CoronaVIRES: A Novel SEIRV Compartmental Model for COVID-19

Lay Jain [a], Shinjini Ghosh [b] and Pawan Goyal [c]
Massachusetts Institute of Technology
[a] layjain@mit.edu, [b] shinghos@mit.edu, [c] pawan14@mit.edu

Abstract

Compartmental models, like the SIR (Susceptible-Infectious-Recovered) or SEIR (Susceptible-Exposed-Infected-Recovered) have been popularly used by epidemiologists to model the spread of COVID-19, and predict future statistics. However, the aforementioned models fail to incorporate the recent introduction of vaccinations that have reduced the mortality rate of COVID patients. We introduce *CoronaVIRES*, an SEIRV (Susceptible-Exposed-Infectious-Recovered-Vaccinated) model that models vaccines and different vaccination strategies. We estimate model parameters by fitting the model using the provided data-sets, and show how the parameter values conform to our expectations of them. Changing *CoronaVIRES* parameters allows us to model the effects of different policies and vaccination schedules. We demonstrate how *CoronaVIRES* can be used to predict future statistics by comparing it to the Baseline SEIR model.

1 Introduction

COVID-19 has ravaged the world over all of 2020, with over 117.3 million cases and 2.6 million deaths, worldwide lockdowns, mask-wearing mandates, prohibited gatherings, and a pause to normal life as we know it. It is thereby of no surprise that investigating every factor related to COVID-19 as well as attempting to model the spread of the disease buoyed to the forefront of research needs immediately, especially in the domains of public policy making. One of the leading benchmark papers in this area [1] look into three models—exponential growth, self-exciting branching process, and the susceptible–infected–resistant (SIR) compartment model, one that has been time-tested for its use in epidemiological modeling.

The SIR model assumes a total population of size N where S is the total number of susceptible individuals, I is the number of infected individuals, and R is the number of resistant individuals. For infectious diseases, a new compartment soon came into play - E, the exposed people. This, too, has been previously used to fit the death data from the 1918

influenza pandemic [2], during which governments implemented extensive social distancing measures, including bans on public events, school closures, and quarantine and isolation measures, much like current times. Hence, the SEIR model naturally made a comeback in modeling the spread and effects of COVID-19. However, all these models assume no existence of vaccines or antiviral therapies. We noticed high negative correlation between vaccine counts in a country and the mortality rate. In light of this observation and multiple new vaccines being approved and administered in countries over the world, we introduce a new model, the SEIRV model, which also includes a vaccinated compartment in addition to the susceptible, exposed, infectious and recovered ones.

2 The *CoronaVIRES* Structure

The CoronaVIRES (aka SEIRV) model includes the standard SEIR/SIR model assumptions of a fixed population sample size of N, where all people start off susceptible to the disease. Eventually, they will all end up either recovered from COVID (but unvaccinated), fully vaccinated, dead, or susceptible (never caught the disease, have not taken the vaccine). The model dynamics are in the figure below. V_1 and V_2 represent people after the first and second dose of the vaccine, respectively. We can just eliminate stage V_1 for people obtaining single-dose vaccines.

