MAT292 Project - Modeling COVID-19 Mitigation

Group 13 - Robyn van Heerden, Amy Liu, Andrew Lu, Chi Lan Tran

October 3, 2025

Abstract

This paper evaluates the effectiveness of SIRU and SEIQR epidemiological models in the context of modeling the COVID-19 pandemic in the United States. While the SEIQR model accounts for quarantined individuals, it is only effective for periods of a few days, making it ineffective for long term covid predictions. On the other hand, SIRU uses reported case and vaccination data to improve adaptability over longer periods. However, both models do not allow for the population in the susceptible (S) category to grow outside of the birth rate, limiting their ability to capture resurgences. To address this limitation, we propose a modified SIRU model that introduces renewed vulnerability caused by new variants and reduced immunity. All models were implemented in MATLAB and evaluated against CDC case data using Mean Absolute Error (MAE) and retrospective peak analysis. Although the proposed model exhibits higher MAE, it accurately captures COVID-19 surges, with zero error in the first peak timing and less than 9% error for the second. Moreover, the peak magnitude error for the first surge reduced from 64% in the SIRU model to 9% in the proposed model. These results show that allowing re-entry into the susceptible population significantly improves a model's ability to predict long-term resurgence behaviour.

1 Introduction

The COVID-19 pandemic has been one of the most significant health crises of modern times, which makes understanding its transmission dynamics both scientifically important and socially relevant. Existing models, such as the basic SIR model [1], have already been adapted and utilized to analyze the spread of COVID-19. While these frameworks provide a useful foundation for understanding disease transmission, they are far too simple to fully capture the real-world complexities of COVID-19 [1]. Key limitations include not accounting for the significant preventative measures that governments put in place to slow down the spread of the pandemic or the inability to reflect surges in cases due to factors like new emerging COVID variants. To address these limitations, we created our own model for COVID-19 in the United States by first analyzing and building on two distinct models (SIRU and SEIQR). Our model focuses on the U.S. because of its large, diverse population and publicly available data, making it ideal for model evaluation and comparison.

Reliability is also a key issue when model results are used to inform real-world decisions. Two methods were used to assess the reliability of select models and our own model: Mean Absolute Error (MAE) and Retrospective Evaluation.

2 Mathematical Model or Theoretical Foundation

We chose to study compartmental models, since they are often created using a system of differential equations [2]. Compartmental models divide the observed population into separate compartments, which can be represented mathematically using a system of differential equations. We chose to study two compartmental models to inform the creation of our own model: SIRU and SEIQR. These models build upon the foundational principles of the basic SIR model, which will be presented first to provide context for the more complex frameworks.

2.1 SIR Model

The SIR (susceptible, infected, recovered) model is the most basic model and is the basis for all other compartmental models. It utilises three compartments and simulates the change in the size of each compartment over time. The SIR model can be represented by the following set of differential equations [1]:

$$\frac{\mathrm{d}S}{\mathrm{d}t}(t) = -\beta S(t)I(t), \tag{1}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t}(t) = \beta S(t)I(t) - \gamma I(t), \tag{2}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t}(t) = \gamma I(t). \tag{3}$$

Where β and γ are rates at which people are infected and recover. These parameters are often estimated based on available transmission data about the disease.

While the SIR model can serve as a starting point to analyze and predict the spread of COVID-19, preventative measures are not accounted for in the simple model, which can greatly affect the outcomes predicted by the model [3]. Two potential models that account for government actions to slow the spread of the disease are SIRU and SEIQR.

2.2 SIRU Model

The SIRU model (susceptibles, asymptomatic infectives, symptomatic reported, asymptomatic unreported) [4] accounts for government intervention in two ways: with a varying transmission rate τ and a parameter for vaccination rates $v(t)$. The SIRU model is particularly powerful because its parameters depend on the number of daily infections as well as how much of the population has been vaccinated, allowing for the model to dynamically correct itself as the disease spreads. The SIRU model can be represented with the following set of differential equations, with parameters explained in table 1:

$$\frac{\mathrm{d}S}{\mathrm{d}t}(t) = -\tau(t, S(t), I(t), R(t)) - v(t)S(t), t \geq t_0, \tag{4}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t}(t) = \tau(t, S(t), I(t), R(t)) - (v_1 + v_2)I(t), t \geq t_0, \tag{5}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t}(t) = v_1 I(t) - \eta R(t), t \geq t_0, \tag{6}$$

$$\frac{\mathrm{d}U}{\mathrm{d}t}(t) = v_2 I(t) - \eta U(t), t \geq t_0, \tag{7}$$

Symbols	Value	Description
$\tau(t, S(t), I(t), R(t))$	See **3. Methodology**	Transmission Rate
t_0	7 (7 March 2020)	Initial time
$v(t)$	See **3. Methodology**	Rate $S(t)$ removed from infection pool due to vaccinations
v_1	0.25/7	Rate $I(t)$ become symptomatic reported $R(t)$
v_2	0.75/7	Rate $I(t)$ become symptomatic unreported $U(t)$
η	1.0/7	Rate $R(t), U(t)$ removed from their pools

