Epidemiological Compartmental Models in Python

Completed as the final project for CAAM 37830 at UChicago

Luke Chen, James Keane, Song Liang

December 2020

1 Abstract

Compartmental models such as the SIR model are commonly used within epidemiology to model the spread of a disease. Depending on the nature of the disease and the goal of the research, certain extensions may be desirable to make the model more realistic. In this paper, we implement various extended versions of the SIR model using systems of differential equations and agent-based modeling in Python. Simulation results derived from our implementations are presented and discussed. Key results are: (1) location and mobility of initial spreaders can have a large impact on the trajectory of the disease, (2) the disease can spread even when people limit their interactions with strangers as long as close contact still occurs with those close to them, (3) interventions such as mask-wearing, vaccination, pervasive social-distancing, and quarantining infectives are very effective in "flattening the curve" of infectives over time. Our simulations also demonstrate that differential equations and agent-based models yield very similar results over large populations and longer periods of time.

2 Introduction

The SIR model is a compartmental epidemiological model which assumes everyone in a population falls into one of three compartments: Susceptible, Infectious, and Recovered (sometimes referred to instead as Removed). The Susceptible population consists of those who have not yet gotten the virus and thus are *susceptible* to becoming infected. Infectious individuals are those who carry the virus and are capable of infecting others for some period of time before eventually transitioning into the Removed group. Lastly, Removed individuals are those who are no longer part of the system because they have either acquired immunity or died from the disease.

The trajectory of the disease generally depends on the parameters b, the per capita number of interactions capable of spreading the disease that occur per day, and k, the rate at which Infected individuals move into the Removed compartment. Notably, the basic SIR model assumes that all agents in the population interact randomly so every Susceptible individual shares an equal chance of becoming infected in each period. I.e. the probability for any given Susceptible individual to become infected in a period is $b * \frac{I}{N}$. Additionally, [1]. For Covid-19, no conclusive figure has been established as the mean duration of infection. However, the US CDC estimates that the majority of cases become no longer infectious within 10 days, though some serious cases can remain infectious for up to 20 days [2].

From the discussion above, we can note that for any period, we can predict if the infection will continue to spread or begin to decrease in the next period. Since the (expected) number of new infections can be given by $S * b * \frac{I}{N}$ and the (expected) number of removals is given by $k * I$, if the ratio of the two, $\frac{S*b}{k*N} > 1$ then the infection will continue to spread since the number of new infections outpaces the number of recoveries; whereas if $\frac{S*b}{k*N} < 1$, the number of recoveries exceed the number of infections and so the infection will begin to die out on its own. Within epidemiology, this relationship is commonly represented through the 'effective reproduction number', $\mathcal{R}_e$ [3].

The SIR model lacks many aspects that may make it more intuitive and closer to reality. First, we may want to add a spatial component to the model so that we can ascertain insights about how the disease evolves

not only time but also geographical location. Second, we add a concept of cohorts to the agent-based version of the model so that the nature of social interactions is more reflective of reality where interactions are less random than posited by the model. Third, we implement methods for modeling the role of interventions (i.e. mask usage and vaccinations) using agent-based modeling to examine the potential of such interventions to alter the trajectory of the disease. Lastly, we implement additional classes to represent ODE models with additional compartments such as the "Exposed" compartment.

3 Variants

3.1 Spatial Component

3.1.1 Spatial PDE Model

To implement a theoretically simulated spatial SIR model, we updated our ODE model into a PDE model. The system of PDEs is:

$\partial s(x,t)/\partial t = -b * s(x,t) * i(x,t) + p * L * s(x,t)$

$\partial r(x,t)/\partial t = k * i(x,t) + p * L * r(x,t)$

$\partial i(x,t)/\partial t = b * s(x,t) * i(x,t) - k * i(x,t) + p * L * i(x,t)$

Here we have a new parameter p which is to weight the diffusion term. A larger p means each compartment diffuses quickly into space.

Based on our earlier findings in our vanilla SIR model, we set b=3 and k=0.1 when testing p's impact on our spatial model.

