Modeling Covid Cases and Deaths with Linear Dynamic Systems

Sherlock Zhang
Introduction to Electrical Engineering
Washington University in St. Louis
St. Louis, United States
jiaxi.zhang@wustl.edu

Brant Yang
Introduction to Electrical Engineering
Washington University in St. Louis
St. Louis, United States
boran.y@wustl.edu

***Abstract*— In our study, we developed a model that simulates the Covid cases and deaths in St. Louis from March 18, 2020, to March 22, 2023. Using linear dynamic system modeling with an improved SIRD model that is split to four different phases, we were able to create a good fit to the cases and deaths data provided. Our final matrix A contains 6 state variables to record populations in different stages, which reacts differently to being infected.**

I. Introduction

In this study, we are aiming to use linear dynamic system to model covid case trajectories for St. Louis. We need to use the base SIRD model with certain number of phases and some alterations to the base SIRD model to reflect different waves of Covid, such as Delta and Omicron, as well as the impact of people getting the vaccines.

In this project, we applied knowledge of linear algebra, such as linear dynamic systems, specifically state vectors, linear combination, matrix-vector multiplication, and vector addition. When processing data, we also calculated the cumulative sum of new cases each time interval as well as mean square error to better see how well our model fits to the data. We also introduced external inputs into the model when we model the interaction between two distinct populations and how government policies affect the rates travel and how that affects case trajectories.

II. Methods

In this project, our goal is to build linear dynamical system to model the COVID-19 cases and deaths in certain region. The model is an extension form of *SIRD* model, which has four state variables representing four different kinds of population, and an matrix to update the state. We extend the model so that it takes re-infection, travelling between two population, and vaccine into account. The objective is to build the model that fits our given data the most.

A. Base Model

In the base *SIRD* model, we have 4 state variables: susceptible, infected, recovered, and deceased population. For each day in the model, part of the susceptible population will be infected, and part of the infected population will be recovered and get immunity, deceased, or didn't get immunity and goes back to susceptible. The recovered and deceased people stay recovered or deceased. The ratios can be stored in the update matrix A (where i is infection rate, e is recovering rate, m is immune rate, and d is the decease rate).

$$\mathrm{A} = \begin{bmatrix} 1-i & e & 0 & 0 \\ i & 1-e-m-d & 0 & 0 \\ 0 & m & 1 & 0 \\ 0 & d & 0 & 1 \end{bmatrix} \quad (1)$$

The update process represents a continuing linear transformation.

$$x_{t+1} = Ax_t \quad (2)$$

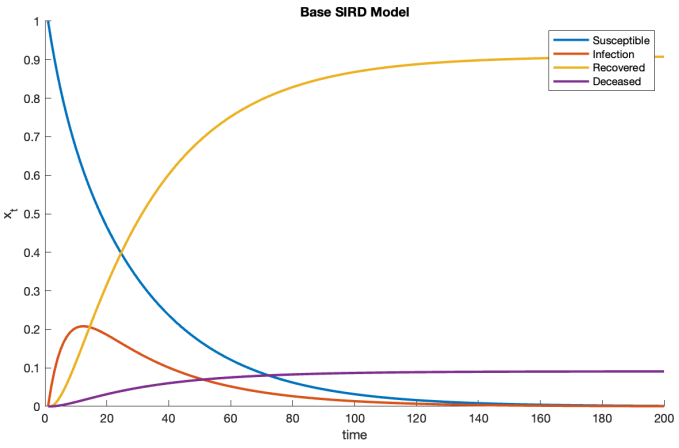

Fig. 1. The Base SIRD model without re-infection considered

However, the base model has a few flaws with the assumptions, and we modified it to better fit the specialties of COVID-19. According to the graph, the line for recovered people is a monotonically increasing curve slowly converging to 1-d, meaning the model assumes that people who recovered from the disease would get permanent immunity. This is clearly not the case for COVID-19 as people can only be temporarily immune to the virus for a period of time and exposed to re-infection again. Therefore, we modified the third column of matrix A and add a new r_e to represent the waning immunity, which allows certain proportion of people to go back to susceptible category after being recovered and immune

to the virus. If the r_e value is high, it implies rapid loss of immunity and higher likelihood of reinfection.

