

VSEIRD Model for Covid-19 Simulation

Vince Chen
UC San Diego
bic015@ucsd.edu
A13696788

Jiayi Luo
UC San Diego
jil818@ucsd.edu
A13721924

Jin Wu
UC San Diego
gew013@ucsd.edu
A13687390

Abstract

There are many compartmental models in the field of epidemiology, and SEIR model is one of them. The SEIR stands for Susceptible, Exposed, Infectious, and Recovered, with each of them marked as one stage of the infection. The model used four differential equations, with each represents a different stage of the COVID-19 infection, and by solving the equations, it is able to generate the simulation based on different parameters. Overall, SEIR model has done quite well in simulating the COVID-19 in early period. However, as time went on, we found it quite inaccurate after a long period of time, where the model deviates from true COVID-19 case curve. In our project, we would like to improve the model by adding more deterministic stages, such as Vaccination and Asymptomatic, on top of the original SEIR model and make it a more rigorous and accurate model.

1. Background

At the end of the year 2019, a novel SARS virus began to spread across the entire world in a short period of time. Victims of this novel virus were found in different countries from different continents almost at the same time. Till now, the origin or the birthplace of this novel SARS virus is still in debate. The World Health Organization later named this virus as COVID-19[1], short for Coronavirus 2019.

Now, the COVID-19 epidemic has been effectively controlled and the situation that it brings is improving in most of the places. However, the epidemic is still raging in some countries unfortunately. For example, recently COVID-19 raged in India and not only impacted the local economy there but also took many individuals' lives. Luckily, the vaccination process is going on well, and more and more people are getting vaccinated. Hopefully, the COVID-19 could be stopped at the end of this year. **ADD MORE CONTENTS HERE LATER**

2. The SEIR Model

It is crucial to simulate the spread of the COVID-19 so that people could foresee the trends and therefore take different preventive measures to best fight against the COVID-19. Therefore, many different models are studied and used to simulate the COVID-19 epidemic.

SEIR model[2] is one of the mostly used model to simulate the COVID-19. It belongs to compartmental models, which usually simplify the mathematical modelling of infectious diseases. Usually, the population is assigned to compartments with labels, and individuals may progress between different compartments[3]. The order of the labels often shows the flow patterns between the compartments.

The model itself runs with ordinary differential equations(ODE), by solving the equations with different parameters, models can try to predict the outcomes related to the epidemic, such as how exactly a disease spreads, the total number of infected during a certain time, or the duration of the epidemic[3]. Therefore, SEIR model is regarded as a powerful tool in the analysis and the prediction of the COVID-19 and is therefore worthwhile to dive deep and study for.

2.1. Structure

For the SEIR model, it has four different compartments(states), which are Susceptible, Exposed, Infectious, and Recovered.

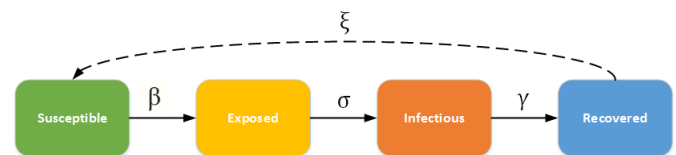


Figure 1: Flow Diagram of the SEIR model

Each compartment mark the transition of an individual involved in COVID-19:

- **Susceptible:** This stage represents the fraction of susceptible individuals, or those who are able to contract

the COVID-19.

- **Exposed:** This stage represents the fraction of already exposed individuals, in which they have been in contact with infected individuals but are not yet infectious.
- **Infected:** This stage represents the fraction of infected individuals. These people are infected themselves and are capable of transmitting the COVID-19.
- **Recovered:** This stage stands for the fraction of those who were infected and recovered now. These people are usually immune to the COVID-19 for a period of time

2.2. Equations and Parameters

The differential equations of the SEIR model are listed below:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$N = S + E + I + R$$

Where β is the transition parameter from Susceptible to the Exposed state, which we usually refer it as the infection rate or the rate of spread. and σ represents the incubation rate, in other word, or the rate that latent individuals becomes infectious and can spread the COVID-19. γ is the recovery rate or mortality rate.