Table 1: Parameters with their values & descriptions. Constants are based on values from the paper, which are based on CDC data [4]

The transmission rate τ depends on the daily number of cases as well as social distancing measures, while the vaccination parameter acts by removing people from the susceptible pool at a rate of $v(t)$ dependent on national vaccination rates at a given time. The daily vaccination rate can be set to zero after a certain percentage of the simulated population becomes vaccinated, allowing the model to predict how different national vaccination rates can affect the spread of COVID-19. These two parameters and their implementation are discussed in further detail in the Methodology section.
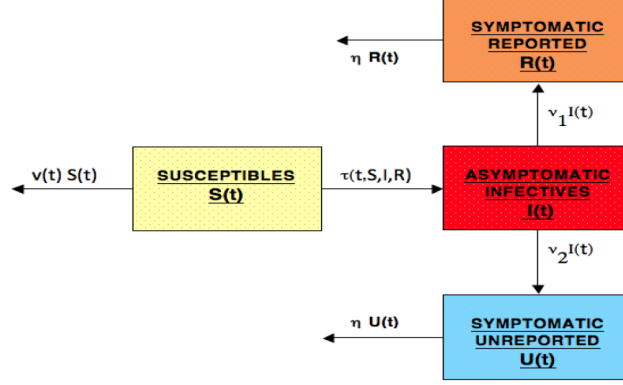
Figure 1: Flowchart for the SIRU model, with arrows representing the flow of people between compartments and labelled by corresponding parameters.

2.3   SEIQR Model

The SEIQR model (susceptible, exposed, infected, isolated, recovered) [5] extends the classical SIR model by introducing both an exposed compartment and isolation compartment to account for the transitions between states and the effects of isolation.

The exposed compartment $E(t)$ represents individuals who have exited the susceptible group but are not yet considered infected. This intermediate state introduces a delay to prevent an instantaneous transfer from $S(t)$ to $I(t)$. The isolation compartment $Q(t)$ represents infected individuals who have been removed from contact with the susceptible population through intervention measures such as quarantine or social distancing. Individuals in this compartment are assumed to have a negligible contribution to the rate of change of the infected population.

The SEIQR model can be represented by the following system of differential equations [5]:

$$\frac{dS}{dt} = A - \mu S - \beta S(E + I), \tag{8}$$

$$\frac{dE}{dt} = \beta S(E + I) - \pi E - (\mu + \gamma)E, \tag{9}$$

$$\frac{dI}{dt} = \pi E - \sigma I - \mu I, \tag{10}$$

$$\frac{dQ}{dt} = \gamma E + \sigma I - \theta Q - \mu Q, \tag{11}$$

$$\frac{dR}{dt} = \theta Q - \mu R. \tag{12}$$

where each parameter is defined as follows:

Symbols	Description
A	Recruitment rate of susceptible individuals (e.g., births)
μ	Natural death rate plus disease-related death rate
β	Rate at which susceptible population moves to infected and exposed class
π	Rate at which exposed population moves to infected class.
γ	Rate at which exposed individuals are isolated
σ	Rate at which infected individuals are isolated
θ	Recovery rate of isolated individuals

Table 2: Parameters and descriptions for the SEIQR model

2.4 Own Model - SIRU Model with Surge

We based our own model on the SIRU model over the SEIQR model. The SEIQR model struggled when attempting to predict daily infections projected more than 30 days into the future. Since SEIQR only relies on a single set of initial conditions, it was unable to account for the evolving factors as the pandemic progressed. On the other hand, SIRU utilizes reported daily infections and vaccinations to guide the model and improve its performance. The accuracy of both models is explored in the Discussion section.

The current SIRU model functions well within its intended scope (projecting only until the end of 2022). However, the model breaks down when projected further since the number of people in the Susceptibles pool $S(t)$ goes to zero as time increases. This is because the differential equation for $S(t)$ features only people being removed from the pool, as seen in Equation 4. This problem can be addressed by imposing an increase to the Susceptibles pool $S(t)$ at a time t_{surge}, which represents a sudden increase in the number of people can can be infected. Physically, surges can represent situations like new outbreaks of COVID variants or reductions in vaccine efficacy.

For our model, we utilized the same system of differential equations (Equations 4 - 7) as the SIRU model while adding a surge at $t_{surge} = 580$ and magnitude $S_{surge} = 100 \times 10^6$. This surge is meant represent the introduction of the Omicron variant to the United States along with reduced social distancing, resulting in a large fraction of the population becoming susceptible to being infected again.