$$A = \begin{bmatrix} 1-i & e & r_e & 0 \\ i & 1-e-m-d & 0 & 0 \\ 0 & m & 1-r_e & 0 \\ 0 & d & 0 & 1 \end{bmatrix} \quad (3)$$

B. Fitting model into real-life data

As the COVID-19 virus has evolved to different variants in three years, which has different infection rate, immune rate etc. A single time invariant system fails to account for such changes. Our group decided to firstly separate the whole timeline into several segments and use different matrix A to account for these differences.

In order to decide where to update matrix A, we firstly looked up the periods of two dominant variants "Delta" and "Omicron" exists, and then put the time left into 2 more time period. Using this method, we eventually separate the timeline into 4 segments shown in table 1.

TABLE I. SEPARATION OF DIFFERENT TIME PERIODS

Pre-Delta period	2020-03-18 to 2021-06-23
Delta period	2021-06-30 to 2021-10-26
Omicron period	2021-10-27 to 2022-03-22
Post-Omicron period	2021-10-27 to 2022-03-22

In this way, 1 initial state and 4 corresponding A matrix is needed to finish the model. The initial state x_0 is defined as the following, where I_1 is the population proportion being infected in the first day:

$$x_0 = [1 - I_1 \quad I_1 \quad 0 \quad 0] \quad (4)$$

Equation (2) is used to calculate the simulated SIRD model. Matrix A is updated with new entries when entering a new period. We get our final modeling results by combining the data from the 4 periods, and get four 1 ×158 vector *S, I, R, D*, representing the population proportion in each category in each week.

$$X = [S \quad I \quad R \quad D]$$

Then, we tune our model to make it fit the provided data in St Louis. We compute the cumulative cases and deaths from our model. The death can be directly extracted from *D*, and the cumulative cases *C* can be calculated by performing cumulative sum to the product of susceptible population fraction and infection rate.

$$C_a = \sum_{a=2}^{n} S_{a-1} \cdot i \quad (5)$$

We plot the modeling cases, deaths and the given cases and death on the same plot and modify the entries in matrix A by hand to make the modeling graph look similar to the given graph. In addition to graphing, we also calculate the mean square error between modeling and give data to evaluate the modeling accuracy.

$$MSE_1 = \frac{\sum_{a=1}^{n}(C_a - C_a')^2}{n} \quad (6)$$

$$MSE_2 = \frac{\sum_{a=1}^{n}(D_a - D_a')^2}{n} \quad (7)$$

These two equations quantify the error between the modeling data and the given data, so when changing the entries in entries of matrix A, we are trying to make both MSE_1 and MSE_2 smaller.

After tuning the model, we tried to implement our own policy by changing the entries in A matrix and get a 25% drop in both cases and deaths in omicron period.

C. Considering travelling

To consider the effect of travel on the case trajectories of two cities, there are several aspects to consider. First of all, the number of people leaving a city are people who are susceptible, infected, or recovered. Deceased population should always be 0. Secondly, the proportion of S, I, and R of the people who leave the city should follow the same proportion of the population the day before, which means B1 and B2 need to be updated for every time interval. Lastly, for each time interval, the total population for the two cities both changes, so there need to be two other vectors recording the populations for these two cities respectively so that we can calculate the proportion of the cumulative cases to the overall population at each time interval. We assume that the infectious rate is 0.3% for St. Louis and 0.2% for the other hypothetical city. The SIRD parameters A1 and A2 are set with arbitrary numbers to represent two cities and the population of the second city is also an arbitrary number. The proportion of the population leaving both cities each time interval is a randomly generated percentage between 0.1% to 0.8%. This travel is modeled through u1 and u2, which represent the number of people leaving each city per day. As time progresses, we calculate the changes in each compartment (SIRD) for both the isolated case (without travel) and the case with inter-city travel. For the latter, we remove the population that is leaving the city and add the population entering the city. We adjust the numbers based on the percentage for each compartment in each city on a given day. The cumulative case counts proportional to the total population of each city are then calculated with and without considering travel to better visualize the impact of traveling in the spread of COVID. The time-invariant linear dynamical system with two external inputs is:

$$xt_{t+1} = A1 * xt_t - B1 * u_1 + B2 * u_2 \quad (8)$$

$$xt2_{t+1} = A2 * xt2_t + B1 * u_1 - B2 * u_2 \quad (9)$$

where xt and xt2 are cumulative population in each compartment of SIRD at each time interval. A1 and A2 are the A matrix for the two cities, u1 is the population leaving St. Louis, and u2 is the population leaving the other city. B1 and B2 are the matrices that set the proportion of the number of people who are susceptible, infected, recovered, or deceased in the traveling population. They are based on the proportion of the population in these compartments on each day. The new total population is also calculated by subtracting u1 and adding

u2 to the current population. The graph shows the proportion of the cumulative case to the population for each time interval.

We also performed the what-if policy analysis by simulating the same time period (t=300) with two phases on the two cities. In the first phase, the u1 and u2 remain the same to represent normal rates of travel between the two cities. In the second phase, however, u1 and u2 become smaller and B1 and B2 have a 0 in the infected compartment. Smaller u1 and u2 mean that people are traveling less because they know the new wave is more dangerous. The 0 in the infected compartment for B1 and B2 means that the government is isolating their cities and does not allow people who are currently infected to travel to another city. This could be interpreted as when the Delta or Omicron wave appeared, the government started to tighten policies regarding travel and did not allow COVID-infected people to leave the city.

D. Considering vaccination

In this part of the project, we are trying to extend our model to consider the effects of vaccine and fit the mock COVID data and further predict the population received vaccine and the vaccine breakthrough population in the given model. We first need to identify the roll-out date of the vaccine and make some correction to our assumptions to our model.

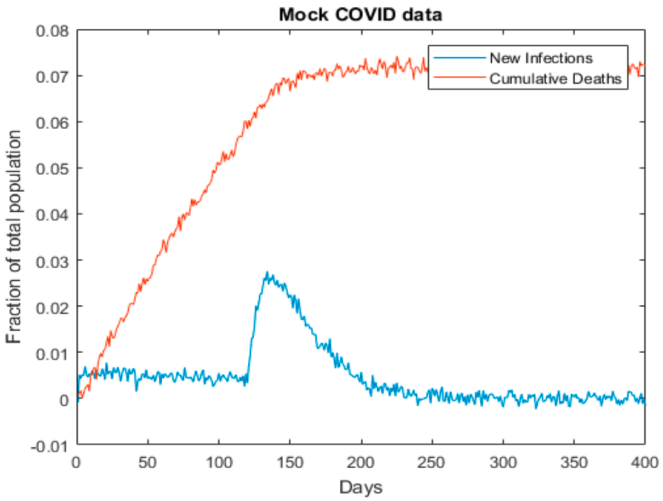

Fig. 2. Mock COVID data

In figure 2, we observe that there is a drastic increase in new cases starting around day 120. We assume this day is the vaccine roll out date because people who receive vaccine have virus injected to their body, and they will have higher rate of infection but lower death rate. In addition, in the figure after day 200 the new infections slowly converge to zero. Therefore, we assume that people who recovered from COVID have permanent immunity to the virus.

Based on these assumptions, we add two state variables to the base SIRD model: *V* representing the vaccine but not infected population, and I' representing the vaccine and infected population. The original *S* variable now represent people who are susceptible and didn't receive vaccine, and *I* represent population who are infected but didn't receive vaccine. The updated state variable x' can be represented by:

$$X = [S \quad I \quad I' \quad R \quad D \quad V] \tag{10}$$

The updated matrix A' can update the new model, where variable i, d, e, m represents the infection, death, recover, and immune rate for people who haven't receive vaccine, i_2, d_2, e_2, m_2 represents the corresponding rate for vaccinated people.