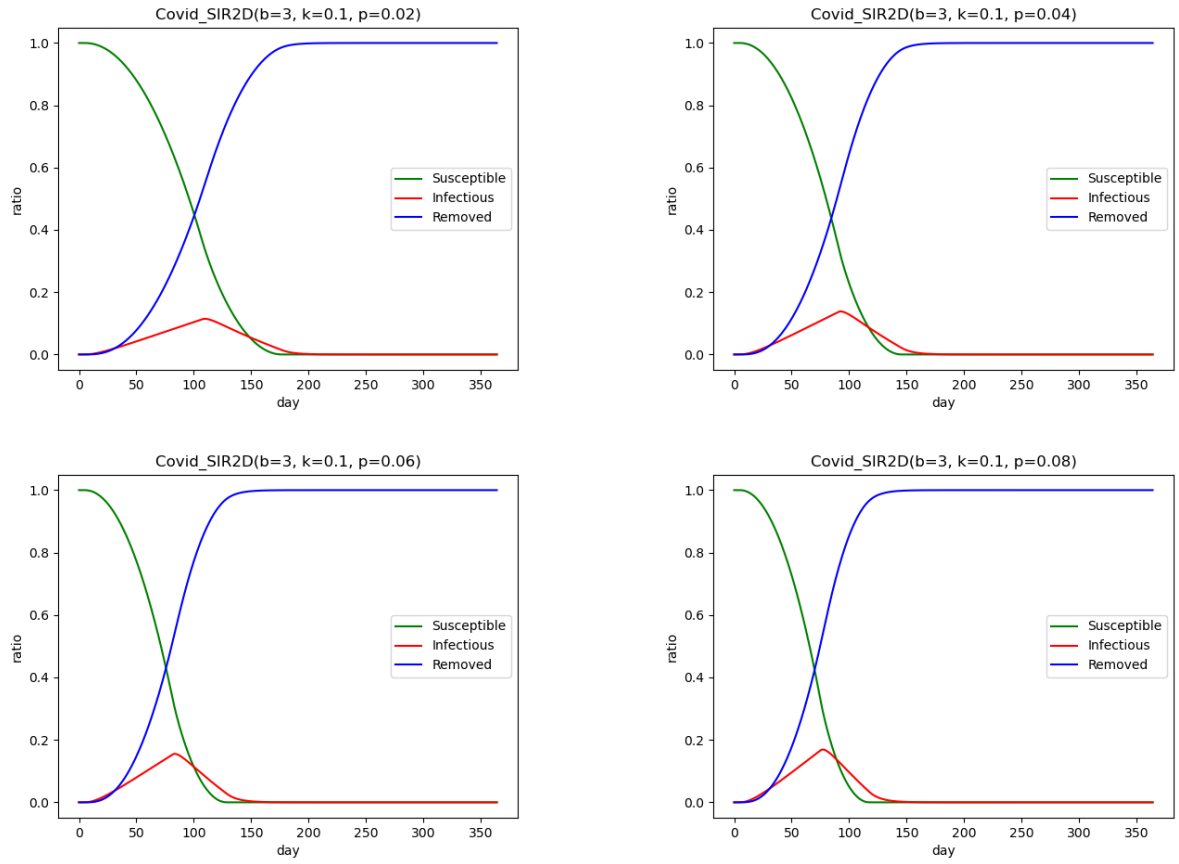

Figure 1: Centrally Initiated Infections Simulated at Multiple p's

From Figure 1, we can find that the recovered population of all plots approximates to 100% as t gets large. This means for different p all population will eventually be infected. However, the ratio of total infected population (recovered and infectious compartments) develops at different speeds. When p is large, the trajectory of spatial SIR model is similar to SIR model. This is because large p means a person can interact with more people, so when p is large enough the trajectory of spatial SIR model would be very similar to that of a non-spatial SIR model.

To differentiate from the vanilla SIR model, we will choose a p of 0.1 to use in our simulations.

By setting *i(x,0)* appropriately, we can investigate how the location of infection outbreak affects our simulations.

By setting *i((0,0), 0)* to be nonzero but other i to be zeros, we can simulate diseases spreading started in a single corner of the square. Likewise, setting *i((M/2,M/2), 0)* to be nonzero but other i to be zeros, we can simulate diseases that begin in the center of the square. Lastly, we can randomly select several points of i to be nonzero, with others being zero, to simulate a randomly spread out disease initialization.

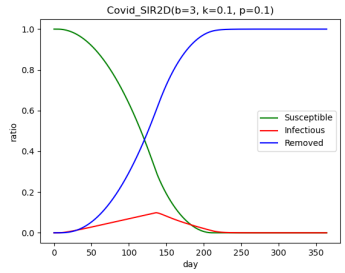

Figure 2: Corner Initiation

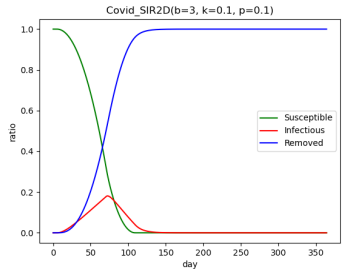

Figure 3: Center Initiation

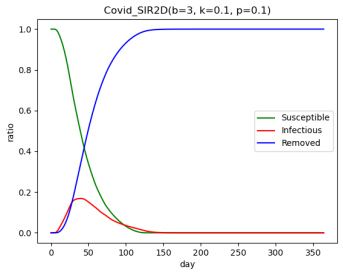

Figure 4: Random Initiation

From these simulations, we can infer that diseases initialized from the corner will take a longer time to develop. The ratio of illness population at one time will be small or the curve be "smoothed". This 'square world' will have a longer time to prepare. When the disease starts at the center of the square, it will take an intermediate time for the disease to spread out. The quickest disease development arises when the outbreak happens in multiple, random, locations.

3.1.2 Spatial Agent-Based Model

In addition to our PDE model, we also implemented a spatial component into our existing agent-based simulations. Upon initialization, agents in our spatial simulations were given two dimensional coordinates to designate their starting position contained within the unit square.

The daily steps of the simulation remain largely the same, with a specified proportion (k) of those infected being removed with the remaining infected interacting, and possibly infecting, the rest of the populace. Rather than designating a set number of interactions for the infected, each agent in the spatial simulation will move once in a random direction each day and interact with whichever agents happen to be where they end up in the unit square. Infected agents will infect whoever they reach and susceptible agents will get infected if they end up near a diseased agent. Recovered agents will not move at all as they cannot influence any part of the simulation in that state.

Daily movements will be scaled by a new parameter (p) and their net of interactions will be determined by the circle with radius q which will effectively replace our previous parameter b. Any daily movements that would see the agent exceed the bounds of the unit square will have them instead remain in the same location for the day.

Based on the earlier SIR model, we've elected to choose $k = 0.1$ and $q \approx 0.02$ to determine the effect of p on simulations. The latter was determined based on the below rough relationship with b:

$$q = \sqrt{(\frac{b}{(N * \pi)})}$$

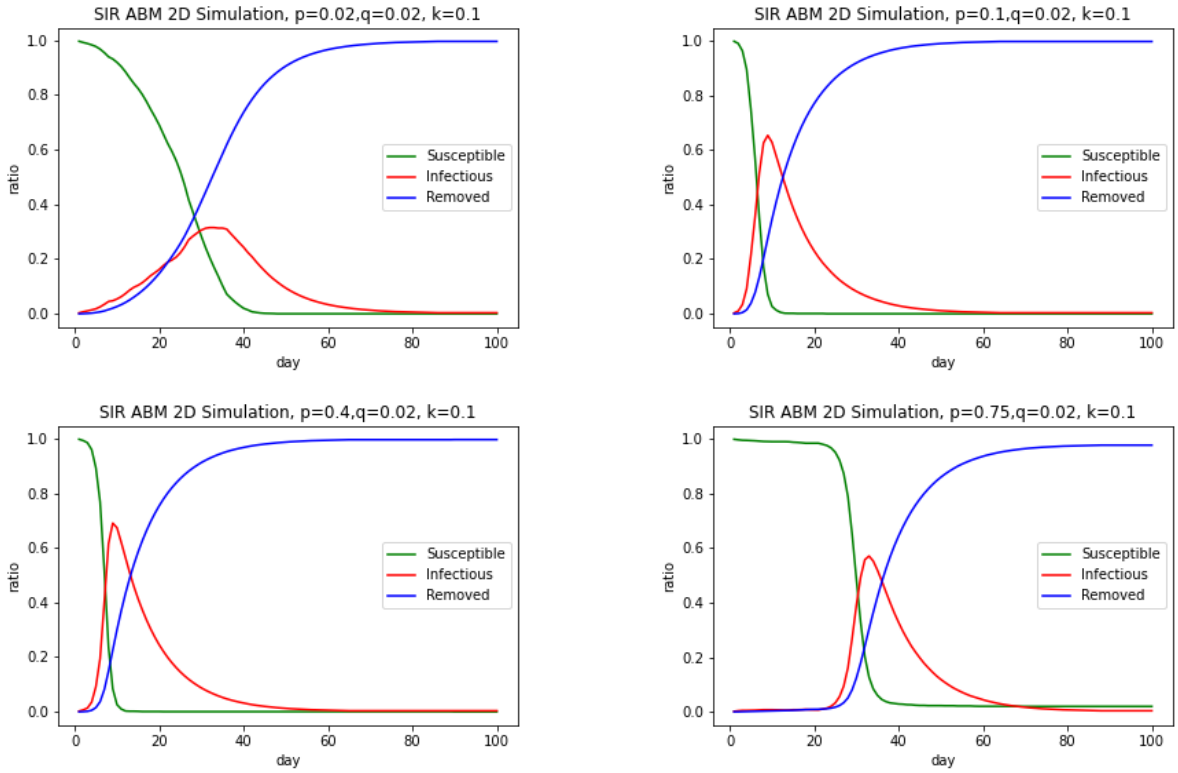

Figure 5: Centrally Initiated Infections Simulated at Multiple Step Sizes

Similar to our PDE model, we see that no value of p is completely preventative of forever delaying a completely infected populace. However, we can note a few differences in the agent-based simulation's relationship with p.

From Figure 5 our slowest growing diseases come from our lowest and highest values of p. This comes from the infected simply making very meager progress across the unit square, or from making such large moves that they exceed the boundaries of the square and do not end up moving. A middle ground sees infected able to cover large parts of the unit square without often stepping out of bounds.

To identify the relationship between our spatial component and the duration of our epidemic, we then can look at different simulations where the infection starts in a Corner, the Center, or at random points across the unit square. For a more moderate virus expansion, we've selected a step size of $p = 0.02$.

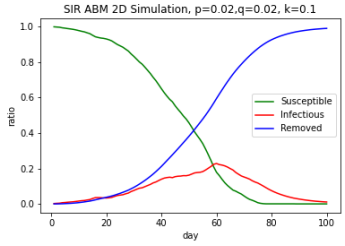

Figure 6: Corner Initiation

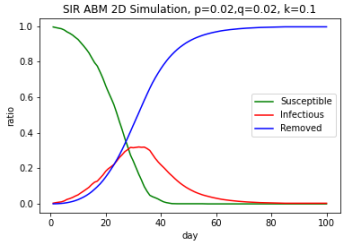

Figure 7: Center Initiation

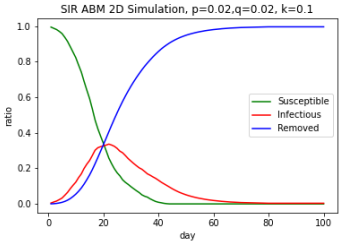

Figure 8: Random Initiation

The agent-based figures largely compare well to the PDE counterparts:

Corner-based diseases are the slowest to perpetuate across the unit square. Even with a small step size, infected agents are very likely to overstep the two edges they start next to which could greatly impede

disease progress. Center-based diseases are faster, simply as they are unlikely to overstep any edges of the square at such a low step size. However, having all of the infected start in one location delays the spread a little bit. Randomly initialized diseases have the easiest time spreading as the infection can start in different communities of the square from the beginning.

3.2 Cohorts

3.2.1 Background

One major procedure for stemming the spread of an epidemic virus is to cut out all unnecessary close contact with infrequent social circles. For example, part of the province of Alberta's COVID-19 Relaunch guidelines was the insistence that social interactions without the proper safety measures (e.g. mask wearing and social distancing) to be limited to a core cohort of individuals[1]. This kind of group often is established as a household or an extended household. Done effectively, a cohort helps insulate your social circle from the virus, and helps contain the virus within your social circle if someone within is infected.

3.2.2 Implementation

Our initial agent-based model considered interactions to occur randomly across the populace. For this extension, we will instead be considering all b interactions to occur within an agent's cohort. The last of these b interactions will have a chance to occur outside the cohort, or non-locally. We will label this outside b chance as the parameter ob.

To loosely identify cohorts, our agent population will be assembled into a two-dimensional, $N\ x\ N$, grid. The fellow agents immediately surrounding an agent will be considered its local interaction candidates. For most agents this will be a group of 8, whereas for the agents on the edge of the grid this could be as low as 3. The simulation will occur in the same format as our vanilla SIR ABM, but with the b interactions affecting only direct neighbors, with the aforementioned chance ob that an interaction occurs with a non-neighbor.

3.2.3 Results

With the below phase diagram, we can see how our disease spread changes based on the value of ob. To judge it in comparison to our initial SIR model, we'll use the consistent parameters b=3 and k=0.1.

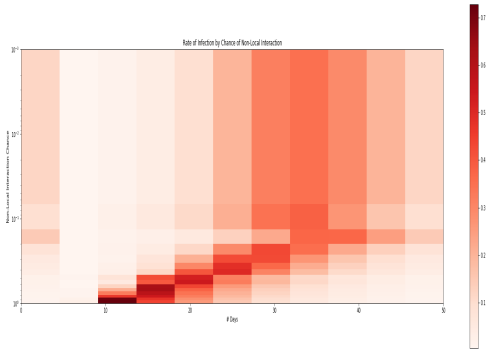

Figure 9: Phase Diagram of Non-Local Interaction Chance by Day. Color corresponds to population infection rate.

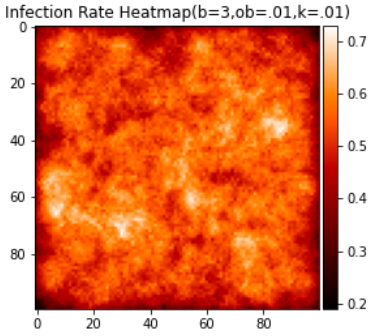

Figure 10: Heat Map of Infection Rate by agent over 100 simulations.

[1]Link to Alberta cohort guidance document

Largely, the implementation of cohorts to the model dulls the spread of the virus. Most simulations do not reach high levels of infection until day 35. Higher values of *ob* lead to faster rates of infection more in line with the spread we saw in our vanilla SIR model. Even in spite of cohorts, we still eventually the infection hitting the population heavily within 50 days.

In addition to seeing how the cohorts affect the spread as a whole, we can also gauge how it affects certain agents differently.

From Figure 10 we can see that the most susceptible agents lie in the central band of our population grid. These individuals have the same number of agents in their direct cohort, but are also closer to every other cohort on the grid than the majority of other agents. Contrarily, those on the edge of the grid have more degrees of social separation than the average agent, and thus have the lowest chance of infection.