3 Methodology

3.1 Computational/Numerical Methods

SIRU Model

The SIRU model was coded in MATLAB and `ode45` was chosen as the ODE solver (since it is the recommended solver for most applications [6]. The initial conditions were set to $S_0 = 331.5 \times 10^6$ (population estimate of USA) , $I_0 = 1, R_0 = 1, U_0 = 1, t_0 = 7$ (March 7, 2020).

Two of the more complicated parameters of this model are Transmission rate (τ) and Rate of Vaccination ($v(t)$). Their key points are discussed below and further details can be found in the Appendix.

Transmission rate (τ) The SIRU model utilizes the daily number of cases to help project future trends. The daily number of cases in the United States was sourced from the CDC's COVID-19 Case Surveillance Public Use Data [7].

Due to the sporadic reporting of COVID, we chose to use the `cdc_case_earliest_dt` dataset, which utilizes the earliest of the Clinical Date (date related to the illness or specimen collection) or the date Received by the CDC to give a best estimate of the actual date(s) of infection. We will also apply further smoothing to the data by taking a rolling weekly average of the reported daily cases, which is a common method utilized in many models [4]. This is discussed further in the Appendix.

The transmission rate has four different forms, depending on the time. For the sake of clarity, ω is taken as 0.03 and further details about how to vary ω are discussed in the Appendix.

For $t < t_D$, where t_D is the last day of available daily reported case data, the transmission rate is given by

$$\tau(t, S(t), I(t), R(t)) = \frac{CS''(t) + CS'(t)}{v_1} + (v_1 + v_2)\left(\frac{CS'(t) + CS(t)}{v_1}\right). \tag{13}$$

Where $CS(t)$ is the continuum cubic spline interpolation of the discrete rolling weekly average number of daily reported cases and is derived in the Appendix.

For $t_D \leq t \leq t_1$, where t_1 represents a date for relaxing social distancing behavior over time, the transmission rate is given by

$$\tau(t, S(t), I(t), R(t)) = \left(\tau(t_D, S(t_D), I(t_D), R(t_D))\right)\left(\frac{(I(t) + 4R(t))S(t)}{(I(t_D) + 4R(t_D))S(t_D)}\right). \tag{14}$$

For $t_1 \leq t \leq t_2$, where t_2 represents a date for *further* relaxing social distancing behavior over time, the transmission rate is given by

$$\tau(t, S(t), I(t), R(t)) = \left(1.0 + \omega(t - t_1)\right)\left(\tau(t_D, S(t_D), I(t_D), R(t_D))\right)\left(\frac{(I(t) + 4R(t))S(t)}{(I(t_D) + 4R(t_D))S(t_D)}\right). \quad (15)$$

Finally, for $t_2 \leq t$, there is no further change to social distancing behavior, and the transmission rate is given by

$$\tau(t, S(t), I(t), R(t)) = \left(1.0 + \omega(t_2 - t_1)\right)\left(\tau(t_D, S(t_D), I(t_D), R(t_D))\right)\left(\frac{(I(t) + 4R(t))S(t)}{(I(t_D) + 4R(t_D))S(t_D)}\right). \quad (16)$$

Rate of Vaccination ($v(t)$)   Individuals are removed from the Susceptibles (S) pool at a rate $v(t)$ given by:

$$v(t) = 0.95\frac{v_{daily}(t)S(t)}{S(0) - CV(t)}, \quad (17)$$

where v_{daily} is the number of people vaccinated that day and is taken as $1,000,000$ after the last date of available vaccination data, t_{Vmax}, and $CV(t)$ is the cumulative number of people vaccinated. For the sake of clarity, t_{Vmax} is taken as 621 to represent a 90% vaccinated population and further details about how to vary t_{Vmax} are discussed in the appendix.

SEIQR Model

The SEIQR model was implemented numerically in MATLAB to simulate the time evolution of the five compartments. Similar to the SIRU model, the system of ODEs was solved using the `ode45` solver.

Initial conditions for the SEIQR model were chosen to represent the early stages of COVID-19 spread in the United States. These values were specified using a normalized system, where each compartment represents a fraction of the total population. The total equilibrium population is given by

$$N_{eq} = \frac{A}{\mu}. \quad (18)$$

Based on external epidemic modeling studies, COVID-19 was treated as a newly emerged disease with minimal pre-existing immunity in the population. Thus, at the initial simulation time $t_0 = 0$, the susceptible population was assumed to be 99% of the total population. The remaining 1% was distributed among the exposed (0.5%), infected (0.3%), and isolated (0.2%) compartments, following standard early-outbreak modeling assumptions. The recovered compartment was initialized to zero, corresponding to the absence of immunity at the start of the simulation.

The SEIQR system was solved over a 30-day interval, as the model was specifically designed for short-term predictions over this time frame [5]. This allows direct comparison of the generated plots with the results presented in the paper.

