of 1 person per 100 days.

2.5 SIIRSVD



This model incorporates many different models. It has several different populations, as seen in the graph above. The majority of work in this model was finding the coefficients that fit the real word example of New York City. There are several coefficients that are important: infection rate, recovery rate, vaccination rate, movement of Infected from Stage 1 to 2 rate, and death rate. For the infection rate we used two different stages to try to replicate the real life data, however due to factors (social distancing and/or lack thereof, masks) we did not put into our model it doesn't match.



We also set up vaccination rates in stages, as it wasn't available for the first 250 days of the pandemic, then followed by an aggressive vaccination plan put in place by the government, then it slowly tapers off. Due to this we can see jumps in the other populations as well. Our end vaccination

percent (.842) is close to that of New York's actual percent of .786. Another important thing to note is that our end death total (.026) is significantly higher than New York's (0.005)

3 Analysis/Conclusion

In conclusion, our various models achieved some success in representing the trajectory of COVID-19 in New York, but each fell short with respect to one factor or another. For example, while our SIRV model was able to accurately estimate the initial effects of the vaccine on population, it failed to provide long term displays of population growth in any accurate way. Similarly, our SIIRVDS model was able to come close to the number of people that got vaccinated in New York, but it overestimated the rate of COVID-19-related deaths.

For the SIIR model, we gain important insights on timing of a quarantine. Answering questions of how early the quarantine needs to be enacted, and how effective it will be requires measurements on rates of interaction between different compartments. In the future, we hope to extend this result to more complicated models that incorporate vaccines, such as SIIRVDS. While we focus on modeling the COVID-19 dynamics of New York City, a perturbation analysis done in the SIIR section gives hope that SIR models and their variants can scale nicely to other situations. Again, this is a result we hope to extend to more complicated models.

If we had more time, we might be able to more successfully hone in on these shortcomings and understand which coefficients were hindering the model and why. We could use permutations of the variety of factors we modeled to see which combination highlights the most important features and most accurately displays the lifespan of the pandemic.

Through modifying the SIR model, we learned that adding more features to a model doesn't always make it more accurate in modeling complex systems. Even though the more the features, the more realistic an environment might be, more features also introduce a greater variation in the behavior of the model.

If we had successfully modeled this phenomenon, we could use this information and its predictive nature to implement more effective preventative measures at a government level. For instance, if we found that identifying various stages of the infection was the most indicative of how fast the disease might spread, we could prioritize understanding the periods of infectiousness and use this as a more prominent metric to differentiate between strands of viruses.