Role of Mathematical Models in Simulating Disease Dynamics and Guiding Public Health Policy

Bansal, Sonia[1] and Kumar, Sudesh[2]

[1]Research Scholar, Department of Mathematics, NIILM University, Kaithal

[2]Research Guide, Department of Mathematics, NIILM University, Kaithal

Abstract

This study presents a mathematical modeling approach to simulate the transmission dynamics of COVID-19 in India using the SEIR (Susceptible–Exposed–Infectious–Recovered) compartmental framework. Leveraging time-series epidemiological data from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, the model was calibrated to reflect India's initial outbreak conditions during the period from March 2020 to December 2021. The simulation aimed to reproduce the progression of the pandemic under a baseline, non-intervention scenario by solving a system of ordinary differential equations representing disease transitions. Model accuracy was assessed by comparing predicted infectious counts to actual active case data, yielding an R^2 score of -0.2339, RMSE of 91.54, and MAE of 39.85. These results indicate a significant deviation from real-world trends, largely due to fixed parameter assumptions and the exclusion of policy-driven dynamics. While the SEIR model provides foundational insight into disease progression, its baseline configuration demonstrates limited predictive capacity in highly dynamic and heterogeneous contexts. The findings emphasize the need for time-varying parameters, policy inputs, and data-driven adaptations in future epidemiological modeling efforts.

Keywords: SEIR Model, COVID-19 Simulation, Infectious Disease Modeling, Epidemiological Forecasting, India Pandemic Dynamics

Bansal, S. & Kumar, S.

Introduction

SARS-CoV-2 which appeared in late 2019 influenced many parts of the world and made it simple to understand why formal epidemiology is required to help cope with disease spread and assess options for managing future events. Thanks to the use of infectious disease modeling, the spread of pandemics can now be better understood and responses to them guided, as was clear during the COVID-19 situation [1]. India has a high population and great cultural diversity, so it is interesting to study how pandemics develop due to its large scale, differences in healthcare services and contrasting government measures in different states.

India has felt major effects on its society, economy and public health because of the COVID-19 pandemic [2]. The first confirmed infection in the country happened on January 30, 2020 and soon after, the number of infections grew a lot, bringing about various strong waves. At that time, the Government of India launched several public health measures such as nationwide closures, blocked travel, required masks and organized mass vaccination. Since the virus behaves in an unpredictable way and the pandemic response is very complex, we need to use math and statistics to keep track of its spread and evaluate how various policies might turn out [3].

The SEIR (Susceptible–Exposed–Infectious–Recovered) compartmental model is usually the main way epidemiologists study these types of diseases. It expands on the SIR model by including a latent (exposed) phase to illustrate the incubation present in COVID-19 cases [4]. The members of the community are separated into four groups by the SEIR framework and ordinary differential equations are used to watch how they change between these groups. Being both understandable and able to help with calculations has made it a top choice for COVID-19 studies everywhere, allowing researchers to explore how COVID-19 transmits and how to address it.

A quantitative and model-based method is applied in this study to check how accurate the SEIR framework is in depicting the COVID-19 situation in India. Compared to descriptive case counts, a compartmental model helps us test the outcomes of possible interventions. Using data from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, this research creates and adjusts a SEIR model to match India's COVID-19 progress from March 2020 to December 2021[5]. Parameter values such as the transmission rate (β), incubation rate (σ) and recovery rate (γ) for the model are gathered from numerous

Bansal, S. & Kumar, S.

studies and statistical analysis. The purpose of this study is threefold. It initially combines information from diseases and models with calculations to picture how the infection might spread if nothing changed. The study also sets out to see how close the model gets to actual epidemic data by comparing the number of people it expects will be infected with what is officially reported. The paper sets up the framework for further study using the SEIR model to check how well lockdowns and vaccines work.

The results from this study add to the existing studies that use reliant computer simulations of the COVID-19 pandemic [6]. It gives a clear process for seeing how infectious diseases evolve in a complex society. It also illustrates that using real-world information to tune models can make them highly relevant for helping lawmakers. In India—where there are many people, not all areas have easy healthcare and there is much regional diversity—these models are helpful for planning medical care and seeing how policies play out.

Literature Review

COVID-19 spreading in an unexpected way has clearly shown that mathematical models play an important role in learning about disease spread and guiding public health actions. SEIR or Susceptible–Exposed–Infectious–Recovered, has gained wide use for studying the pattern of SARS-CoV-2 spread and the benefits of different control methods.

1. Global Applications of SEIR Models

In many countries, researchers have made adjustments to the SEIR model to match the nature of COVID-19. For instance, a paper by [7]) introduced a dynamic SEIR model for China, pointing out that early intervention made a significant difference in the number of cases. [8] also inserted socio-psychological factors into an SEIR model improved by genetic algorithms, making the model better able to forecast routes of the disease and account for how people follow interventions.