The issue with using direct neighbors as a cohort is that only implements loose cohorts. It is somewhat helpful in stemming the disease spread, but if everyone in an agent's cohort is also part of separate cohorts, there exists a maneuverable path for the disease to spread between any two agents. Part of the province of Alberta's stricter guidelines was to insist that citizens participate in one single core cohort. That is, the same people should be in a singular small group. This distinction separates cohorts that consist of households to those that consist of households and additional social relations.

To model a stricter variation of cohorts, we instead assigned every agent in our population to a group of 9 individuals. Now instead of local interactions occurring with their neighbors, it occurs only with neighbors that have the same given cohort number.

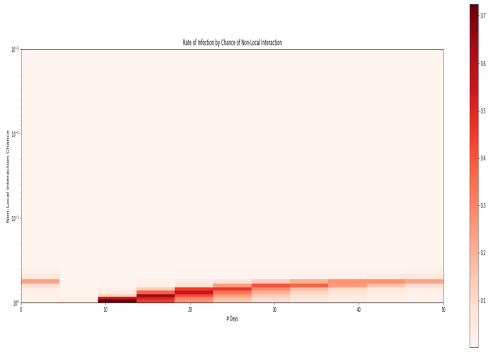

Figure 11: Phase Diagram of Non-Local Interaction Chance by Day. Color corresponds to population infection rate.

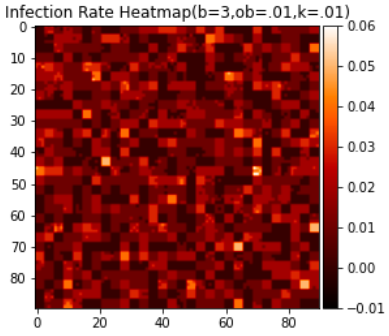

Figure 12: Heat Map of Infection Rate by agent over 100 simulations.

The non-local interaction chance becomes drastically more important in producing high levels of infection. As viewable in Figure 11 When *ob* is low, infection does not get high in the first 50 days. This stricter cohort enforcement is far more effective than the looser guidelines initially provided.

Further, Figure 12 is a lot less smooth as infections that occur happen and stick within a certain cohort block. We can see that, in general, very few cohorts end up hitting a concerning rate of infection. Also of note is that anytime a cohort member does get infected, the rest of the 9 members generally follow suit.

3.3 Viral Spread Interventions

3.3.1 Background

As we have seen with Covid-19, interventions are often sought out and employed to limit the spread of the disease. Most notably with Covid-19, mask usage has become a relatively wide-spread practice, and is even oftentimes officially mandated. The motivation for mask-wearing is that it reduces the emission of respiratory

droplets which are responsible for the spread of the virus. Generally, mask wearing by infected individuals is thought to reduce their risk of infecting others by about 50% - 80%; mask wearing by susceptible individuals is thought to reduce their risk of being infected by about 30% - 50%. Similarly, vaccines are just now starting to be implemented as a method of controlling the spread of the virus by immunizing the susceptible population. Imminent vaccines have an estimated efficacy rate of around 95%, meaning immunized individuals are 95% less likely to catch the virus conditional on exposure [4]. These values inspire the parameter values set for the simulations.

3.3.2 Implementation

This extension is implemented by adding functionality to the methods available to the agent-based modeling (ABM) methods available in `sir.abm`. Specifically, it introduces the *masked, infectivity,* and *risk* attributes to each person. Respectively, the attributes specify if a person wears a mask or not their relative infectivity (if they are infected) and relative risk of receiving the disease (if they are susceptible). Functions are available that update the population by making m proportion of the unmasked population wear masks and u proportion of the masked population lose their masks. Additionally, the effectiveness of the masks can be input through the *infectivity* and *risk* arguments. Simulations incorporate updating of mask usage with constant parameters in each period.

3.3.3 Results

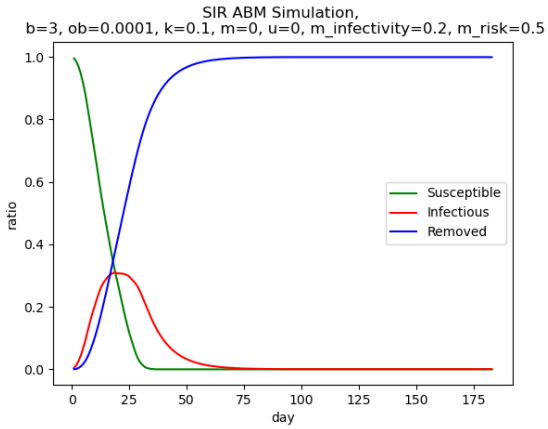

Figure 13: ABM results with no interventions on a population of 10,000 with 15 initial infected

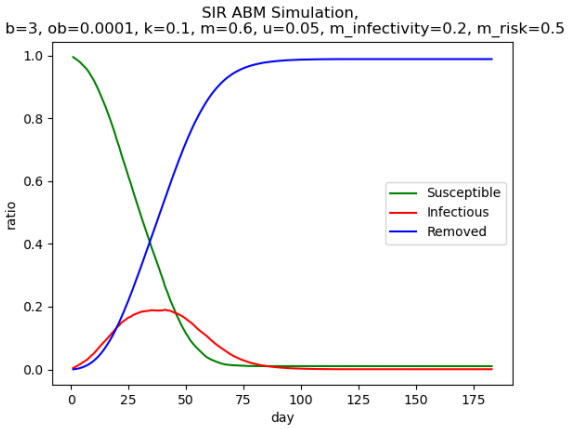

Figure 14: ABM results with mask usage on a population of 10,000 with 15 initial infected

Figure 13 provides the baseline simulation when there is no spread intervention within a population. Figure 14 shows the simulation with all the same parameters except for 60% of the unmasked population dons masks each period and 5% of the masked population stops wearing masks. We can see that the peak number of infections is substantially lower with mask usage and the incidence of the disease is more spread out over time. In real life, this would translate to less strain on the health system and thus lives saved. However, ultimately, the entire population eventually moves into the removed category, indicating everyone either gets the disease and gains immunity or dies. Figure 15 shows what would happen if a vaccine (with 95% efficacy) is introduced when a large portion of the population is already infected as is the case right now[2]. We can see that even if almost 20% of the population is infected at the start, almost 20% of the population manages to make it through the course of the disease without ever catching it. In real life, this would result in many lives being saved. Additionally, we can notice that the duration of the disease is shortened substantially as it becomes largely non-existent within 50 days.

[2]The graph may seem bad because around 50% are infected at the peak, however, Figure 20, found in the Appendices, shows that just by virtue of a high number of initial infectives, a high peak will result.