The above equations have N as the denominator. Therefore, it's a fraction of the total population N , the population of all the states sums up to the total population N .

Besides these parameters, many other parameters are set before launch. R_0 is another important parameter that sets the basic reproduction number. Specifically, R_0 is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection[4], which can be regarded as the capability of a person in specific region to infect with other individuals. Therefore, R_0 value is different according to different regions.

2.3. Limitation

In homework 3, we used Youyang Gu's SEIR model to simulate and predict confirmed cases, deaths and recoveries. After performing Youyang Gu's SEIR Covid-19 simulator, we noticed that the predicted number of confirmed cases is much greater than the observed value after the day calculated with the best parameters, although it can simulate the observed number before the date with best parameters. This fact infers that this model can be an acceptable simulator rather than an appropriate prediction model, since it will over-predict more and more values as time goes by. After conducting more experiments and studying on SEIR model, we concluded that this four-states model might be too simple to measure diseases in complex situations such as Covid-19. From studying SEIR model, we found out that the most significant part of this type of model is the ordinary differential equations that corresponds to each state in the model. The key task is to find the solutions of the ordinary differential equations with the observed data points so that the curves generated from the solutions will best fit the observed datasets, therefore stimulate the past and current situations in real-life and predict the tendency of all the states in the near future.

At the end, we figured SEIR model is not the best fit to measure COVID-19 confirmed cases, mainly because when the virus first broke out, the growth rate in the first few months was much higher than when after the disease got controlled, according to report by CDC. To fit this rapid increase of infected cases, the normal distribution simulated by SEIR will also be enlarged, so that the peak value of confirmed cases predicted by SEIR will be exaggerated and the time point where confirmed cases begin to decrease will start later, compare to that under regular conditions. Moreover, original SEIR model does not contain factors that represent ways or policies to alleviate the epidemic such as vaccinations, social distancing or quarantine. Since the epidemic spread so fast, it's urgent for government to take actions to alleviate this spreading disease. When the government takes action in real life, the number of confirmed cases will start to drop early, so on the normal distribution the peak value will then shift to the left.

Furthermore, there are a group of people who was first infectious to COVID-19 and eventually dead. Those people may also affect how the model simulate and predict.

In our project, we will add two new state called dead and vaccinated to the standard SEIR, which we then believe can predict the confirmed cases of COVID-19 more accurately. We also add natural birth and natural death to the model. With these new states and modification, we believe that the result predicted will not be too far from the real dataset, even if there is a rapid increase of cases during the beginning of the pandemic.

3. Related Work

3.1. Trial on YYG-SEIR Simulator

It is recommended that our work should extend from the YouyangGu-SEIR model[5] in our Homework 3, we therefore decided to initially work on the original YouyangGu-SEIR model and see what additional interesting area that we can dig into.

We tested the yyg-seir simulator with recent dates and compared the results with the actual cases of that date that Johns Hopkins University provided, and two set of data is largely different, meaning yyg-seir simulator being inaccurate. The simulator itself significantly over predicted the expect infection cases. For example, the model predicted that there will be over 200 millions infection cases across the United States by 03/01/2021 but the actual data indicated that the total infection cases was 28.6 millions. Therefore, large discrepancy appeared between the yyg-seir simulator results and the real data, and since the number had been very different from the real number, we decided to dive deeper into the source code and see if we could continue on top of the yyg-seir simulator and improve the accuracy of its prediction.

It is natural that our mind came to the individuals who are immune to the COVID-19 due to different reasons such as Vaccination or recovered from it. We therefore dive deep into the source code of Youyang Gu's model, and we noticed that a variable called immunity was used when running the simulation, which stands for the the immunity across a specific region, and the immunity multiplier for the US is 0.5, and he then use the following equation to calculate the immunity rate of the population:

$$immunity_perc = (1 - infected_perc)^{immunity_multiplier}$$

However, after our calculation, we found that the immunity percentage is very high and close to 1 even if the infected population is high (we fed 0.1 as infected population because it's the rate that the United States has currently), and the result showed that the immunity percentage is 0.94. Despite the fact that the immunity of the simulator is set to high, the prediction result still overpredicted a large amount. Therefore, we tried to decrease the immunity percentage by adding more weight on the power of the above equation, and the result still came out bad.