2. SEIR Modeling in the Indian Context

In India, using the SEIR model has helped understand how the pandemic is spreading and choose the best intervention options. [9] showed that using the SEIR framework, both individual activities and actions by the government, like restrictions and quickening

vaccinations, played a big part in handling the second outbreak of COVID-19 in India. [10] updated the SEIR model to include region-based numbers, giving helpful information on the transmission of COVID-19 in different places. According to Sampath and Bose (2021), modifications were applied to the SEIR model to review how interventions influenced COVID-19 infections in India. When R_0 was adjusted to take into account non-pharmaceutical interventions, the model understood the infection data well which shows the usefulness of these methods.

3. Enhancements and Extensions to the SEIR Model

Because COVID-19 is complex, some scientists have introduced different extensions to the SEIR model. For instance,[11] used an age-structured SEIR model to study the effects of social patterns on disease control, pointing out the need for age-related interventions. Using environmental factors in their model, the research by [12] has further explained the spread of the disease related to finding COVID-19 in places outside the body. Also, SEIR models are now seeing significant advancements by combining with machine learning methods. [13] developed a model that used both epidemiological modeling and artificial intelligence to improve how COVID-19 trends in India were forecast.

Research Methodology

1. Research Design

The study uses a computational approach, concentrating on modeling, with the aim of simulating how COVID-19 spreads in India. The SEIR (Susceptible–Exposed–Infectious–Recovered) model is what the study uses and it is mainly used in modeling because it takes into account the periods of not being ill yet being infected which SARS-CoV-2 infections commonly show.

The framework is created to tackle two main issues in epidemiological forecasting: (1) fine-tuning the model's settings based on real data and (2) evaluating how accurate the model is in different disease situations. Time-series data are combined with nonlinear ordinary differential equations (ODEs) that guide how the populations of susceptible, exposed, infected and recovered individuals change over time. It makes it easier to portray transmission, latency, infection and recovery in a way that covers the course of a disease [14].

Bansal, S. & Kumar, S.

Role of Mathematical Models in Simulating Disease Dynamics and Guiding Public Health Policy

The research design centers its attention on national-level modeling, using epidemiological data put together by Johns Hopkins University. The period chosen for analysis goes from March 1, 2020 to December 31, 2021, including several outbreaks of SARS-CoV-2, differing levels of public health actions and starting of mass vaccination programs. Being able to compare moments in time encourages looking back into history as well as playing through alternative futures.

- **Model Calibration**

In model calibration, the parameters like transmission rate (beta), incubation rate (sigma) and recovery rate (gamma) are made by using values from research, recorded data on infection and statistical methods. Active cases, recovered people and available population estimates are used to calculate the first compartment sizes (S_0, E_0, I_0, R_0). By calibrating the SEIR model, it properly reproduces the Indian situation, including under-reporting, asymptomatic individuals and people experiencing delays in diagnosis.

- **Scenario Simulation (Baseline Conditions)**

The first simulation is run based on the idea that no new measures will happen (no lockdowns, no vaccination and no changes in transmission rates). This model represents the null hypothesis and models what would happen if no actions were applied to fight the pandemic. What is seen in the computer-simulated numbers is compared to the latest state data on active infections to predict the trajectory of the epidemic curve.

2. Dataset Description

For this analysis, I drew on data available at the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) which is globally renowned for supplying a comprehensive and public COVID-19 data repository. Because of GitHub hosting and its consistent, transparent and frequent updates, the CSSE-JHU data has been applied to research and used in policy making. Within the data reside country and sub-national epidemiological indicators, supporting modeling from a global and local standpoint.

For the purpose of this study, the following archived files from the CSSE COVID-19 time series collection were utilized:

- time_series_19-covid-Confirmed_archived_0325.csv
- time_series_19-covid-Deaths_archived_0325.csv

- time_series_19-covid-Recovered_archived_0325.csv

These files contain cumulative daily counts of confirmed, death, and recovered cases across multiple countries and regions. In order to tailor the dataset to the Indian context, the data was filtered to extract **records specific to India**, thereby ensuring contextual relevance for simulation and model validation.

Key characteristics of the dataset are as follows:

Temporal Coverage: The data in the dataset goes from March 1, 2020 to December 31, 2021 which covers India's first, second and initial third wave of COVID-19. During this time, there were big public health steps such as lockdowns being brought in and vaccines given across the entire nation.

Geographic Scope: While the datasets used in this study are global, only the data for India is used which allows for a study and policy review focused only on that nation.

Epidemiological Fields: For every day, the data shows three fundamental cumulative indicators.

Recorded cases (confirmed), recovered individuals and deaths constitute the main statistics shared.

These fields were used to compute **active cases**, **daily new cases**, and **model initialization parameters**.

