Modeling COVID Dynamics in Populations

Nix, Ines
Electrical and Systems Engineering
Washington University in St. Louis
St. Louis Missouri, United States
i.n.nix@wustl.edu

Fandozzi, Raynah
Electrical and Systems Engineering
Washington University in St. Louis
St. Louis Missouri, United States
f.raynah@wustl.edu

Jiang, Christine
Electrical and Systems Engineering
Washington University in St. Louis
St. Louis Missouri, United States
christine.jiang@wustl.edu

Since March 2020, the rapid spread of COVID-19 left the United States rushing to contain the virus and understand its transmission. Several factors influencing the infection rate including mask mandates, vaccines, quarantines, and emerging variants necessitate an in-depth analysis to document the transmission rates. In our study, we developed multiple models, such as SIRD modeling, to track the virus's spread in relation to key events and policy actions. SIRD modeling divides the population into four categories - susceptible infected, recovered, and deceased - allowing us to analyze and optimize simulated infection and death rates with real data. However, some trends were challenging to capture due to the initial model's simplicity. Accordingly, we introduced two additional state variables focused on vaccination and breakthrough infection rates, providing a more comprehensive framework for the pandemic dynamics. By modeling how infection and death rates respond to specific policies over time, our project offers insights into the critical effects of crisis preparation and response. These findings are valuable for informing future epidemic monitoring and response strategies.

I. Introduction

By the time the CDC declared the COVID-19 pandemic over on May 11th, 2023, the United States had faced over 193 million cases [1], adding to the $4.6 trillion in response and relief costs [2]. Thus, understanding how the virus variants and government policies adversely impacted susceptibility, infection, and mortality rates is crucial. Applying skills we learned in class and lab, such as matrix multiplication, to SIRD modeling in linear dynamical systems enhances our knowledge of electrical and systems engineering. Using tools such as MALTAB's *fmincon* and state-space model function, *ss*, to assess COVID-19 case data, we gain insights into the impacts of policy changes and virus variants on populations. By optimizing our SIRD modeling to reflect changes in susceptible, infected, recovered, deceased, vaccinated, and breakthrough-infected groups, we aim to minimize error and understand which policies contribute to resilience or vulnerability. These findings can guide future epidemic mitigation strategies and support better communication on preparedness and response.

II. Methods

A. Data Sources

The census data we analyzed for this project comes from the Annual Resident Population Estimates (CO-EST2021-ALLDATA) United States Census Bureau. This includes census information, such as population estimates for each county. The COVID-19 case data is sourced from The New York Times, based on reports from state and local health agencies. The COVID case and death data are contained in arrays called "`cases_STL`" and "`deaths_STL`" respectively, in MATLAB.

B. Overview of Methods

Our analysis of the COVID-19 data in St. Louis is based on fitting a SIRD model using a linear dynamic system, with the *fmincon* function employed to create a 4x4 A matrix. For the COVID-19 mock data, we expanded the model to a SIRDVV_b framework, also using a linear dynamic system and hand-tuning to create a 6x6 A matrix. The values within these matrices allowed us to draw key conclusions about the trends in the COVID data modeled. To visualize these records and their significance, we used MATLAB version 2024a.

C. Data Preprocessing

In Parts 1 and 2, we divided the COVID case and death counts by the population of St. Louis to express them as proportions of the total population, making the model easier to interpret. It's important to note that this data is cumulative, so the matrices we created represent cumulative counts over the time periods analyzed. In Part 3, we worked with cumulative death counts and new infections, which differed from the data used in the earlier parts. To address this, we calculated current daily deaths by taking the difference between cumulative counts on consecutive days. Processing was more complex for new infections, so instead of transforming the data to current daily counts, we adapted our model comparison approach, as discussed in Section II E.