After previous trial, we supposed that the hyperparameters could be finer tuned in order to have a closer prediction. Therefore, we used Grid Search on some of the hyperparameters of the SEIR simulator. Specifically, Grid Search was applied to search for the best combination of the following hyperparameters: mortality_multiplier, immunity_multiplier and days_before_death. However, even the best combination of these hyperparameters will result in more than 150 millions cases, which indeed improved,

but still deviated far more than actual data. Therefore, Grid Search on the hyperparameters yielded unsatisfactory result as well.

3.2. Trial on SEIR+ framework

Meanwhile, we also started to dig into seirsplus model[6] that was recommended in project guideline. SEIR+ model initialized a Barabasi-Albert random graphs and use the SEIR to simulate the process of the infection. The simulation results based on the Barabasi-Albert random graphs successfully predicted the trend of the real COVID case. However, it overpredicted the data as well. Therefore, we implemented the SEIR+ model with different graph networks. The networks that we've tried are listed below:

- Watts-Strogatz Random Graphs
- Erdos Renyi Random Graphs

However, after trying different graph networks, we compared the simulation result curve and we found out the Barabasi-Albert random graph has the fittest simulation curve, which in turn marks that Barabasi-Albert random graph is the best fit for SEIR+ model, the reason for this is that Barabasi-Albert random graph is the most suitable network describing the social network. Despite the fact that social distancing is enforced during the pandemic, the most common way of spreading the COVID-19 is still through social contact. Therefore, social network can best describe the COVID-19 network as well and is therefore most accurate network graph to simulate the SEIR+ framework. But the result still overpredicted the actual cases.

3.3. New Thought

After spending a large amount of time on the literature reviews and different trails, we aimed our focus on modifying the differential equations of the original SEIR model, and we put importance on the effect that the Vaccination could bring in to the simulation of the COVID-19. Therefore, we propose our own simulation model for the COVID-19, the VSEIRD Model, more detailed information about VSEIRD model will be explained in section 4.

4. Our VSEIRD Model

4.1. Structure

For the VSEIRD model, it has six different compartments(states), which are Susceptible, Exposed, Infectious, Recovered, Dead and Vaccinated. Compared to standard SEIR model, our proposed VSEIRD model has two more states, which are Dead and Vaccinated. We also take natural birth and death into consideration, which will make this model closer to real-world situation.

- **Dead:** This stage represents the fraction of individuals who were infectious to COVID-19 and eventually died.
- **Vaccinated:** This stage represents the fraction of already vaccinated individuals, who will have a stronger immune system to fight against COVID-19.

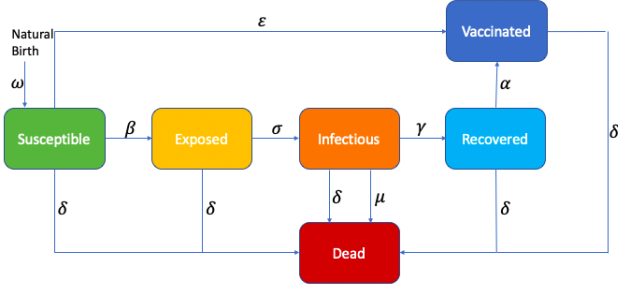


Figure 2: Flow Diagram of the VSEIRD model

Letter μ is the rate controlling the transfer from state Infectious to state Dead and letter α and ϵ are the rates controlling the transfer from state Recovered to state Vaccinated and from state Susceptible to state Vaccinated. Based on our knowledge, only those are not tested positive on COVID-19 or those are not exposed to an COVID-19 infectious individual can get vaccinated. Natural birth is controlled by letter ω and natural death is controlled by letter δ .