$$A = \begin{bmatrix} 1-\mathrm{i}-\mathrm{v} & \mathrm{e} & 0 & 0 & 0 & 0 \\ \mathrm{i} & 1-e-\mathrm{m}-\mathrm{d} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1-e_2-d_2 & 0 & 0 & i_2 \\ 0 & \mathrm{m} & 0 & 1 & 0 & 0 \\ 0 & \mathrm{d} & d_2 & 0 & 1 & 0 \\ \mathrm{v} & 0 & e_2 & 0 & 0 & 1-i_2 \end{bmatrix} \tag{11}$$

The same equation (2) will be used to update the state of the system. Before the roll out date, *v* is set to 0 so the model works the same as base SIRD model, and after the roll out date, v is set to the proportion who receive vaccine.

Then we change the variables in matrix A to calibrate the data. We are using the same plotting and calculating the MSE method to calibrate our model, but the modeling new cases C is algorithm is updated.

$$C_n = S_{n-1} \cdot i_{n-1} + Vn - 1 \cdot i_{2\,n-1} \tag{12}$$

Finally, after calibrating our model, we can calculate vaccine population (vaxpop) and vaccine breakthrough population (vaxbreak):

$$vaxpop_n = S_{n-1} \cdot v \tag{13}$$

$$vaxbreak = I' \tag{14}$$

III. Results and Discussion

A. Results

Figure 3 shows the SIRD model after adding the reinfection rate r_e = 0.5. Figure 4 and 5shows the comparison of model based and given cumulative cases and cumulative death after tuning the entries in matrix A, and table 2 shows the entries in matrix A that the mode use in different period of time. Figure 6 shows the comparison of model based new cases and cumulative deaths with the mock data, and table 3 shows the entries the model uses.