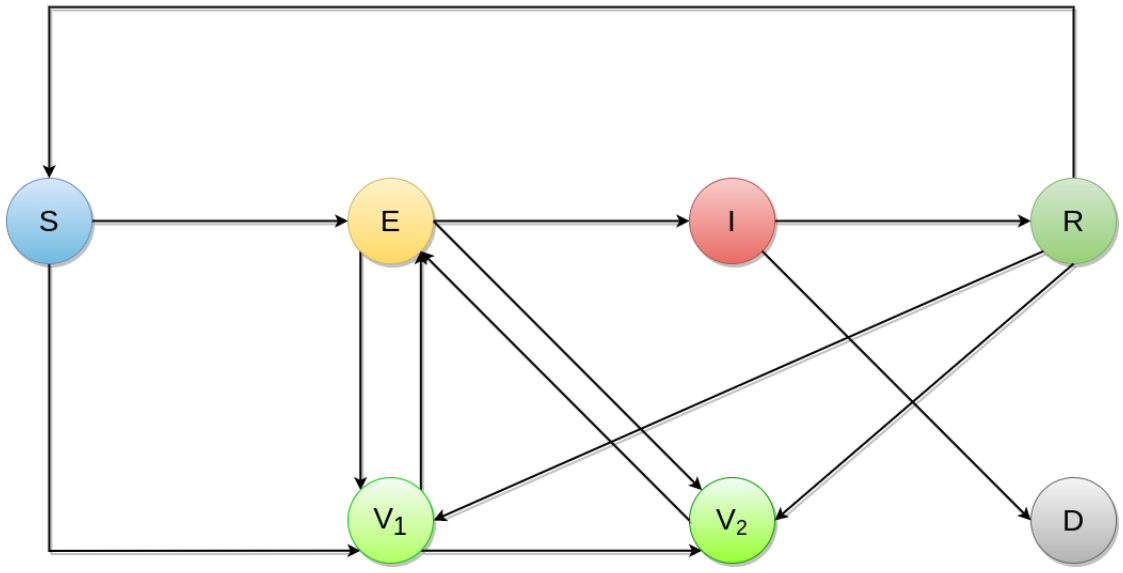


Figure 1: SEIRV Model Dynamics

The following are the SEIRV model equations, which thoroughly explain each dynamic above.

$$\dot{S} = \alpha R_S - \frac{S}{N}\beta I - \frac{S}{N}\chi E - \rho S$$

$$\dot{V_1} = \rho S + \rho R_S - \frac{V_1}{N}\beta I - \frac{V_1}{N}\chi E - \phi V_1$$

$$\dot{V_2} = \phi V_1 + \phi' R_1 + (1-\delta_2) I_2 - \frac{V_2}{N}\beta I - \frac{V_2}{N}\chi E$$

$$\dot{E_1} = \frac{V_1}{N}\beta I + \frac{V_1}{N}\chi E - \theta E_1$$

$$\dot{E_2} = \frac{V_2}{N}\beta I + \frac{V_2}{N}\chi E - \theta E_2$$

$$\dot{E_S} = \frac{S}{N}\beta I + \frac{S}{N}\chi E - \theta E_S$$

$$\dot{I_1} = \theta E_1 - \delta_1 I_1 - (1-\delta_1) I_1$$

$$\dot{I_2} = \theta E_2 - \delta_2 I_2 - (1-\delta_2) I_2$$

$$\dot{I_S} = \theta E_S - \delta_S I_S - (1-\delta_S) I_S$$

$$\dot{R_1} = (1-\delta_1) I_1 - \phi' R_1$$

$$\dot{R_S} = (1-\delta_S) I_S - \rho R_S - \alpha R_S$$

$$\dot{D} = \delta_1 I_1 + \delta_2 I_2 + \delta_S I_S$$

where

$$I = I_1 + I_2 + I_3$$

$$E = E_1 + E_2 + E_3$$

$$N = S + V_1 + V_2 + E_1 + E_2 + E_S + I_1 + I_2 + I_S + R_1 + R_2 + R_S + D$$

3 Estimation Of Model Parameters

The parameters of the model may be considered as proxies to various interactions in the society.

Parameter	Description
α	Temporary Immunity rate
β	Contact and infection rate of transmission per contact from infected class
θ	Transition rate of of exposed individuals to the infected class
δ_i	The death rate of the Infected classes
ρ	The vaccination rate

Our model performs as expected on changing these parameters. For example, increasing interaction β, or decreasing immunity $1\alpha^{-1}$, results in more new positive cases in a day. Increasing vaccinations, ρ, reduces the death count. Similarly, increasing the disease fatality (δ) or reducing vaccine effectiveness $(1-\delta_1, 1-\delta_2)$ increases the death rate.

We use `scipy.curve_fit` in order to fit the model parameters after numerically solving the differential equations. We have fitted the parameter values for a few countries. A comparison of these values would tell us how different countries responded to the pandemic. For example, the β values of US (0.04) and Italy (0.033) are higher compared to other countries (around 0.02). This indicates more mixing among people in these countries, leading to more spread of infection. The δ_2 values tend to be lower than δ_s, indicating that the vaccine indeed has an effect. However, some regions fail to satisfy this general trend. Our model is applicable to places where the population is approximately uniformly vaccinated (There are no clusters of vaccinated vs. non-vaccinated people). Hence, applying CoronaVIRES to the world as a whole would not give good results.

3.1 *Cleaning and Preprocessing the data*

We mainly focused on owid-covid-data and restricted our timeframe to vaccination phase, namely from Jan 1, 2021 till present (we downloaded updated data from the owid github to get the latest numbers). We then looked for top ten countries in terms of number of days since the country started its vaccination. We removed all the entries for which either vaccination or new_cases or new_deaths data was unavailable and then used this final processed data to fit the curve

3.2 *Standardization of y labels*

We used new_death_smoothed_per_million, new_cases_smoothed_per_million and total_deaths_per_million to fit our model. The primary motivation to choose first two was the stability of derivatives over absolute value (thus preferring new_cases over total_cases) while the thought process behind total_deaths_per_million was to give more weightage to predicting correct deaths over correct infected cases. We further standardized our model for unit population (thus dividing all the above relevant statistics by 10^6 and setting $N = 1$ in the above SEIRV model equations). This was a crucial step in fitting our model parameters as it made all hyper parameters in the range (0,1), thus making much easier for scipy.curve_fit to find the right fit.