D. Using SIRD

To model with SIRD, we used a linear dynamic system based on matrix multiplication of a matrix A to get from day t to day t+1.

$$x_{t+1} = Ax_t \tag{1}$$

$S \to S$	$I \to S$	$R \to S$	$D \to S$
$S \to I$	$I \to I$	$R \to I$	$D \to I$
$S \to R$	$I \to R$	$R \to R$	$D \to R$
$S \to D$	$I \to D$	$R \to D$	$D \to D$

TABLE I
A MATRIX FOR SIRD

Each element of the A matrix represents the fraction of each population category—susceptible, infected, recovered, or dead—that contributes to another category (see Table I). Thus, each column represents a distribution across the four SIRD populations and should sum to one to maintain proportionality.

To create the optimal 4x4 A matrix for the first and second parts of the case study, we utilized the *fmincon* function. We defined two main constraints for the optimization: each element in A must lie between 0 and 1 (representing fractions of the population), and each column of A must sum to 1. The latter constraint was harder to do but was achieved using an A_{eq} constraint and a b_{eq} constraint which are shown in the code.

Our objective function minimized the sum of squared errors between the model's weekly estimates and the actual data. Specifically, we minimized the sum of squared differences between the model's estimated weekly infections and real weekly infections, as well as the estimated weekly deaths and real weekly deaths.

To improve model realism, we set certain elements of A to zero. For instance, individuals in the "dead" population cannot transition back to the susceptible, infected, or recovered populations, so the fourth column of A must be fixed at (0, 0, 0, 1). We ensured that these values were zero in the initial A matrix passed to *fmincon*, so the optimization only affected the relevant non-zero elements.

With these constraints and the objective function, the algorithm generated an accurate A matrix to model the initial COVID strain [Table II], the Delta [Table III], and the Omicron [Table IV] waves of the pandemic.

0.9984	0	0.0012	0
0.0016	0.9645	0.8901	0
0	0.0353	0.1087	0
0	0.0002	0	1

TABLE II
TRANSITION MATRIX FOR INITIAL COVID-19 MODEL

0.9796	0	0.0473	0
0.0204	0.7253	0.0937	0
0	0.2745	0.8590	0
0	0.0002	0	1

TABLE III
A MATRIX FOR DELTA WAVE

0.9716	0	0.0001	0
0.0284	0.4687	0.2635	0
0	0.5312	0.7364	0
0	0.0001	0	1

TABLE IV
A MATRIX FOR OMICRON WAVE

E. Using SIRDVV$_b$

A significant limitation of the basic SIRD model is its oversimplification, as it doesn't account for vaccination effects or their impact on population dynamics. To address this, we extended the SIRD model to create the $SIRDVV_b$ model, which includes variables for susceptible, infected, recovered, deceased, vaccinated, and breakthrough infections among the vaccinated population. Similar to previous models, we applied a linear dynamic system,using matrix multiplication with the expanded $SIRDVV_b$ matrix A [Table V] to analyze the data.

In Part 3, we were tasked with analyzing vaccinations within a population, including breakthrough infections among vaccinated individuals. For this mock data, we assumed that vaccines were introduced at week 125, based on a noticeable leveling-off in the cumulative death curve around that time, which suggests a decline in death rates due to vaccination. Before week 125, we applied the original SIRD model to represent the data [Table VI]. For MATLAB syntax, we used a 6x6 matrix format but set all values related to V and $SIRDVV_b$ to 0.

After week 125, we introduced a new 6x6 matrix that now included the vaccinated population [Table VII]. In this model, we assumed that only susceptible individuals could receive vaccination. Initially, we used the *fmincon* function to optimize the A matrix, but encountered numerous errors with the new data, particularly in implementing the error function, as the data now included new infections rather than cumulative counts. To address this, we adopted a manual tuning approach for both matrices, evaluating our model's fit by visually comparing it to actual cases and deaths on a graph. A limitation here was that our model tracked current infections, while the real data reflected new weekly cases. The current infections in the model are calculated based on the following transitions:

$S \to S$	$I \to S$	$R \to S$	$D \to S$	$V \to S$	$V_b \to S$
$S \to I$	$I \to I$	$R \to I$	$D \to I$	$V \to I$	$V_b \to I$
$S \to R$	$I \to R$	$R \to R$	$D \to R$	$V \to R$	$V_b \to R$
$S \to D$	$I \to D$	$R \to D$	$D \to D$	$V \to D$	$V_b \to D$
$S \to V$	$I \to V$	$R \to V$	$D \to V$	$V \to V$	$V_b \to V$
$S \to V_b$	$I \to V_b$	$R \to V_b$	$D \to V_b$	$V \to V_b$	$V_b \to V_b$