3.2   Evaluation of Model Reliability

Two methods were utilized to evaluate the reliability of SIRU, SEIQR, and our own model: Mean Absolute Error (MAE) and Retrospective Evaluation.

3.2.1   Mean Absolute Error (MAE)

Mean Absolute Error is a common metric utilized to measure model prediction accuracy. MAE calculates the average of how far off the predicted number of daily cases are from the actual number of daily reported cases and can be calculated by

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|, \tag{19}$$

where n = number of data points, $\hat{y}_i$ = predicted value, and y_i = actual value.

In this case, the predicted value is the number of daily cases predicted by the ODE system and the actual value is the 7-day rolling average of the daily reported cases from the CDC (derivation features in the Appendix). For both models, MAE was only calculated for data points that the models were intended to predict (SIRU: end of 2021, SEIQR: 30 days). The evaluation of MAE beyond the intended projection period is not meaningful, as both models predict zero or a constant amount of reported cases after their final forecast date.

3.2.2 Retrospective Evaluation

While MAE is a useful metric, it only provides a global view of how well a model performs. Much of the usefulness of models stems from their ability to predict key dates, such as peak death or hospitalization rates. An emerging technique to evaluate how well models predict these key dates is Retrospective Evaluation. Two key metrics utilized by Retrospective Evaluation are `PeakDate_Error` and `PeakMagnitude_Error` of the number of daily cases. `PeakDate_Error` measures the error between the predicted versus actual date of peak cases, while `PeakMagnitude_Error` measures the error between the predicted versus actual magnitude of peak cases. While the formula for these functions can be quite complex [8], we can capture the core idea by simply taking the difference between predicted and actual peak dates/magnitudes. In other words, we will represent `PeakDate_Error` and `PeakMagnitude_Error` by

$$\texttt{error} = \frac{|\hat{y}_i - y_i|}{y_i}, \tag{20}$$

where $\hat{y}_i$ = predicted value, and y_i = actual value.

Key peaks used to evaluate model reliability can be chosen by eye. Two key peaks were chosen: day 544 with 156767 cases, and day 677 with 708495 cases. The first peak was used to evaluate how well the SIRU model and our own model performed as a baseline. The second peak was used to evaluate the ability to predict beyond the first wave.
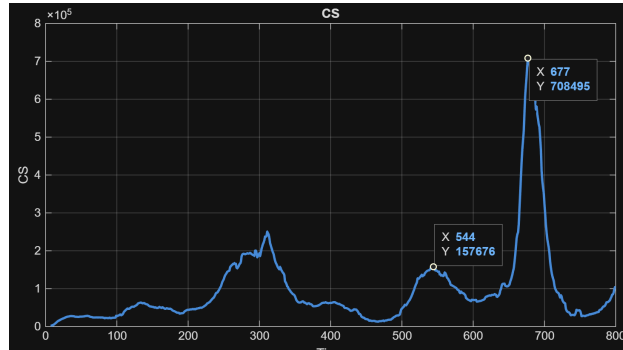

Figure 2: Two peaks shown of 7-day rolling average daily reported cases (CR).

This method was only applied to SIRU and our own model, since the SEIQR model did a poor job of predicting trends and did not exhibit any peaks we could identify.

3.2.3 Note on Qualitative Evaluation of SEIQR

Due to the nature of the SEIQR model, numerical methods of were less helpful in evaluating the models effectiveness, as it was not made to fit a specific population but rather explore the general trend caused by quarantine on the pandemic. We were able to produce graphs that appear very similar to the graphs in the paper (see Section 8.3), however since the paper does not state the values of the parameters that it uses we

could not be certain we had the correct values. The paper also uses the NSFD numerical model to graph the function where we used the `ode` function from MATLAB which uses RK4 and RK5. We suspect these differences account for the differences in the curves we created from the curves featured in the article.

The main drawback of this model is that it does not allow for the recovered population to feedback in the susceptible population, leading to the only increase in the susceptible population to be the birth rate. Additionally this model assumes that everyone in quarantine stays in quarantine, as the only way for the quarantined population to decrease is if they die. This does not account for the people that would break quarantine for various reasons. This lack of mobility between compartments was addressed in the model we proposed.

4   Results

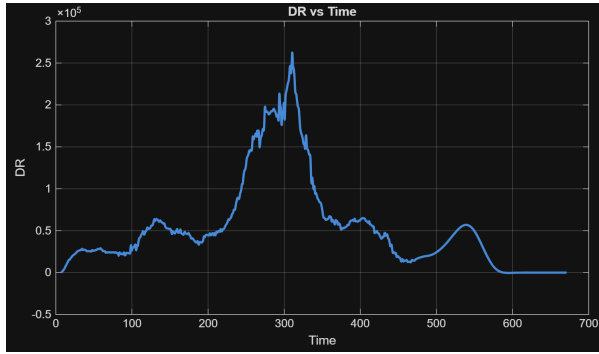The following graphs were obtained from the three models' projections.