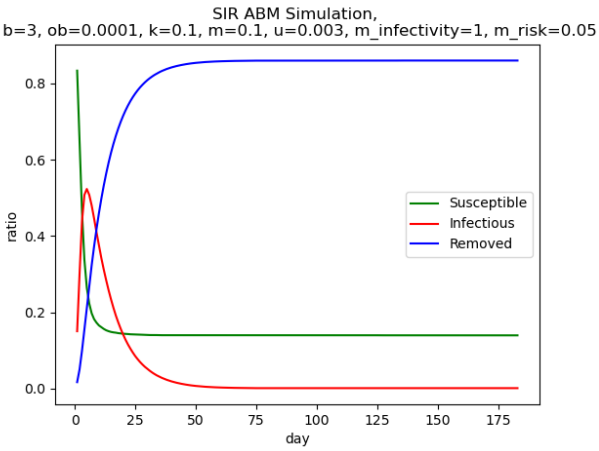

Figure 15: ABM results with vaccine on population of 10,000 with 500 initial infected

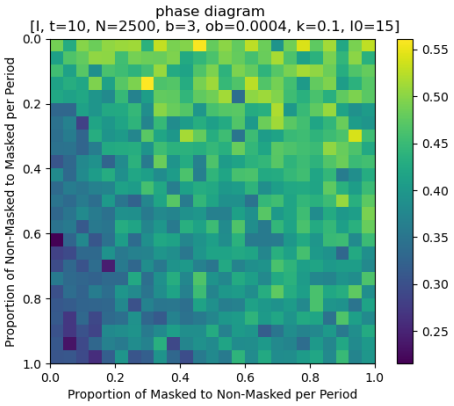

Figure 16: Phase plot of mask updating proportions. Color is scaled according to infection level at t=10.

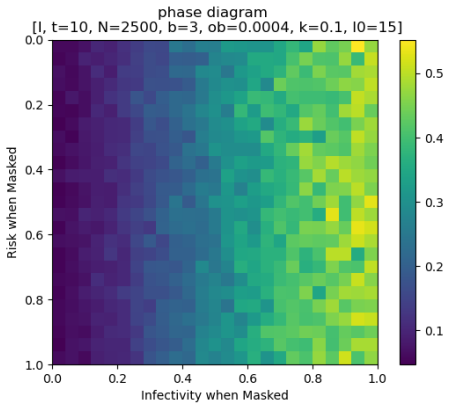

Figure 17: Phase plot of mask efficacy . Color is scaled according to infection level at t=10.

The phase plots in Figures 16 and 17 show cross-sections of the results of simulations with different combinations of mask parameters. Figure 16 shows that greater mask usage and mask persistence (fewer masked people giving up masks) leads to fewer cases. Figure 17 shows that greater mask efficacy, especially when worn by the infective, can have a massive, well-defined impact on number of cases.

3.4 Additional Compartments

3.4.1 Background

Depending on the disease being modeled or behavioral response to the disease, a variety of additional compartments may be added to the basic SIR model. Table 1 lists some of the possible extensions. In general, almost any combination of compartments can be used according to their suitability to the disease being modeled. For example, models that end with 'S' can be used to model diseases where re-infection is possible (i.e. immunity is only temporary). The full list of combinations available in the present version of the package is: SIR, SIS, SIRD, MSIR, SIRC, SEIR, SEIS, MSEIR, MSEIRS.

Letter	Compartment Name	Explanation
S	Susceptible	See Introduction
I	Infected (and infectious)	See Introduction
R	Removed	See Introduction
C	Carrier	Some "recovered" individuals have the chance to become infectious again without being infected by another person.
E	Exposed	Individuals who interact with an Infected first go through a phase of non-infectivity
M	Maternally-derived Immunity	Some diseases may feature maternally-derived or natural immunity by a portion of the population
Q	Quarantine	After becoming Infected, a portion of individuals are moved into quarantine and prevented from interacting with others until they recover
D	Deceased/ Death	Explicitly shows death instead of combining them in the Removed category

Table 1: Legend of compartments and their interpretations

3.4.2 Implementation

Each of the above-listed combination of compartments is implemented as an ODE class which can be used with ODE solver tools. One exception is the MSEIQRDS model which can be created through the MSEIRS class. Functions for generating solutions in array form, time plots and phase plots are implemented.

3.4.3 Results

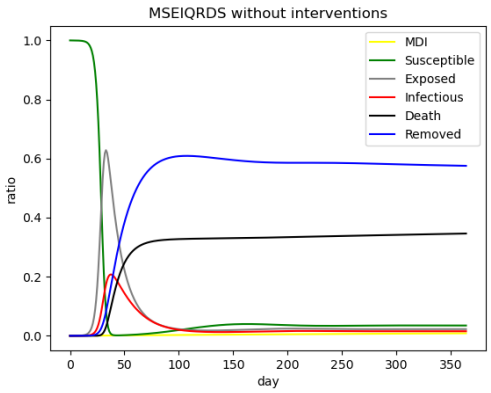

Figure 18: The trajectory of an MSEIQRDS model with no interventions in place (q = 0 and b = 3)

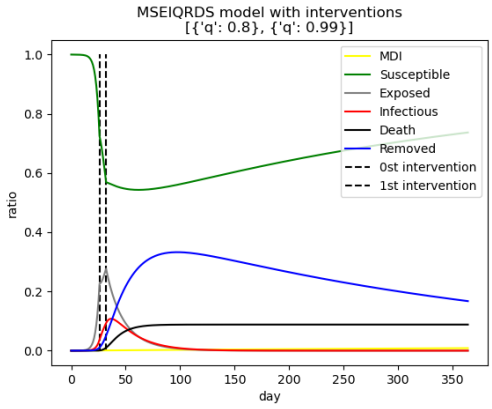

Figure 19: The trajectory of an MSEIQRDS model with staggered, strict quarantine policies in place

Figures 18 and 19 plot the trajectory of the disease with and without quarantine interventions respectively. Figure 22 in the Appendices shows the trajectory of the disease with social distancing instead of quarantine interventions. In general, we can see that the interventions do a great deal to decrease the proportion of the population that gets the disease at any point. Especially notable is the large decrease in deaths that result from the intervention. Additionally, we can see that with interventions, even without permanent immunity, the disease ultimately disappears and does not re-emerge.