3.3 *Results*

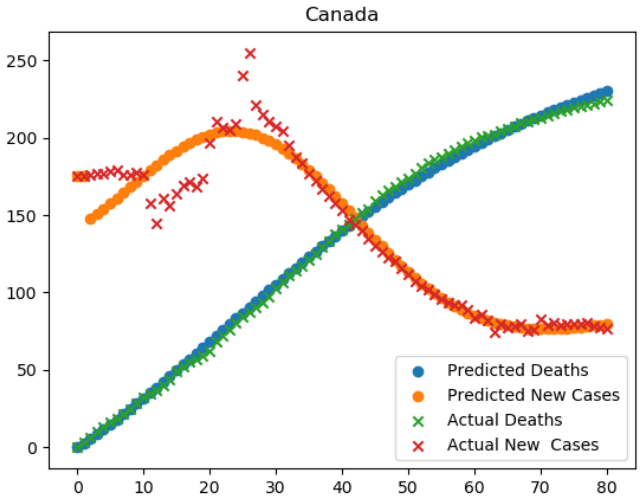


Figure 2: Simulation of SEIRV for Canada

The major countries with significantly large vaccination programs were USA, Canada, Israel, Italy, France, Bahrain, Chile and Denmark. We tried to individually fit the model for each of these countries and the results for Canada, USA, Italy and Bahrain are shown in Fig 2, 3, 4 and 5 respectively.

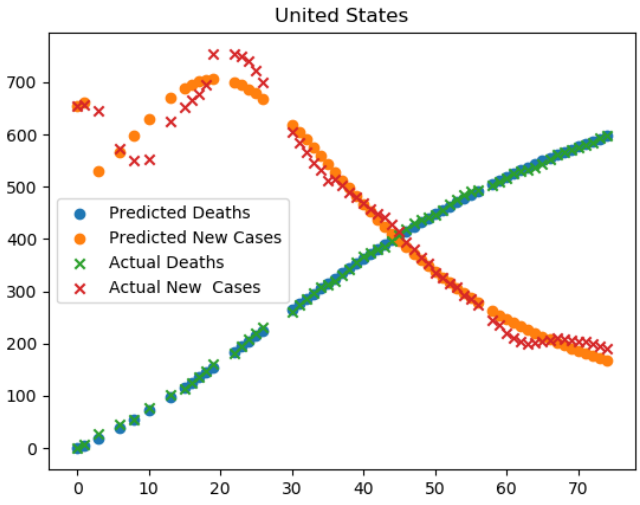


Figure 3: Simulation of SEIRV for USA

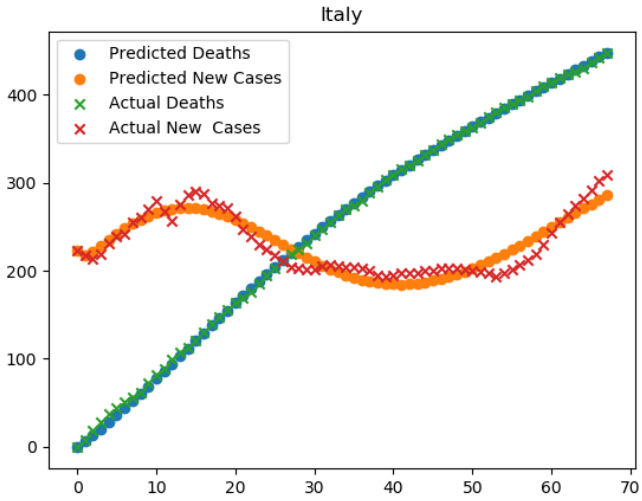


Figure 4: Simulation of SEIRV for Italy

4 Predicting the future

We can use our model to predict future statistics. To this end, we train our model with a fraction of past data-points and compare the predictions with the remaining data-points. We have used the SEIR model (without Vaccinations) as a baseline. We have calculated the root mean squared error for both the models and the results are presented below. In