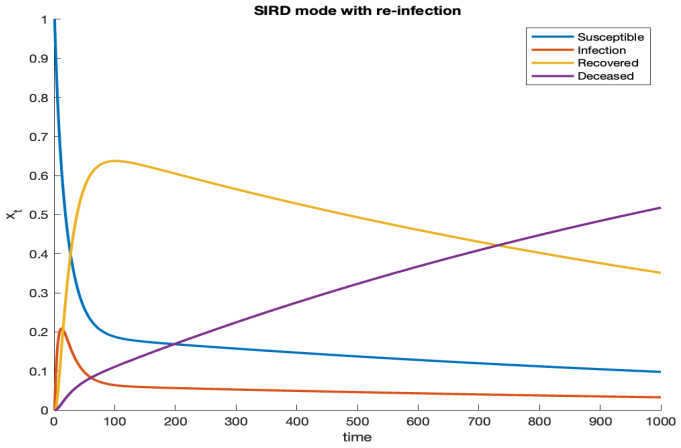

Fig. 3. SIRD model with re-infection

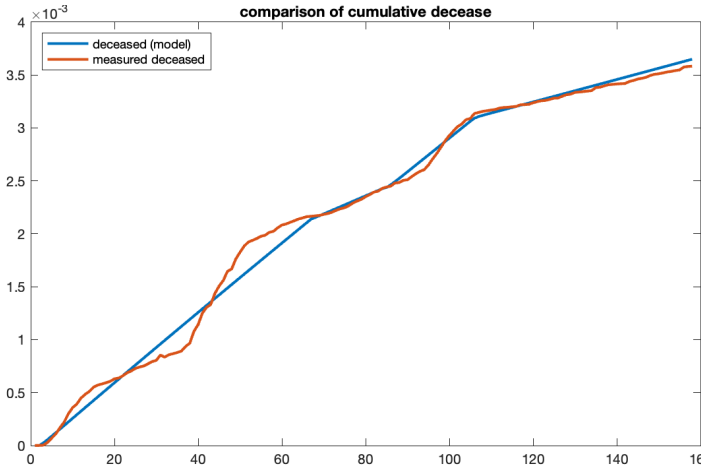

Fig. 4. Model cumulative death and given cumulative death in St Louis

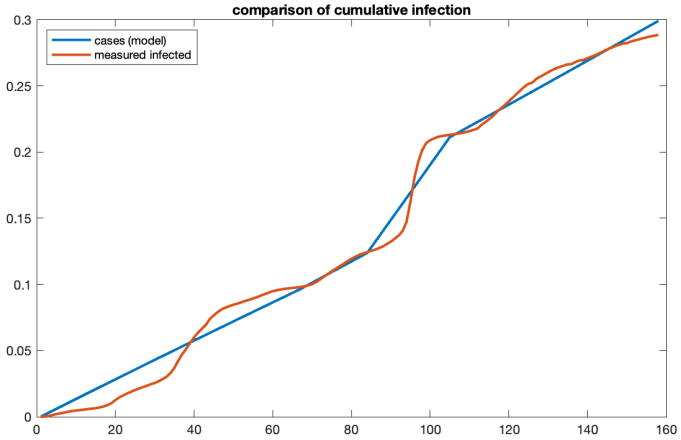

Fig. 5. Model cumulative infection and given cumulative infection in St Louis

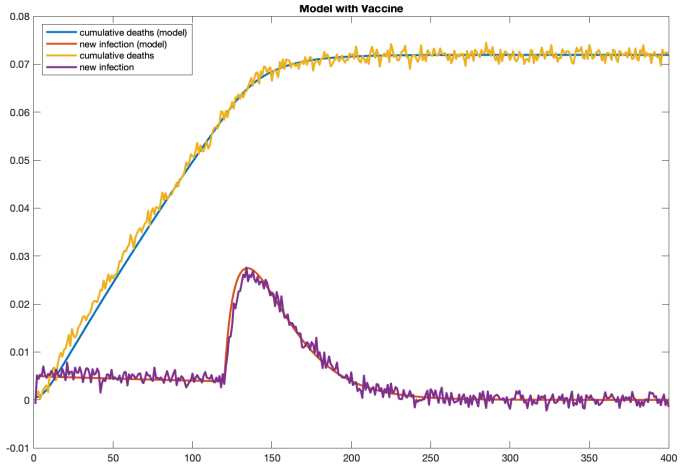

Fig. 6. Model with vaccine compared to mock data

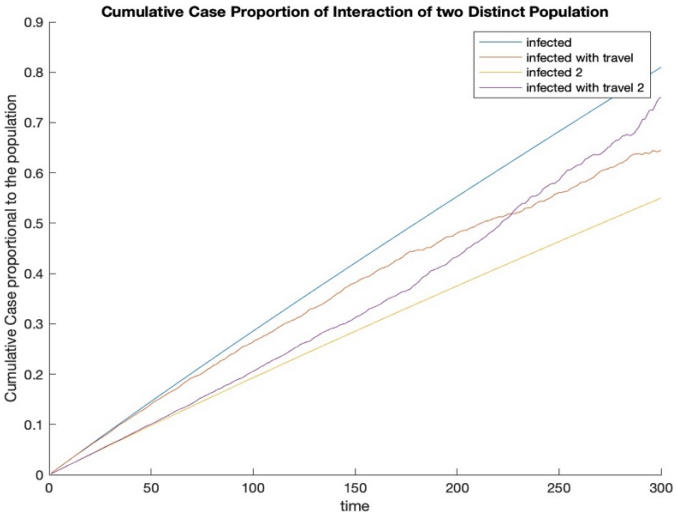

Fig. 7. Cumulative Cases Proportional to Population For Each City(with & without traveling)

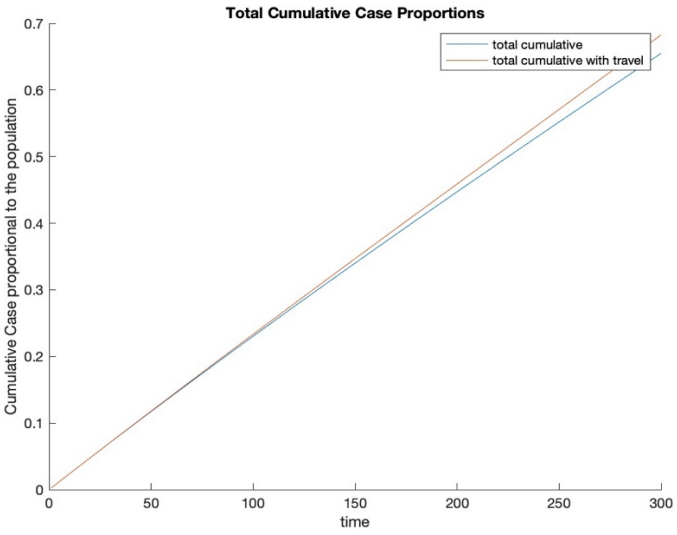

Fig. 8. Combined Cumulative Cases Proportional to The Total Population of The Two Cities (with & without travel)