Figure 3: SIRU Daily Reported cases, with model projection taking over after day 472.
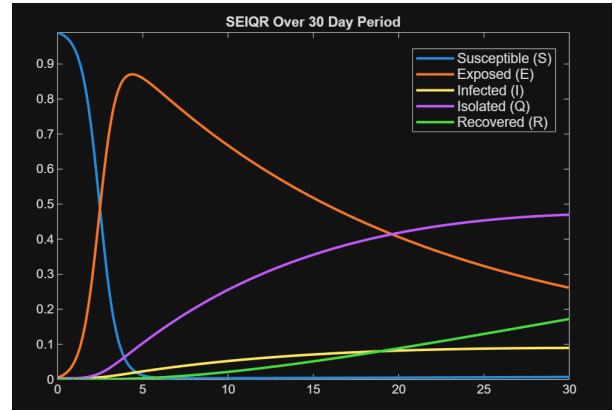

Figure 4: SEIQR model outputs from given starting parameters over 30 days
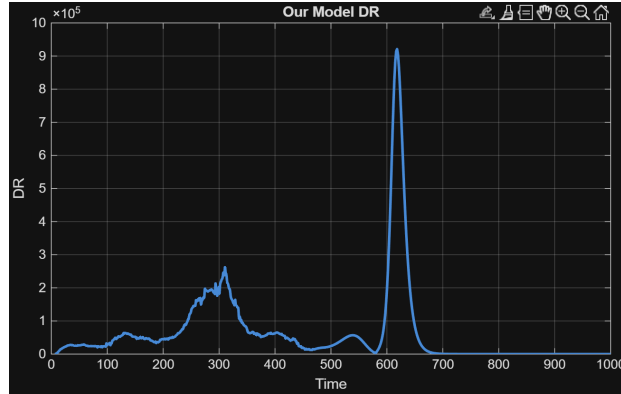

Figure 5: Our model Daily Reported cases, with model projection taking over after day 472, accurately capturing second peak from actual data seen in Figure 2

4.1 Model Reliability

Metric	SIRU Model	SEIQR Model	Own Model
MAE	15171.02	11376.81	72621.17
`PeakDate_Error_1`	$\frac{|540-544|}{544}=0.735\%$	N/A	$\frac{|544-544|}{544}=0\%$
`PeakDate_Error_2`	N/A	N/A	$\frac{|618-677|}{677}=8.71\%$
`PeakMagnitude_Error_1`	$\frac{|56975-156767|}{156767}=63.7\%$	N/A	$\frac{|56975-156767|}{156767}=63.7\%$
`PeakMagnitude_Error_2`	N/A	N/A	$\frac{|921036-708495|}{708495}=30.0\%$

Table 3: Model Reliability Comparison. Metrics include Mean Absolute Error (MAE) and relative errors for peak date and magnitude.

5 Discussion

Evaluating our own model through Retrospective Evaluation, we find that our own model was able to accurately predict the peak of covid cases that occurred in early 2022. From Table 3, it can be seen that the `PeakDate_Error_1`, which is for the first peak, is better for our model than the SIRU model. Additionally, our model performs the same as SIRU in terms of `PeakMagnitude_Error_1`. Notably, our model had relalivelty low `PeakDate_Error_2`, considering the model is predicting long term (around 150 days ahead). Additionally, the error for `PeakMagnitude_Error_2` is even lower than for the first peak, correctly predicting what would be the most severe peak for the entire pandemic.

However, our model performed poorly in terms of MAE. This suggests that MAE is not necessarily an accurate method to evaluate models, as it punishes models that predict peaks that are not on time rather than those who predict no peaks at all. The predicted peak date was also 59 days ahead of the actual peak date, suggesting that our current S_{surge} may be over-correcting for the impact that a new strain like Omicron has on infection rates.

The correct quantitative behavior of the model, as shown by low `PeakDate_Error` and `PeakMagnitude_Error`, suggests that the addition of a surge component to the Susceptibles pool of the existing SIRU model could be an effective way increase a model's ability to predict large peaks in daily cases. However, more work needs to be done to increase the accuracy of our current models predictions in terms of MAE. Further work could include a surge term that depends on government action in more detail, like how the SIRU's transmission rate depends on both vaccination rates and social distancing regulations. Further work could also consider decreasing vaccine effectiveness as an alternative to using S_{surge} for a less jarring way to increase the population of the Susceptibles pool.

6 Conclusion

From our investigation, the SIRU and SEIQR models do not sufficiently model COVID-19 for longer periods of time. The major failing in both the SIRU and SEIQR model is that the only way that the susceptible population increases is through the birth rate, when in reality there are several ways that a recovered person may become susceptible again. The model we proposed, SIRU with surge, allows for the susceptible population to meaningfully increase. This increases the time the model is accurate by accounting for new strains or sudden increases in the susceptible population due to various reasons. Our own model also handled peaks in the infected population better than SIRU and more accurately predicted further into the future. Having accurate epidemic models is critical for governments when making decisions on mask or quarantine regulations, so having a dynamic model that can account for these policy changes is imperative. While our proposed model is still not ideal, the more accurate we can become the more equipped we will be when dealing with future outbreaks.