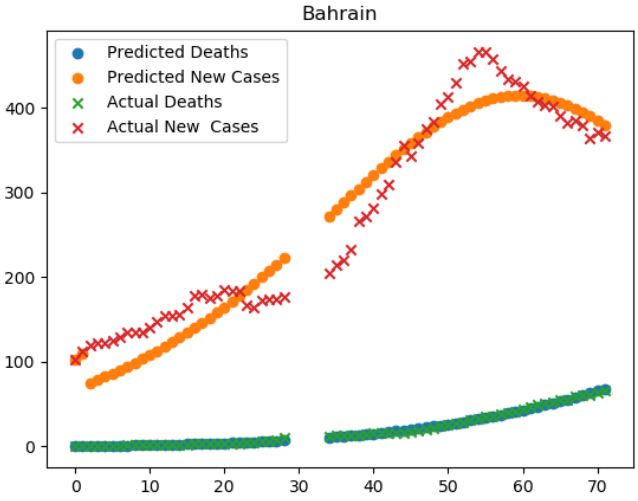
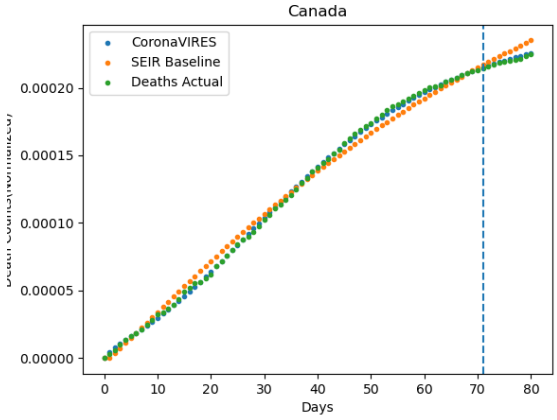
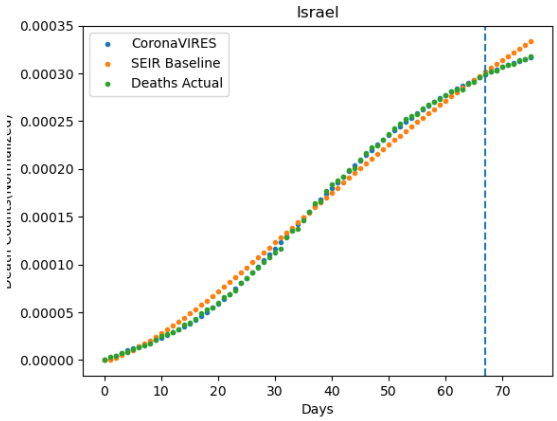


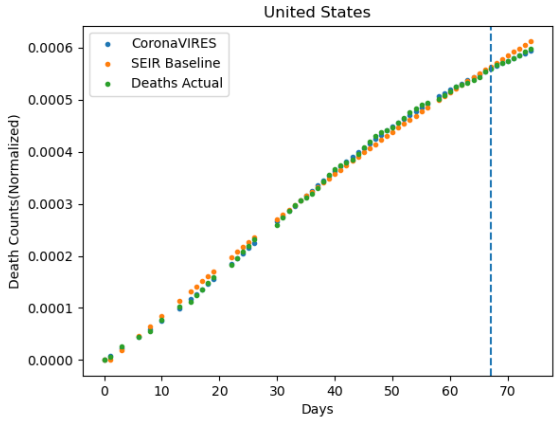
Figure 5: Simulation of SEIRV for Bahrain

order to aid the comparison, we also plot the predictions. In the figures below, the dashed blue line indicates the termination of our training process.

Model Error$\times 10^6$ **/Country**	Chile	United States	Canada	Israel
CoronaVIRES	2.07	2.19	1.23	5.91
SEIR Baseline	9.98	11.22	7.51	10.62







The results above indicate that incorporating vaccinations makes a better predictor for the statistics. A part of this result, however may be due to the fact that adding the vaccine compartment adds a new free parameter for the model. A more rigorous analysis of this comparison would need to take this into account and require more time that we had for the Datathon.

5 Conclusion and Future Work

We have proposed and evaluated *CoronaVIRES*, a compartmental model that incorporates vaccination in order to predict covid statistics. We have showed that this model performs better than a similar SEIR model. *CoronaVIRES*'s usefulness extends beyond what we have had the time to present for this weekend Datatthon. For example, one may examine the effects of different vaccine strategies on the death count. One may also analyze the difference between a two-dose (Pfizer, Maderna, etc.) and a single-dose vaccine (JohnsonJohnson, for

example).

References

[1] Andrea L. Bertozzi, Elisa Franco, George Mohler, Martin B. Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.

[2] Martin C. J. Bootsma and Neil M. Ferguson. The effect of public health measures on the 1918 influenza pandemic in u.s. cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.