4 Conclusion

This paper documents the Python implementation and usage of various epidemiological compartmental models. From the simulations produced by our implementations, we can see that adding various components to the model generally does not change the core functioning of the model but can yield differing results. The results from the spatial modeling extension demonstrate how the initial conditions of the disease in terms of not only number of infectives but also position and mobility of those infectives can affect the trajectory of the disease. In practical terms, the spatial model helps illuminate why the disease tends to hit a peak earlier in more densely populated, mobile, central locations such as metropolitan cities and travel hubs. Beyond physical location, "social location" can also matter a great deal to an individual's risk of being infected. If people reduce or even cut-off contact with people outside of their immediate social circle, they oftentimes still face a very high risk of catching the disease at some point. This is because members of one person's cohort may also belong to other people's cohorts, effectively linking everyone in the population. We can see that relatively smaller cohorts (i.e. those at the edge of the graph) tend to have a greater number of degrees of social separation between them and other individuals in the population. The results of the cohort simulations imply that social distancing must happen not only between strangers but also between close friends and even family members to be truly effective in preventing disease spread. Apart from social distancing, mask usage, vaccination, and quarantining are also promising avenues of disease spread alleviation. An especially interesting result is that the behavior of the Infected population tends to be the driving factor in the trajectory of the disease. Masks are most effective when they are used efficaciously by the infected population. Quarantining of infected individuals leads to massive reductions in the number of Infected and Deceased.

Our results improve upon the basic SIR model by allowing for additional considerations and more realistic scenarios. That said, there are still many limitations even to the models we have implemented. In the spatial model, it may be unrealistic to assume that everyone is uniformly distributed on a grid. It may be that certain areas of the grid have much higher population density than others. In the cohort model, it is more realistic to think that not all cohorts are linked in some way. Additionally, if testing is widely available and people are generally rational, the number of interactions even within tight social circles may decrease during times of high infection rates. In terms of masks and vaccination, it may not be realistic to think that fixed proportions of masked (vaccinated) and unmasked (unvaccinated) populations switch statuses each period. Lastly, as diseases begin to dissipate, quarantine and social distancing policies may become weaker, causing their effects to be less strong than presented in this paper.

Most of the limitations are based on the idea that interventions and behavior changes are unlikely to be fixed throughout the course of a disease. People will tend to adjust their behavior in accordance with the perceived risk in the moment. As such, it would be extremely interesting to incorporate infection rates into the parameters for social-mixing, mask usage, vaccination, quarantining, and social-distancing within each period. This would likely attain results that more closely mirror the reality of Covid-19 where policies, behavior, and infections interact with each other to produce the numerous "waves" of cases that have been observed [5]. Additionally, weighting connections between individuals to reflect physical and social distance may also help explain specific patterns of cases in terms of sub-populations.

References

[1] Maia Martcheva. *An Introduction to Mathematical Epidemiology*, volume 61 of *Texts in Applied Mathematics*. Springer US, Boston, MA, 2015. ISBN 978-1-4899-7611-6 978-1-4899-7612-3. doi: 10.1007/978-1-4899-7612-3. URL http://link.springer.com/10.1007/978-1-4899-7612-3.

[2] CDC. Coronavirus Disease 2019 (COVID-19), February 2020. URL https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html.

[3] Patsarin Rodpothong. Viral evolution and transmission effectiveness. *World Journal of Virology*, 1(5): 131, 2012. ISSN 2220-3249. doi: 10.5501/wjv.v1.i5.131. URL http://www.wjgnet.com/2220-3249/full/v1/i5/131.htm.

[4] Hilary Brueck. How to interpret the 'efficacy' rates of coronavirus vaccines. URL https://www.businessinsider.com/what-is-vaccine-efficacy-how-well-shots-prevent-infections-2020-11.

[5] Coronavirus Second Wave? Why Cases Increase. URL https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus.

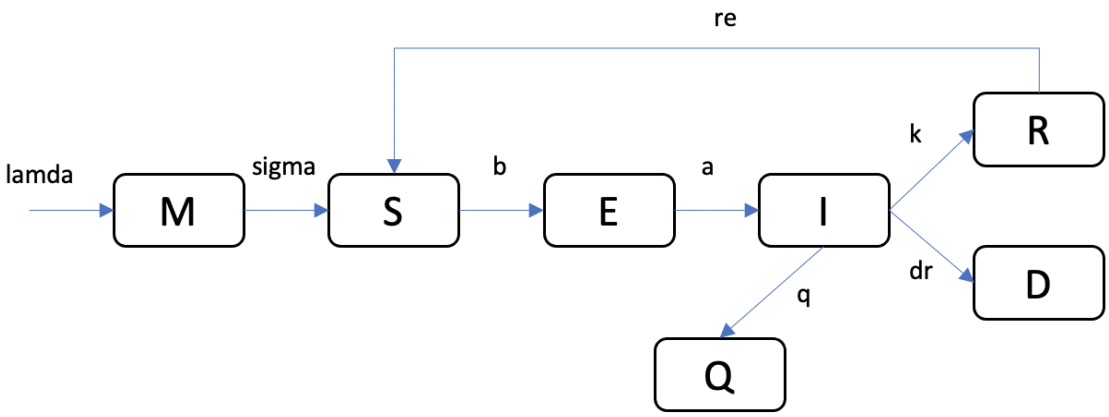

Figure 21: Diagram showing the flows and parameters of the MSEIQRDS model

5 Appendices

5.1 Mask Appendix

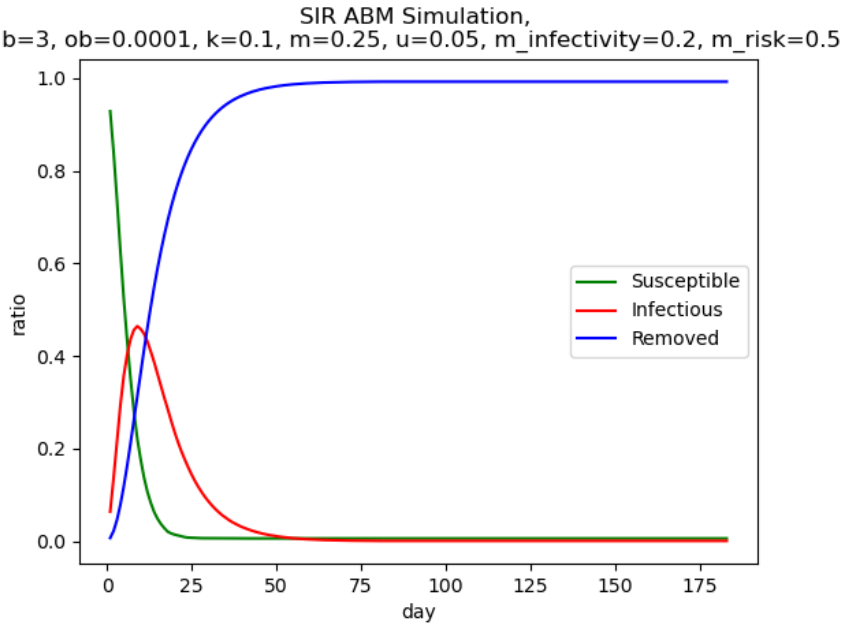

Figure 20: ABM results with no intervention and a large infected population (200/10,000). A similar peak infection rate is achieved here with only 200 initial infectives compared to the 500 used in the vaccination simulation, demonstrating the power of the vaccine in getting the disease under control quickly.