4.2. Equations and Parameters

The differential equations of the VSEIRD model are listed below:

$$\frac{dV}{dt} = \alpha R + \epsilon S - \delta V$$

$$\frac{dS}{dt} = -\frac{\beta SI}{N} - \epsilon S + \omega N - \delta S$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E - \delta E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I - \delta I$$

$$\frac{dR}{dt} = \gamma I - \alpha R - \delta R$$

$$\frac{dD}{dt} = \mu I + \delta(S + E + I + R + V)$$

$$N = S + E + I + R + D + V$$

5. Experiments and Results

According to latest news from CDC, COVID-19 seems to be alleviated in most countries around the world after people have been fully vaccinated these days. In US, as people started to take vaccines, the average daily increase of confirmed cases of the latest week is generally controlled to 20k, which is a massive progress compared to the situation few months ago. Nevertheless, the epidemic in some countries and regions is much worse than before, such as India. From recent news, in May, the peak average daily increase in last seven days increases to nearly 400k in India. This rapid spreading of COVID-19 leads us to focus on simulation and prediction using our model on dataset in India.

5.1. Dataset

The dataset we use in our project is from The Ministry of Health and Family Welfare in India, which can be found here (<https://api.rootnet.in/>). This dataset contains all the confirmed cases, deaths and recoveries in India as the time goes by. It currently saves data points in 458 days with 03/10/2020 as the start date of COVID-19.

5.2. Experiments

We tried four different models (SEIR, SEIRD, VSEIRD, VSEAIRD) and compare their final performance both on training and validation. We input data in the first 50 days to fit our model, which we call training process of the model. After the model finishing fitting data, we use it to do prediction on the next 400 days, which we call validation process of the model. All results of these three models both on training and validation are shown in subsection Results.

We set the initial population to 1380000000 which is the same as current population in India. We set initial number of exposed people to 1000 and initial number of infectious people to 47, which follows the COVID-19 data reported in India. For parameters used between states, we set α to 0.02, σ to 1/5.2, γ to 1/2.9, ϵ to 0.02, μ to 0.034, β to $4 * \gamma$, ω to 0.00174 and δ to 0.000000073, which follows the official estimation from government.

In our experiment, we also perform parameter estimation using least square algorithm. we use Minimize in lmfit to minimize the least square error of the observed data and our simulation result to find the parameters best fit the observed data and eventually build our model based on these best parameters.

With these self-built models, we finally can do prediction on the data in the near future and see how these models perform and what the limitations of these models.