TABLE V
A MATRIX FOR $SIRDVV_B$

- $I \rightarrow I$ (the amount of infected people who did not recover or die from the previous day),
- $V_b \rightarrow V_b$ (the amount of breakthrough-infected people who did not recover or die from the previous day),
- $S \rightarrow I$ (susceptible individuals becoming infected),
- $R \rightarrow I$ (recovered individuals becoming reinfected),
- $V \rightarrow V_b$ (vaccinated individuals becoming breakthrough infected).

To compare this with the new infections data, we focused on the sum of:

- $S \rightarrow I$ (susceptible individuals becoming infected),
- $R \rightarrow I$ (recovered individuals becoming reinfected),
- $V \rightarrow V_b$ (vaccinated individuals becoming breakthrough infected).

With these methods, we created matrices to model the time periods for the mock data in Part 3 [Table VI and Table VII].

III. Results and Discussion

A. Part 1

Part 1 of the case study was all about understanding linear dynamical systems and how to model them in MATLAB. Initially, we were given a matrix that did not have reinfections [Fig. 1], and we were told to change it so that it did. The element in the matrix we changed was the recovered to infected element, changing it from 0 to 0.2 [Fig. 2].

These two figures illustrate how the presence or absence of reinfections affects the long-term dynamics of the model. In the model without reinfections, we observe that the infection rate and death rate stabilize more quickly, with both ratios reaching a plateau. By contrast, in the model with reinfections, these trends remain less predictable, with the death rate continuing to rise over time.

To determine the long term behavior of these two models, we examined the trends on the right side of each graph, focusing on the trajectory of each population group over time. For both models, the susceptible and infected population both seemed to trends towards zero as time progresses. However, the recovered and dead populations exhibit different long-term behaviors depending on whether reinfections are included.

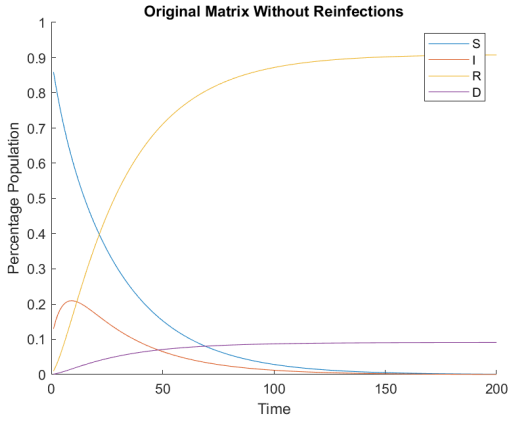

Fig. 1. Part 1 Without Reinfections

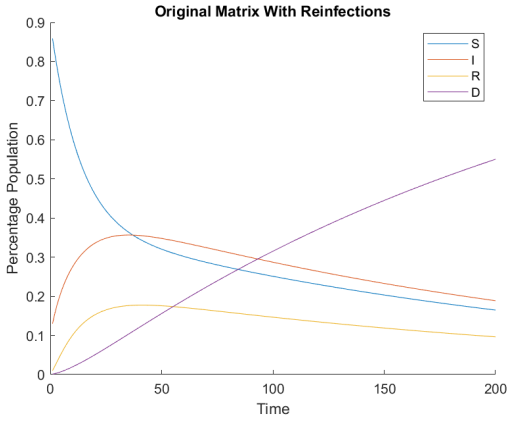

Fig. 2. Part 1 With Reinfections

0.998	0	0.39	0	0	0
0.002	0.6	0.06	0	0	0
0	0.3998	0.55	0	0	0
0	0.0002	0	1	0	0
0	0	0	0	0	0
0	0	0	0	0	0

TABLE VI
A MATRIX FOR WEEKS 0-125 OF MOCK DATA

In the model without reinfections, the death rate reaches a steady plateau at approximately 0.1, indicating that 10 percent of the population eventually dies, while 90 percent recovers. Conversely, in the model with reinfections, the death rate never plateaus, instead continuing to rise until it approaches 1, indicating that the entire population ultimately succumbs to infection and dies. In the end, in the absence of reinfections, the majority of the population recovers, whereas in the presence of reinfections, the model predicts an eventual outcome of complete mortality.