5.2 Additional ODE Model Plots

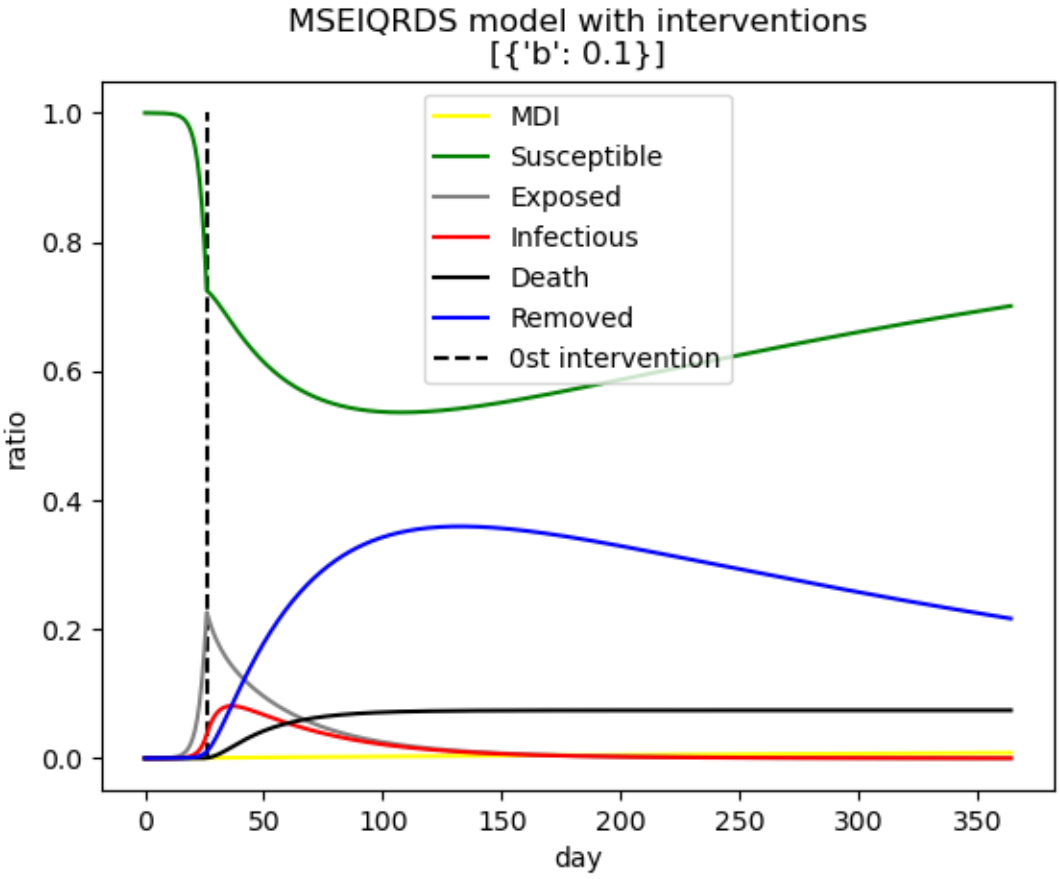

Figure 22: Results of model with social-distancing intervention

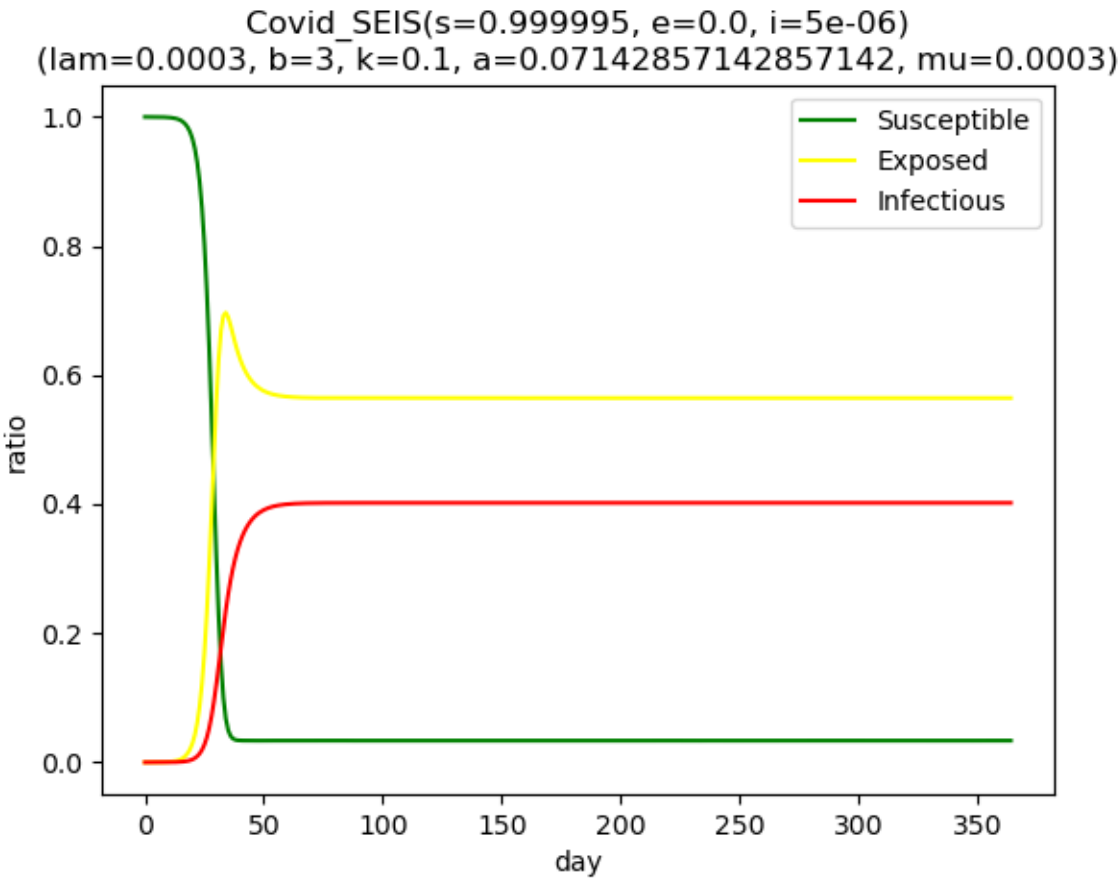