7   References

[1] Hadeel AlQadi and Majid Bani-Yaghoub. Incorporating global dynamics to improve the accuracy of disease models: Example of a covid-19 sir model. *PLOS ONE*, 17:e0265815, 04 2022.

[2] Bastien Reyné, Nicolas Saby, and Mircea T. Sofonea. Principles of mathematical epidemiology and compartmental modelling application to covid-19. *Anaesthesia Critical Care Pain Medicine*, 41:101017, 02 2022.

[3] Yue Xiang, Yonghong Jia, Linlin Chen, Lei Guo, Bizhen Shu, and Enshen Long. Covid-19 epidemic prediction and the impact of public health interventions: A review of covid-19 epidemic models. *Infectious Disease Modelling*, 6:324–342, 2021.

[4] Glenn Webb. A covid-19 epidemic model predicting the effectiveness of vaccination in the us. *Infectious Disease Reports*, 13:654–667, 07 2021.

[5] Anwar Zeb, Ebraheem Alzahrani, Vedat Suat Erturk, and Gul Zaman. Mathematical model for coronavirus disease 2019 (covid-19) containing isolation class. *BioMed Research International*, 2020:1–7, 06 2020.

[6] MATLAB Help Centre. Choose an ode solver.

[7] Centers for Disease Control and Prevention. Covid-19 case surveillance public use data, 2020.

[8] Kimberly A Dautel, Ephraim Agyingi, and Pras Pathmanathan. Validation framework for epidemiological models with application to covid-19 models. *PLoS Computational Biology*, 19(3):e1010968, 2023.

8   Appendix

8.1   7-Day Rolling Average Spline Derivation

A standard method to handle sporadic data reporting, like that of the CDC for number of COVID-19 cases per day, it to take a rolling weekly average. A continuum cubic spline interpolation of the discrete rolling weekly average data points was taken, which is required for `ode45` to function. `ppval` was used to pull discrete data points from the continuous curve for MAE evaluation.

```matlab
function CS = create_CS(daily_cases, t_end)
    weekly_avg = movmean(daily_cases, 7);
    weekly_avg = weekly_avg(1:t_end);
    t_integer = 1:length(weekly_avg);
    CS = spline(t_integer, weekly_avg);
end
```

Figure 6: Code for Rolling Average Spline Curve of Daily Reported cases.

8.2   SIRU Parameter Details

The following parameters can be varied to produce different projections based on assumptions about social distancing regulations and vaccination rates.

8.2.1 Transmission rate (τ) Details

Levels of Social Distacing (ω) Differing levels of social distancing imposed by the government can be represented by ω, with values of $0.03, 0.025, 0.02,$ and 0.15 representing increasing levels of social distancing measures.

8.2.2 Rate of Vaccination ($v(t)$) Details

National Vaccination Rate (t_{Vmax}) Differing national vaccination rates affect the rate at which people move between pools. For the sake of simplcity, the daily number of vaccinations is taken to be 1,000,000 after a certain amount of time and is then set to zero at t_{Vmax}. t_{Vmax} with dates of November 11, 2021, October 26, 2021, and October 9, 2021 represent 90%, 85%, and 80% of the national population being vaccinated, respectively.

8.3 SEIQR Plot Comparison with Paper

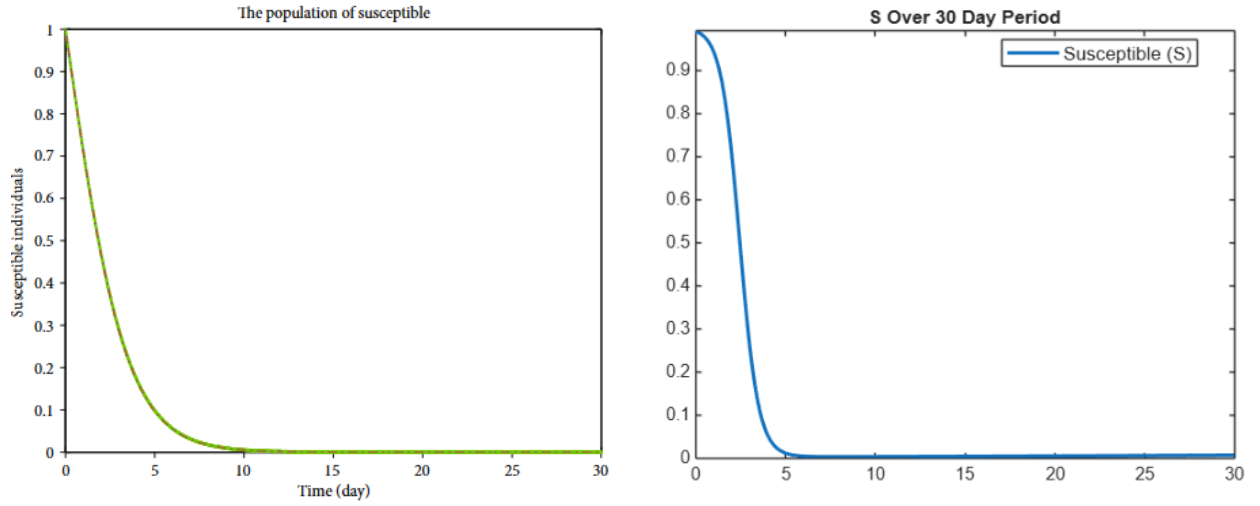The following figures compare our MATLAB outputs with the corresponding plots from Zeb et al. [5].