B. Part 2

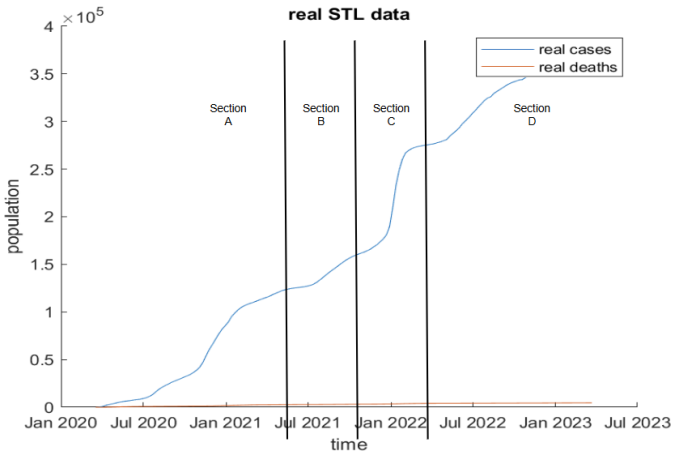

Fig. 3. Full STL COVID Data

0.9454	0	0.001	0	0	0
0.0041	0.99499	0.8423	0	0	0
0	0.005	0.1567	0	0	0
0	0.00001	0	1	0	0.000001
0.0505	0	0	0	0.999	0.8
0	0	0	0	0.001	0.199999

TABLE VII
A MATRIX FOR WEEKS 125-400 OF MOCK DATA

The STL COVID-19 data reveals four distinct phases [Fig. 3], of which we will analyze the first three and highlight their differences. In Section A, covering the period before the Delta wave, COVID-19 cases increased steadily, with a peak around January 2021. We modeled this phase using the *fmincon* function and got a resulting A matrix [Table II]. Using this matrix, we generated the a graph displayed in Fig. 4.

This graph demonstrates the capability of *fmincon* in modeling data. While it provides a reasonable estimate, it does not capture the smaller fluctuations in the data.

We applied a similar process in Sections B and C, using *fmincon* to generate an optimized A matrix for each. We chose to model these sections separately for three main reasons. First, as shown in the initial graph, data trends shift noticeably across the four sections, necessitating distinct models since our matrices do not vary with time. Second, viruses like COVID-19 evolve over time, often developing new strains with unique transmission characteristics. Creating separate matrices for each section allowed us to capture these differences and gain specific insights into the behavior of each strain. Lastly, COVID-19 policies and government interventions, such as the introduction and widespread adoption of vaccines, evolved significantly over time. Figures 5 and 6 present our *fmincon* models for Sections B and C, respectively.