Figure 23: SEIS model results

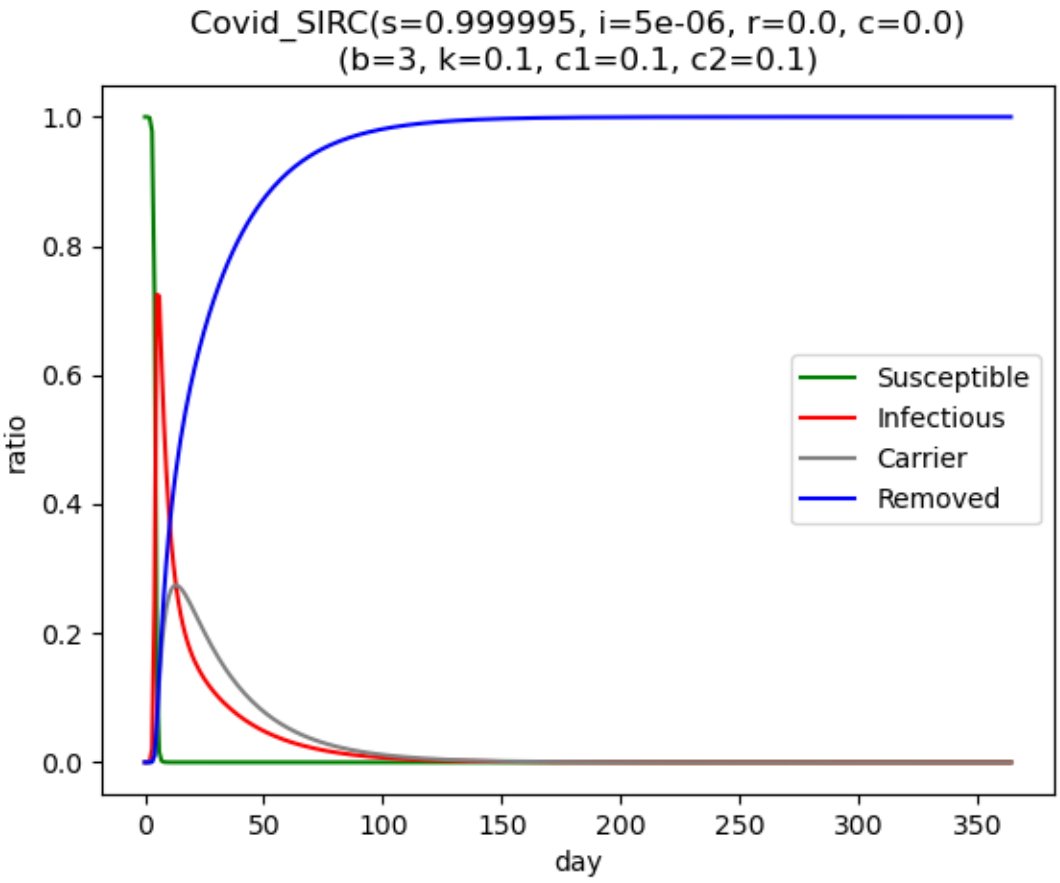

Figure 24: SIRC model results

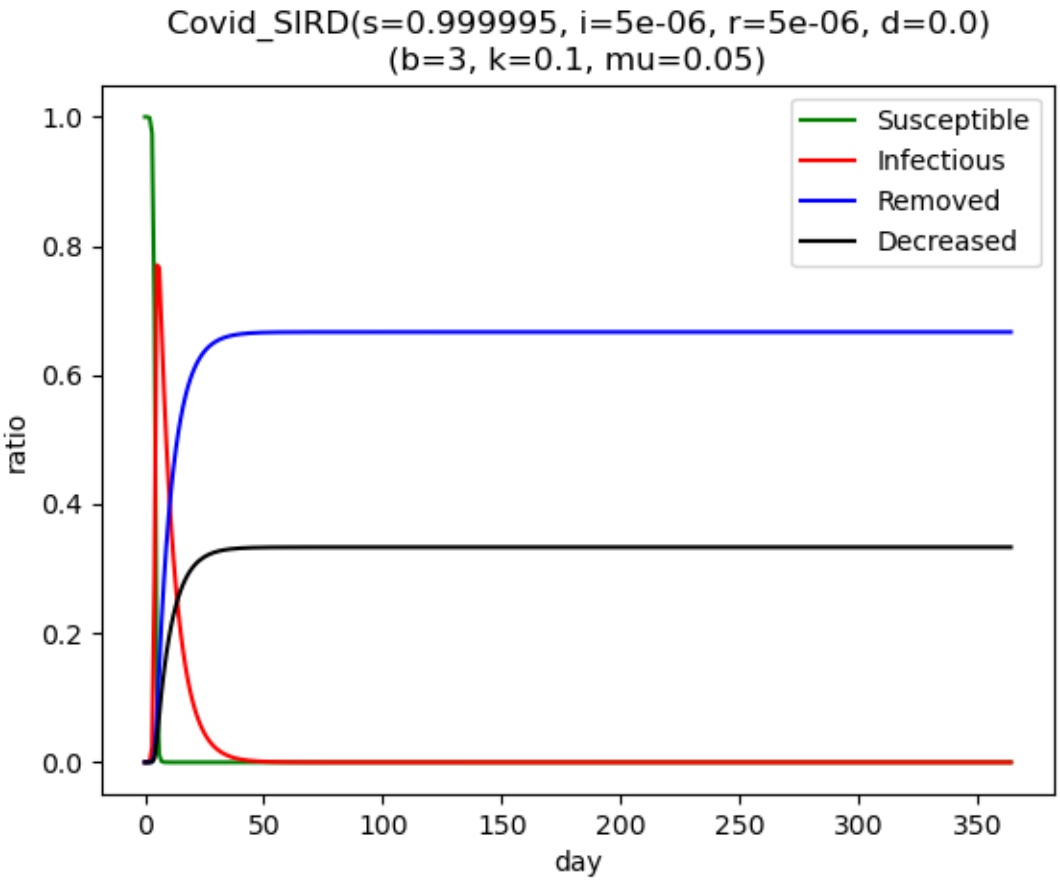

Figure 25: SIRD model results