5.3. Results

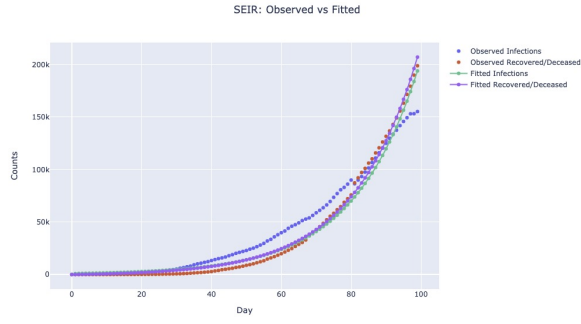


Figure 3: Training Result of the SEIR model

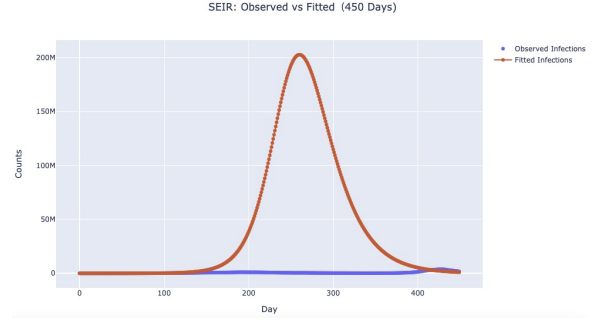


Figure 7: Validation Result of the SEIR model



Figure 4: Training Result of the SEIRD model

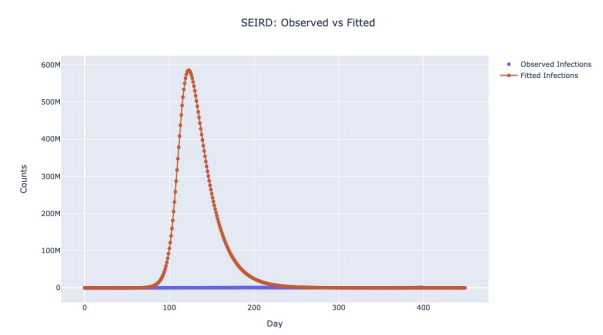


Figure 8: Validation Result of the SEIRD model

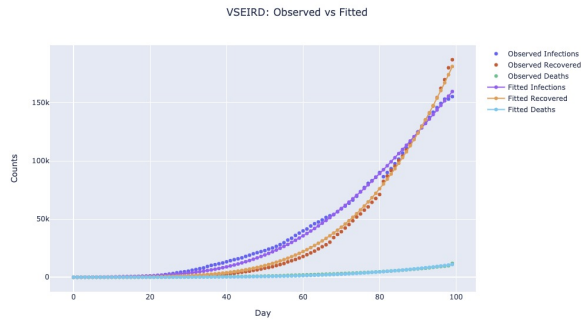


Figure 5: Training Result of the VSEIRD model

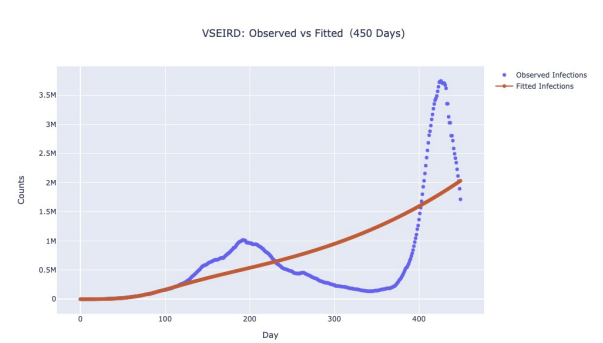


Figure 9: Validation Result of the VSEIRD model

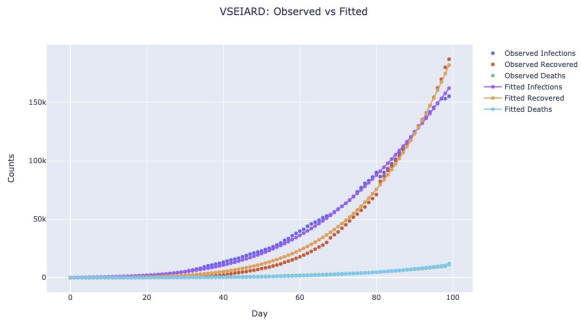


Figure 6: Training Result of the VSEIARD model

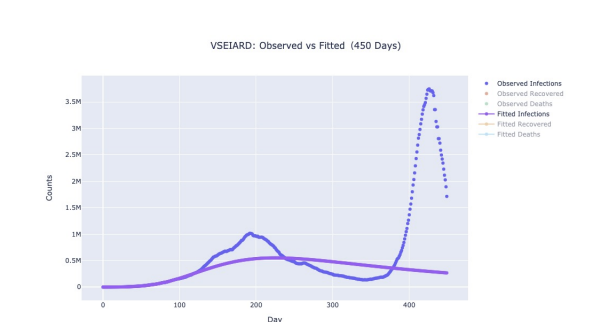


Figure 10: Validation Result of the VSEIARD model

5.4. Analysis

The results of three models are shown above. From the training curves of the three models, we can see that the fitted infections of both the SEIR and SEIRD models are not as accurate as the VSEIRD model, which is because the Vaccinated state is not taken into the account. As we can see from the VSEIRD model, all of the fitted curves have little discrepancy compared to the observed(true) curves. Therefore, VSEIRD model is the best one in training stage.