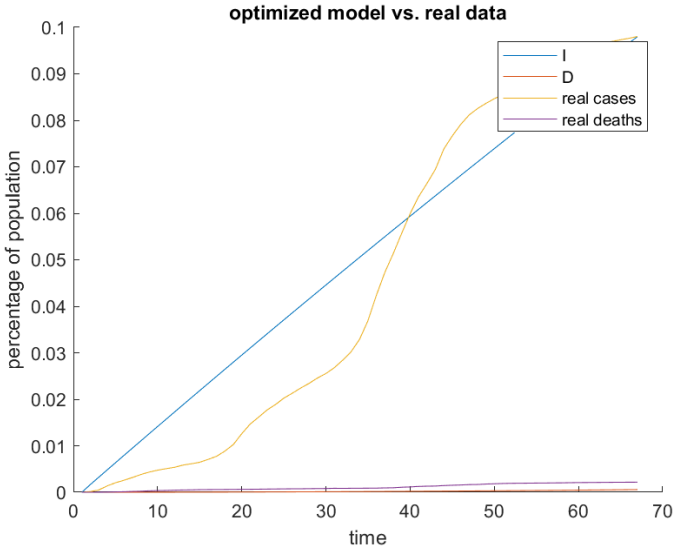

Fig. 4. Section A: Model vs Actual Data

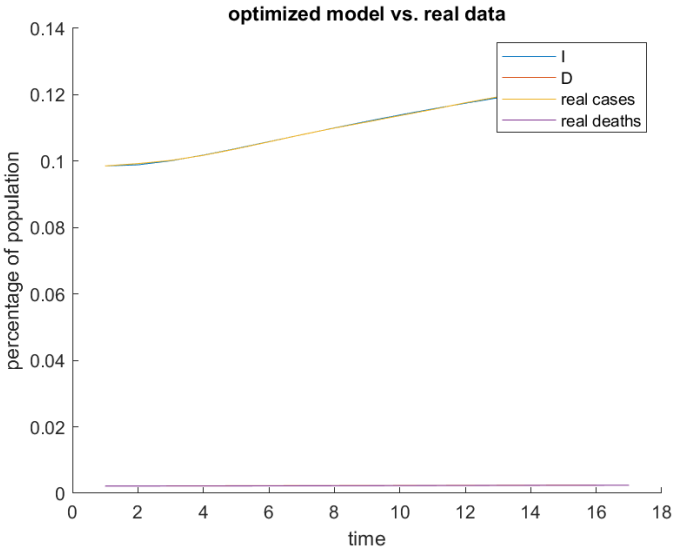

Fig. 5. Part 2 Section B: Optimized Model vs. Real Data

Fig. 6. Part 2 Section C: Optimized Model vs. Real Data

Figures 5 and 6 illustrate that while the model accurately represents the data in Section B, corresponding to the Delta wave, it struggles to capture some of the broader trends present in Section C, the Omicron wave. This discrepancy highlights a significant limitation of the SIRD model—its tendency to oversimplify, which limits its effectiveness in modeling the complexities of real-world COVID-19 data.

As a virus evolves, it often becomes more contagious and less lethal, as its primary objective is to replicate and survive within the host population for as long as possible. We produced 2 matrices that successfully modeled the individual Delta (Table III) and Omicron (Table IV) waves. There are three big differences in those matrices we are paying attention to. First is the I to D element for both. According to the ADA [3], the symptoms of the Delta variant are more severe than the ones

of the Omicron variant resulting in a higher risk of severe illness and death. This shows up in the optimized A model through the infected to dead element. In the Delta model, this value is 0.0002 which is two times the size of the element in the Omicron model which is 0.0001. This matches up with the ADA data since there are more people that have been infected with the Delta variant that end up dying compared to the Omicron variant. Also, according to ADA, Omicron is more transmissible than Delta. This shows up in the recovered to the infected element of both models since we expect the number of reinfected people after having the virus would increase with Omicron since it is more transmissible. In our A matrices, this value was 0.0937 in Delta and 0.2635 in Omicron. The Omicron value is higher than the Delta value which is what we expect with the difference between the strains since Omicron is more transmissible and you would be more likely to get reinfected. The difference in the infected to recovered element suggests a change in policies between the two waves. This element in the Delta model was 0.2745 and was 0.5312 in Omicron. This means that the amount of people recovering in Omicron once infected is greater than that of Delta, meaning that the amount of people are recovering faster with Omicron. This could be due to the increase of vaccines between the two periods [Fig 1].

Figure 1 shows the difference in vaccination rates in the 2 time periods for the different waves of COVID. The Delta wave is on the left half of the graph before October 26th, and the Omicron wave is on the right half after October 26th. There is a clear increase in the magnitude of vaccines between the two waves which could be an explanation of the increase in recovery rate from the Delta to Omicron matrices.

In part (iii) of this analysis, we implemented a quarantine policy for infected individuals, requiring those who test positive to self-isolate at home for 10 days before rejoining society. This policy aims to reduce recovery time for infected individuals and lower the infection rate by limiting contact between infected and healthy individuals.

