COVID-19 Modelling and the Challenges of Modelling Complex Systems

Team AANK

Aadhithyan K, Apoorv Pandey, Ninad Hemant Huilgol, and Umate Kartik

E0-259 COVID-19 Project

28 November, 2021

1 Introduction

The COVID-19 pandemic has been a great challenge for healthcare systems and a major socio-political issue. Due to the widely acknowledged severity of the pandemic, large amounts of data has been gathered about various aspects of the pandemic. This has been an invaluable asset for researchers interested in modelling the pandemic to understand its behaviour and to make predictions about the future course of the pandemic. In this project, we attempt to model the COVID-19 pandemic in the state of Karnataka using the SEIR epidemic model. Using our model, we attempt to predict the course of the pandemic into the future to gain insights into the pandemic.

This project is an archetypal example of modelling a complex system composed of several interacting aspects. Detailed modelling of the pandemic requires the consideration of biological, socio-political, and economic systems. This involves collecting data from various sources, reconciling inconsistencies between them, and using the combined insights from these fields to understand the pandemic and its evolution through time.

2 Methods

The pandemic in the state of Karnataka was modelled using the SEIR model. From experiments, it is known that the mean incubation period for COVID-19 is 5.8 days and the mean recovery period is 5 days. The contact rate parameter β needs to be estimated at various time periods of the pandemic. In addition, the contact rate varies throughout the state, and so our model considers each district of Karnataka as a separate unit. Additional units were considered for various subdivisions of Bengaluru – being one of the largest cities in the country, it is appropriate to subdivide it into smaller units.

Immunity waning was also considered in the model using the Weibull model. The expected number of people who lose their immunity at any time t is given by

$$\Delta w(t) = \sum_{t'=0}^{\infty} \Delta_+ r(t - t') W(t - t'; \kappa, \tau)$$

where $\Delta_+ r(t)$ is the new recoveries at time t and $W(t; \kappa, \tau)$ is the Weibull distribution with shape factor κ and scale factor τ, given by

$$W(t; \kappa, \tau) = \frac{\kappa}{\tau} \left(\frac{t}{\tau} \right)^{\kappa - 1} \exp \left[- \left(\frac{t}{\tau} \right)^{\kappa} \right]$$

The gradient descent algorithm was used to optimize the effective contact rates of each unit. Two different loss functions were used, given by

$$l_{\text{single}}(\beta, \kappa, \tau; t_0) = \sum_{\text{districts}} \left[\log(i_{\text{pred}}(t_0)) - \log(i_{\text{actual}}(t_0))\right]^2$$

$$l_{\text{multiday}}(\beta, \kappa, \tau; t_1, t_2) = \sum_{\text{districts}} \sum_{t=t_1}^{t_2} \left[\log(i_{\text{pred}}(t)) - \log(i_{\text{actual}}(t))\right]^2$$

to optimize for the expected number of cases on a single day, and the expected number of cases over an interval of time respectively. Here the various i's represent the proportion of the population in the infectious state at any given time. In addition, the predicted infected and recovered fractions were compared with data gathered in the second seroprevalence survey conducted in Karnataka. Mobility was also considered in the stochastic phase of the pandemic.

Six phases of the pandemic were modelled as shown in Table 1

Phase	**Start date**	**End date**	**Loss function**
End of first wave	2020-10-11	2020-11-01	l_{single}
Interim period	2020-11-01	2021-02-28	l_{multiday}
Start of second wave	2021-03-01	2021-03-15	l_{single}
Stochastic phase	2021-03-15	2021-04-07	l_{multiday}
Second wave	2021-04-07	2021-04-22	l_{multiday}
Future	2021-04-22	2022-12-31	

Table 1: Phases of pandemic modelled

3 Results and Discussion

Unfortunately, we were unable to optimize our model to accurately predict the course of the pandemic. As shown in Figure 1, the predicted case counts do not match the observed case counts. Due to the large deviation observed, we do not consider it fruitful to attempt to extrapolate to the second wave and to the future using our model.

Nevertheless, a number of conclusions may be drawn from this exercise

1. The SEIR model fails to accurately describe the progression of the pandemic when case counts are low, such as the stochastic phases between the two waves. This is because the theoretical basis for the SEIR model requires modelling discrete events (infections) in a continuum manner, which necessitates high case counts.

2. The initial conditions for the simulator, derived from seroprevalence data, are quite dissimilar to the reported cases at the same time, even with CIR correction. This is because the CIR obtained from the seroprevalence study corresponds to the ratio of total cases reported throughout the pandemic to the total seroprevalence. Due to this, the CIR obtained is averaged over the entire course of the pandemic, and does not correspond to the time-local CIR required to correct for a shortfall in testing.

   We attempted to estimate the CIR as the ratio of reported active cases to the ratio of active infections from the seroprevalence survey. However, this leads to very large CIR values that lead to absurd conclusions elsewhere, such as CIR-corrected case counts in many BBMP zones being over twice the population of the zone itself. This indicates that the naive model we used, where the CIR is inversely proportional to the number of tests performed, is unsuitable.