Figure 7: Susceptible population over time from the SEIQR paper [5] (left) vs our MATLAB implementation (right).
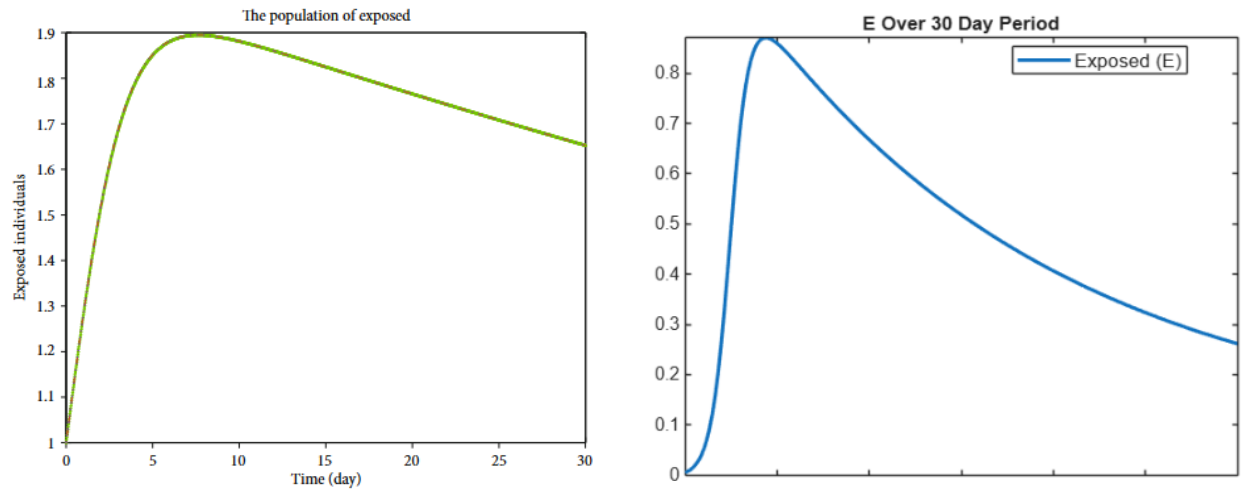
Figure 8: Exposed population over time from the SEIQR paper [5] (left) vs our MATLAB implementation (right).
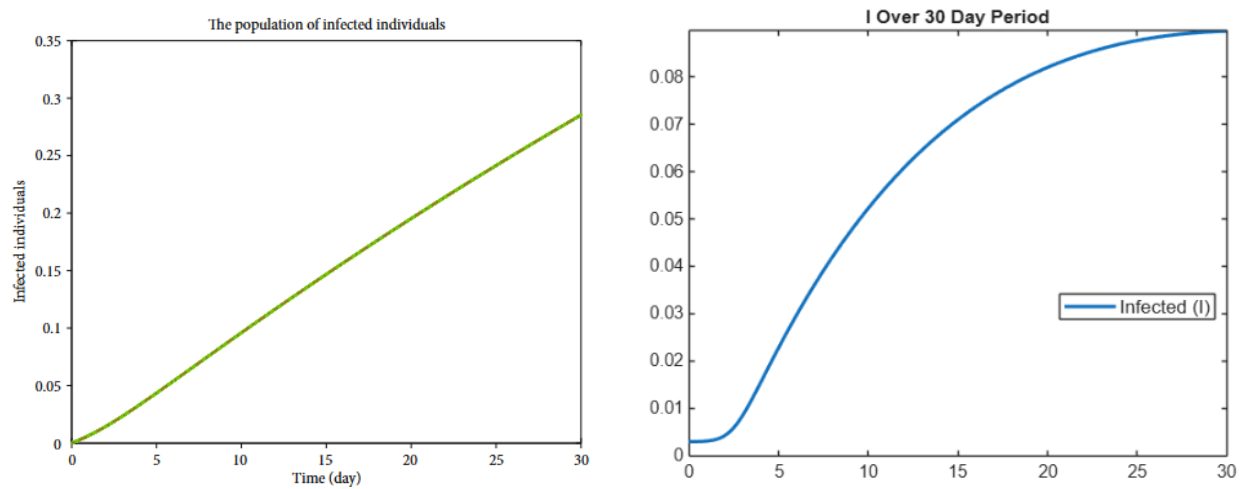