This would manifest in our model by changing the susceptible to infected element, the recovered to infected element, and the infected to recovered element. Since the recovery time

would decrease, the infected to recovered element would increase. Then, since the number of people getting infected decreases, the recovered of infected and susceptible to infected elements also decreases. However, the susceptibility to infected elements would decrease more than the recovered to infected elements since the rate at which recovered people get infected is already lower because they have previously been exposed to the virus and have immunity. This quarantine policy is both, realistic and widely implemented globally, including in St. Louis [4]. The observed effects of similar policies align with the changes in infection and recovery rates discussed [5].

One limitation of our methods in Part 2 of this project is that our models were specifically tailored to match data from St. Louis. While this approach provides valuable insights for St. Louis, it may not be directly applicable to other cities across the country or the world. Without analyzing case data from a broader range of areas, our models' accuracy may be limited, and mandates deemed effective for St. Louis may not produce the same results in other regions.

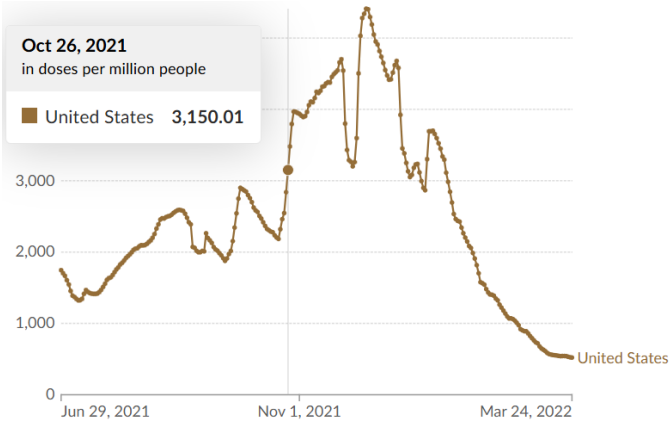

Fig. 7. Vaccination rates in America Delta vs. Omicron

Another limitation of our model arises from the oversimplification of the SIRD and linear dynamical systems framework we used. Our model does not account for any inputs, meaning it overlooks population movement into and out of counties, as well as travel patterns. These factors are crucial, as they are present in the real data but not reflected in our model. While we use a linear dynamic system without inputs, the full equation of a linear dynamic system is:

$$x_{t+1} = Ax_t + u_t \tag{2}$$

The u term would be inputs. If we were to use this rather than no inputs for our model, we could better estimate the amount of cases and deaths which would allow for a more accurate model and more accurate conclusions to be drawn from it.

C. Part 3

In Part 3, we incorporated the effects of vaccination and breakthrough infections into our simulations. The transition rate from vaccinated individuals to breakthrough cases is analogous to the transition rates from susceptible to infected and recovered to infected. However, we expect the rate of vaccinated individuals who experience a breakthrough infection ($V \rightarrow V_b$) to be lower than the rate of susceptible individuals becoming infected ($S \rightarrow I$), as one of our model assumptions is that vaccination reduces the likelihood of infection. This expectation is reflected in our model matrix (see Table VI), where the $V \rightarrow V_b$ transition rate is 0.001, half the rate of $S \rightarrow I$ (0.002), aligning with our assumption of reduced infection rates due to vaccination.

Additionally, we expect differences in mortality rates between vaccinated and unvaccinated individuals, specifically between the infected-to-deceased transition and the breakthrough-infected-to-deceased transition. In our model, the rate of $I \rightarrow D$ is 0.00001 while the rate of $V_b \rightarrow D$ is 0.000001. This decrease in mortality rate for vaccinated

individuals aligns with the anticipated protective effect of the vaccine in reducing COVID-related deaths [**?**].