For the validation result, we can see that the original SEIR model over predicted a lot more, just like what we analyzed in `yyg-seir` simulator. Therefore, the original SEIR model is not accurate in terms of describing the COVID trend in India, so does the SEIRD model, even with the Dead state added, the model cannot predict well, and the SEIRD model performs even worse than SEIR model. Lastly, our validation result of the VSEIRD model makes much more sense as it is not over predicting the actual case, and the result is meaningful in predicting the actual case. However, since we use data of first 100 days to train the model, it will predict the result based on the scenario of the first 100 days case. Therefore, it is within expectation that our VSEIRD model did not catch the second outbreak of the COVID-19 (roughly from day 380 to day 420).

Last, we added another stage, the A(Asymptomatic). So the new model is VSEIARD. However, we don't see the result do better than the VSEIRD model. We believe it should contribute to the better performance of the model. Due time constraint, the parameters was not tuned to the optimal value. The new set of differential equation of VSEIARD is:

$$\frac{dV}{dt} = \alpha R + \epsilon S - \delta V$$

$$\frac{dS}{dt} = -\frac{\beta SI}{N} - \epsilon S + \omega N - \delta S$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E - \delta E - \rho E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \mu I - \delta I$$

$$\frac{dA}{dt} = \rho E - \theta A - \delta A$$

$$\frac{dR}{dt} = \gamma I + \theta A - \alpha R - \delta R$$

$$\frac{dD}{dt} = \mu I + \delta(S + E + I + A + R + V)$$

$$N = S + E + I + A + R + D + V$$

Note the curve is actually going down as time passes. The reason for this is that the model is trained with the first 100 days of data. and we see that the actual data has an initial bell shape curve, which causes the prediction to go down.

6. Conclusion

In conclusion, we believe that the VSEIRD model has the best performance over all other models because we are running out of time to fine tune the parameters in VSEIARD model. However, even the best model could not precisely draw the prediction curve that matches the actual curve.

The mathematical differential equation is rigorously designed and implemented, and the result did simulate COVID-19 case. However, it is after all an mathematical simulation, so it cannot take all the real world variations and parameters into the account. In the end, although the simulation mode successfully marks the curves of the real world data, it is vulnerable and subject to many factors. For example, the Variant COVID-19 had began to spread across the India, and the fatality rate of the Variant is much higher than the original COVID-19. Moreover, the health infrastructure has been crashed down by the COVID-19 temporally during early April, which is why the actual curve went up drastically during 380 days - 420 days. These real world scenarios cannot be properly foreseen by the mathematical model. Therefore, we believe it happens if the model itself cannot produce accurate curve.

Overall, we have improved the original SEIR model by adding some of the more deterministic states and factors, such as Vaccination and Asymptomatic, to make the simulator closer to the real world case. However, it is nearly impossible for a mathematical model to predict the real world case precisely. Hopefully one day more advanced technique could be applied in the field of foreseeing pandemics, and we also wish that India, as well as the rest of the countries, could overcome the COVID-19 pandemics sooner and everything shall be restored to normal.

References

- [1] Thirumalaisamy P Velavan and Christian G Meyer. The covid-19 epidemic. *Tropical medicine & international health*, 25(3): 278, 2020.
- [2] Michael Y Li and James S Muldowney. Global stability for the seir model in epidemiology. *Mathematical biosciences*, 125(2):155–164, 1995.
- [3] Wikipedia contributors. Compartmental models in epidemiology — Wikipedia, the free encyclopedia, 2021. URL https://en.wikipedia.org/w/index.php?title=Compartmental_models_in_epidemiology&oldid=1025589582.
- [4] Wikipedia contributors. Basic reproduction number — Wikipedia, the free encyclopedia, 2021. URL https://en.wikipedia.org/w/index.php?title=Basic_reproduction_number&oldid=1027130669.
- [5] Youyang Gu. `yyg-seir-simulator`. <https://github.com/youyanggu/yyg-seir-simulator>, 2020.
- [6] Ryan McGee. `seirs+ framework`. <https://github.com/ryansmcgee/seirsplus>, 2020.