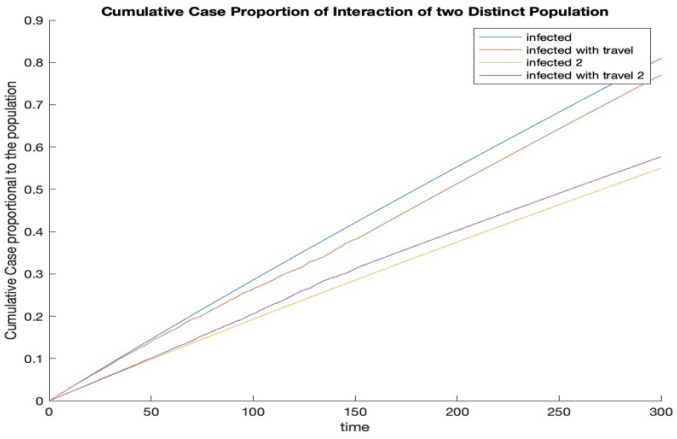

Fig. 9. Cumulative Cases Proportional to Population For Each City with two phases (with & without traveling)

TABLE II. RATE CONSTANT DERIVED FROM ST LOUIS MODELING

	i	*e*	*m*	*d*	r_e
Pre-Delta	0.0015	0.3	0.39	0.016	0.005
Delta	0.0017	0.35	0.35	0.0075	0.005
Omicron	0.0044	0.4	0.25	0.005	0.05
Post Omicron	0.0017	0.5	0.2	0.0045	0.05

TABLE III. RATE CONSTANT DERIVED FROM MODEL WITH VACCINE

	i	*e*	*m*	*d*
No vaccine	0.005	0.1	0.05	0.02
vaccine	0.15	0.07	0.1	0.00005

B. Analysis of adding re-infection

After adding the re-infection rate to the base model, the deceased population does not converge to a certain number as time increases. Instead, the growth rate of cumulative death is approximately linear as time increases. This is a more accurate prediction for COVID-19 deaths because the given St. Louis data cumulative death graph looks more like a combination of line segments with different slopes instead of exponential functions.

C. Interpretation of rate constant and policies

The rate constant derived from 4 different matrix A can help us analyze the different between virus variants. For example, the infection rate of Pre-Delta and Delta period is similar, but the omicron period it increases by over 100%. Moreover, the death rate of pre-delta period is high, and it gradually decreases in the latter periods. We did some research and verified that this observation is consistent with the real-life situation: the virus in the earlier period (such as alpha or delta) has lower infection rate and higher death rate, while those in latter period (omicron and later) have higher infection rate but lower death rate.

The rate constant can also be interpreted in another way related to policies in St. Louis. For example, according to our research on local COVID policies in the past, during the pre-delta and delta period, government implemented stay-at-home orders in late March 2020. Non-essential businesses were closed, and social distancing measures were put in place. Such policies can drastically decrease the rate of infection. However, during the omicron period, the government canceled the lockdown, and there was a trend towards returning to normal activities, with businesses and schools operating more regularly. This corresponds the increase infection rate in omicron period.