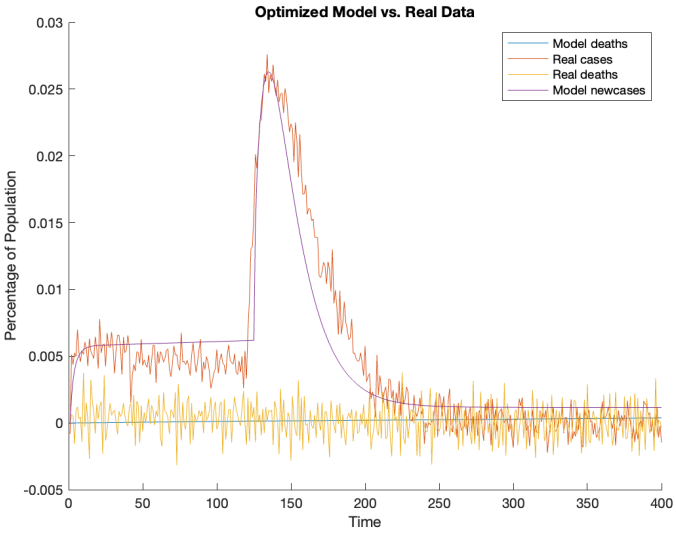

Fig. 8. Optimized Simulation vs Real Data

Lastly, we established a fixed vaccination rate for the population, represented by the transition rate from susceptible to vaccinated ($S \rightarrow V$) in the matrix, which was set at 5.05%. This rate is both reasonable and aligns with our model's assumptions.

As in Part 2, a limitation of our model in Part 3 lies in its oversimplification, particularly due to the absence of inputs in the linear dynamic system we used for modeling. This creates challenges, as our mock data does not represent a steady state, regardless of the conditions under which it is modeled.

IV. Conclusion

Drawing our analysis to a close, our modeling framework highlights the complex dynamics of COVID-19 transmission and policy impacts across the United States. Amid the urgency of the pandemic, where trillions of dollars were spent on response efforts, understanding and mitigating the spread of COVID-19 became a priority. Using the provided data in combination with MATLAB analysis we developed a SIRDVV_b model that captures how different states - susceptible, infected, recovered, deceased, vaccinated, and breakthrough infections - interact with varying policies and conditions. Our design strategy, which incorporated iterative adjustments to optimize the simulation of the given datasets, was structured to test how different scenarios impact transmission and recovery rates. This quantitative approach allowed us to approximate pandemic patterns with meaningful accuracy. A strength of our study lies in its ability to highlight how targeted interventions might influence outbreak dynamics. However, a key limitation is the SIRD model's simplified structure, which may restrict accuracy by excluding additional variables, such as those accounting for regional movement and broader policy impacts. This research contributes to the broader understanding of COVID-19 transmission and the effectiveness of crisis response measures. Our findings underscore the importance of adaptive and regionally tailored public health strategies to manage future outbreaks effectively. Future work could expand this framework by incorporating additional factors, such as socioeconomic data and healthcare access, to improve predictive accuracy. In conclusion, as we face a future where public health events become more frequent, integrating infection trajectory modeling into local and national planning will be essential for fostering community resilience and minimizing future pandemic impacts.

References

[1] WHO, "COVID-19 cases — WHO COVID-19 dashboard," datadot, 2024. https://data.who.int/dashboards/covid19/cases.

[2] U. S. G. A. Office, "COVID-19 Relief: Funding and Spending as of Jan. 31, 2023 — U.S. GAO," www.gao.gov, Feb. 28, 2023. https://www.gao.gov/products/gao-23-106647.

[3] "Omicron Vs Delta Comparison: Risks, Difference, Treatments — Ada Health," Ada, Jun. 09, 2023. https://ada.com/covid/covid-19-omicron-vs-delta-symptoms/

[4] "City of St. Louis, MO," stlouis-mo.gov, 2024. https://stlouis-mo.gov/government/departments/health/communicable-disease/covid-19/guidance/isolation-quarantine-transmission-guidance-august-2023.cfm (accessed Nov. 08, 2024).

[5] "Early COVID-19 shutdowns helped St. Louis area avoid thousands of deaths," ScienceDaily, 2021. https://www.sciencedaily.com/releases/2021/09/210901113710.htm (accessed Nov. 08, 2024).

[6] Brechje de Gier et al., "Effect of COVID-19 vaccination on mortality by COVID-19 and on mortality by other causes, the Netherlands, January 2021- January 2022," Vaccine, vol. 41, no. 31, Jun. 2023, doi: https://doi.org/10.1016/j.vaccine.2023.06.005.