3. The antibody waning parameters were guessed to be $\kappa = 3.67$ and $\tau = 120$ days. These parameters never changed significantly when training, indicating that the effect of antibody waning in all the phases we considered was quite small.

4. The error in our predictions were largest in the various zones of Bengaluru. This is likely due to the fact that the zones, being highly connected, show much greater mobility among themselves compared to any other districts in the state.

Below, we list some changes that we theorize would lead to better outcomes in future endeavours in pandemic modelling

- While the SEIR model is a simple and elegant model, it cannot be applied to all aspects and phases of a pandemic. Its utility is greatest during the waves of a pandemic, and it cannot accurately predict the interval between waves. In these phases, a probabilistic approach to modelling the dynamics may be a better alternative.

- Due to the large time separation between the waves of the pandemic, we may treat treat them as independent and model them separately. This also means that it will be difficult to estimate antibody waning purely by modelling; however we may resort to biological experiments to gain more accurate insights. Furthermore, this also indicates that using the SEIR model beyond the end of the second wave is ill-advised, since new waves of the pandemic likely arise through random processes as opposed to the deterministic nature of the SEIR model.

- The COVID-19 pandemic is unique among epidemics as it has become a major socio-political issue in society. In practical terms, this means that various external forces such as lockdowns and other travel restrictions alter the natural progression of the disease. Therefore, the contact rate parameter β is very difficult to estimate accurately, as external forces can cause very dramatic and rapid changes in β.

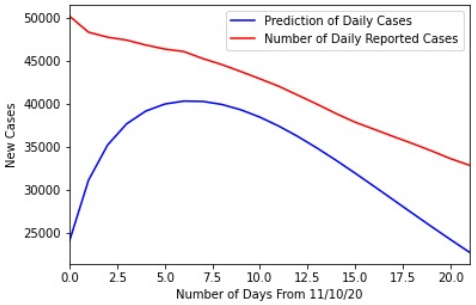
- Mobility is very important in epidemics, especially in a disease as contagious as COVID-19. Our naive mobility matrices treat all districts of Karnataka equally, which is perhaps an inappropriate assumption to make.

- Our naive CIR model is insufficient for modelling more than one wave of the pandemic. The CIR is influenced by several factors such as the strain on the health sector, the extent of testing, the efficacy of tests, political influence, and even the definitions of infection. Therefore, a more robust model is needed which considers, at the very least, the fact that tests are not performed uniformly randomly on the entire population.

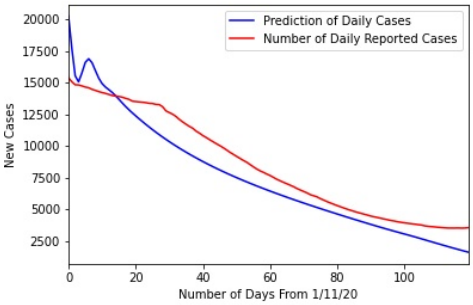
4 Conclusion

While we ultimately failed to produce a good model for the COVID-19 pandemic in all phases, we have gained insights into the strengths and weaknesses of the SEIR model. Our previous assignment on this topic illustrates the utility of the SEIR model during the peak of the pandemic's waves. This project illustrates the shortcomings when attempting to use the same model for the stochastic phases.

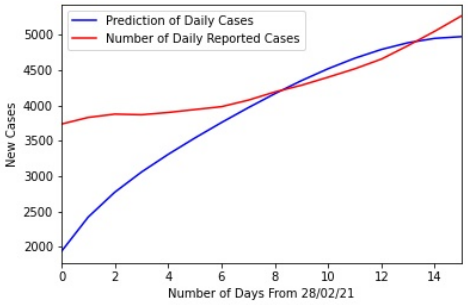
It is always disheartening to post a negative result on a problem we have worked hard on solving. However, especially for a subject with such importance as the COVID-19 pandemic, we must be very prudent in our use of analytical modelling methods so as to ensure the veracity and validity of the conclusions we draw from them. Ultimately, models are merely a tool for understanding nature, and we must accept that our models are fallible and limited in their applicability.



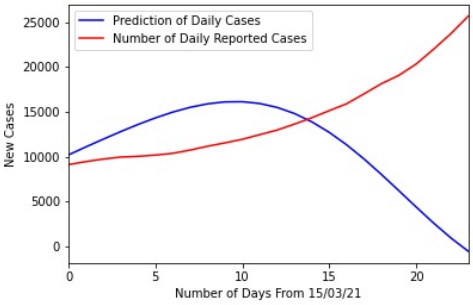
(a) Phase 1 predictions



(b) Phase 2 predictions



(c) Phase 3 predictions



(d) Phase 4 predictions

Figure 1: Optimized model predictions compared to actual case counts (CIR corrected)