Considering the effects of policies making on the COVID cases and deaths, we tried to come up a policy that can lead to a 25% drop in both cases and deaths. Our policy is to continue the lockdown and shutting the unnecessary business during the omicron period, which in turn would lead to a lower infection rate. We implement different infection rate and find the minimum infection rate decrease is 0.0012, in order to achieve the 25% drop. However, this policy might not be feasible as the social cost of implementing it is huge. Business and people have been restricted for 80 weeks before the omicron period, and it would bring even huge loss for both economy and people if we don't return to normalcy and cancel the lockdown, as people cannot work off-line and students cannot go to school. This cost is too great for the policy to be worthwhile, considering the death rate of the omicron period is already quite low. Although it is quite costly to implement the policy, the model proves that lockdown is an good way of reducing infected cases and deaths.

D. Analysis of travelling

Based on figure 7, we see that when one population experiences a higher infected proportion, the other population experiences a lower infected population. If the travel rates increase, there will be a higher difference between the population with travel and the isolated population. We also calculated the cumulative cases for the combined populations of the two cities in figure 8 and found out that when travel is introduced, the cumulative cases for the combined populations of the two cities would always increase due to cross-infection.

From figure 9, we can see that starting at the second phase (t=150), the line for cumulative cases proportion for both cities have almost the same slope as the case trajectories without travel, which effectively slows down the infection rate.

E. Interpretation of model with vaccine

In this model with vaccine, we compare the rate constant for population who have received vaccine and who haven't to show the impact of vaccine. The death rate of vaccinated people drastically decreased compared to the original death rate, and more people are getting immunity rather than recovered but not immune. These observations are consistent to our research, that vaccine brings a lower death rate and higher immunity rate to protect people.

However, the infection rate is much higher after people are vaccinated. This is out of our expectation, and inconsistent to our research results. We provide an alternative interpretation that vaccine is not the only determinant of the infection rate, and the policy might increase the infection rate as well. Maybe during the period of people receiving vaccine, the government ended lockdown because of vaccination and people are more likely get infected.

F. Limitations

The linear dynamic system we use in modeling is mostly time invariant, which means the update matrix A is same all the time (or changes with relatively low frequency compared to the long-time span). This is because we are using the hand tuning method to calibrate our modeling results to fit in the given real-life data or given simulated data. Therefore, we can only map out the approximate modeling for COVID cases and deaths, which cannot be highly reliable.

One limitation to the model of traveling is that only two populations are interacting, whereas, in real life, multiple cities interact simultaneously, which could be incredibly more sophisticated. In addition, the number of people traveling is randomly generated, which is likely not accurate in many cases. We would need more data to know approximately how many people are traveling between two cities to more accurately model their interactions.

Moreover, there are a lot of assumptions we made in order to make our model work, and those assumptions may not necessarily be consistent with real life situation. For example, when considering the model with vaccine, we assumed permanent immunity after recovery, which doesn't necessarily to be true in real-life situation.

IV. Conclusion

A. Summary

In conclusion, by separating the timeline into four segments with four distinct A matrices and adding I' and V to our matrix A, we can model the actual trajectories of the cases and deaths relatively well. We considered the effect of different stages with outbreaks, interactions between two distinct populations with government policies at different phases, and vaccines affecting infections and deaths. We also calculated both the vaccine population and breakthrough infection population as a function of time.

B. Future Plan

In order to improve our modeling, we should try to use the time variant system to give more accurate model for COVID cases. This can help government analyze and predict the epidemic situations more accurately.

Moreover, future models should incorporate better calibration methods that can tune the model better than the current hand-tune result. We should try the fmincon function and see how to provide better optimization.

References

[1] S. Boyd and L. Vandenberghe, *Introduction to applied linear algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press, 2018. doi: 10.1017/9781108583664.

[2] "COVID-19 Standards and guidance," *stlouis-mo.gov*. https://www.stlouis-mo.gov/government/departments/health/communicable-disease/covid-19/guidance/index.cfm#:~:text=Currently%2C%20there%20is%20no%20mask,www.cdc.gov%5D.

[3] Navarrete J, Barone G, Qureshi I, et al. SARS-CoV-2 Infection and Death Rates Among Maintenance Dialysis Patients During Delta and Early Omicron Waves — United States, June 30, 2021–September 27, 2022. MMWR Morb Mortal Wkly Rep 2023;72:871–876. DOI: http://dx.doi.org/10.15585/mmwr.mm7232a4