Figure 9: Infected population over time from the SEIQR paper [5] (left) vs our MATLAB implementation (right).
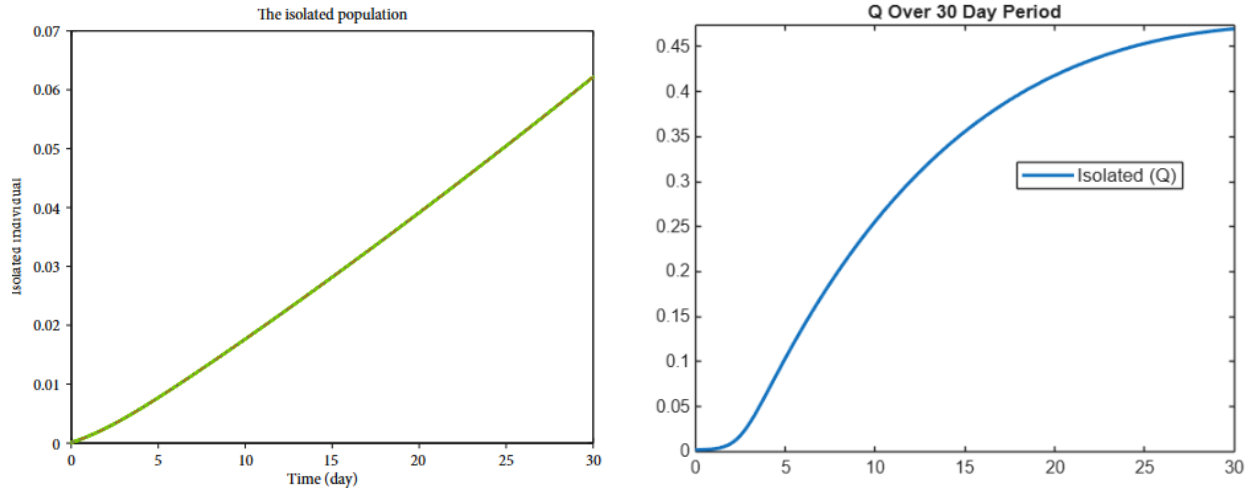
11

Figure 10: Isolated population over time from the SEIQR paper [5] (left) vs our MATLAB implementation (right).
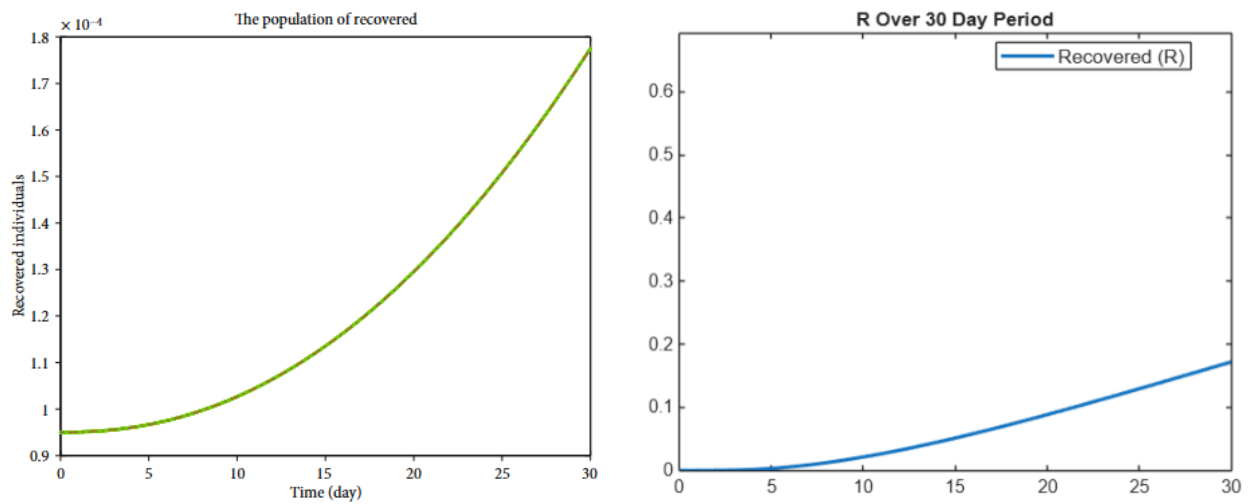

Figure 11: Recovered population over time from the SEIQR paper [5] (left) vs our MATLAB implementation (right).