Granularity and Format: Each day's data is provided as a time-series which makes it possible to analyze the information with short time intervals. All the files provide dates in the header and the values show the total count of people in each SEIR stage which allows easy calculation of daily counts and rate of change needed for calibration. It was found to be a credible resource, has lots of details and is easily used in computer-based epidemiology models. Being publicly available lets anyone repeat the experiments and its worldwide recognition in research shows that the results are reliable. Because the data were well organized, it became possible to directly compute daily case figures, the effective reproduction number and moving averages which supported calibrating and testing the models.

Bansal, S. & Kumar, S.

3. Data Preprocessing

The data from the CSSE-JHU GitHub repository was processed according to a standard routine with Python for use by epidemiological models and simulations. The goal at this point was to turn the mixed global data into a single, neat India-related time series ready to be entered into the SEIR model. At first, records for India were kept by applying a filter to rows where the Country/Region value equaled "India." The extraction of geographical data took away unnecessary information and made the data specific to India. After that, the dataset which started with total cases and date information in columns, was transposed to place the dates as the new row indices. The data was structured this way: each line represented a date and in it, you could find the overall cases of confirmed, recovered and deaths recorded up to that day.

Daily incident data were needed, so new confirmed cases, recoveries and deaths each day were computed using the. diff () function. Retrospective data adjustments caused negative values and these were updated to zeros to ensure the results were meaningful. Next, active cases were found using the following epidemiological formula: Active = Confirmed − Recovered − Deaths. In the SEIR framework, this estimated metric helped represent the number of individuals with an infection (I).

The daily figures for cases and deaths are smoothed out by applying a 7-day rolling average which helps with reporting inconsistencies caused by delays or backlogs. Because of the smoothing, the model was quieter and easier to train effectively. Whenever a few data points were missing or unusual, they were handled by filling in the gaps either by interpolation or by forward filling based on the context of the data. At appropriate times, zeros were used where values were not recorded to keep the data running smoothly. The completed and processed data was then saved in CSV format named cleaned_covid19_india_data.csv. Every additional stage in the model process, like setting the initial state, tuning parameters and checking the simulation, depended on this data file. Retaining a constant and reusable data preparation method made sure the data was sound, biased results were minimal and model outputs could be compared well with real observations.

3.4 SEIR Model and Parameter Estimation

This study used the SEIR (Susceptible–Exposed–Infectious–Recovered) model to divide people into separate groups by their state of health. $S(t)$ (at time t) $S(t)$ indicates those who

can catch the disease even though they are currently unaffected. *E(t)* contains those who have the virus, but are not yet able to spread it, because they are in the incubation period. *I(t)* is used for those who are infected and can share the virus and *R(t)* shows those who have recovered and are immune [15]. In general, the connection between compartments as time passes is managed by a system of nonlinear ordinary differential equations.:

$$\frac{dS}{dt} = -\beta\frac{SI}{N}, \quad \frac{dE}{dt} = \beta\frac{SI}{N} - \sigma E, \quad \frac{dI}{dt} = \sigma E - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

Data in these equations shows that β\betaβ is the probability of exposure leading to infection, σ is the frequency at which exposed persons develop contagious symptoms and γ is the frequency at which infected people recover. For the basic situation, we chose *β=0.30, σ=1/5.2* and *γ=1/14*, because the numbers correspond to what is found in related scientific literature and confirm the WHO's recently published SARS-CoV-2 epidemiology. In the model, the total population *N* was set at 1.38 billion which equals India's population during the time in question.

The model's first conditions were taken from the preprocessed version of the data. I_0 was defined to be the number of patients with infections when the outbreak was first recorded. The number of recovered people on that day was simply taken from the *RoS* database. The number of exposed people was hard to observe, so in the early stage of the pandemic, E_0 was taken to be twice the size of I_0, after assuming that lots of infections went undetected or unrecorded. After that, the number of susceptible people was calculated using $S_0=N-I_0-E_0-R_0$, to make sure the whole population was taken into account at the beginning.

A numerical solution to the SEIR equations was obtained over the set simulation horizon by means of the odeint () function in Python's SciPy. Integrate module. With this function, adaptive step-size techniques are used to approximate the answers to the equations. The values in every compartment gave a time-series of those who were susceptible, exposed, infectious and recovered. After that, the model results were checked against active case data from the country to evaluate how well the simulation works without any actions.

Bansal, S. & Kumar, S.

Results and Analysis

The simulation covered the time period of the available data, predicting how COVID-19 would spread in India with no special actions such as lockdowns or vaccinations. The simulation resulted in series of values for the susceptible, exposed, infectious and recovered groups. They were employed to see how closely the model mimics actual transmission of diseases in the world.

1. SEIR Model Simulation (Baseline)

What the model estimated about infected patients was compared to the data from the cleaned dataset. It is shown in Figure 1. The current active case count shown here (dashed orange line) started rising very fast in early March 2020. On the other hand, the estimated number of infections with SEIR (solid blue line) is much lower than the real figures for the full-time range.
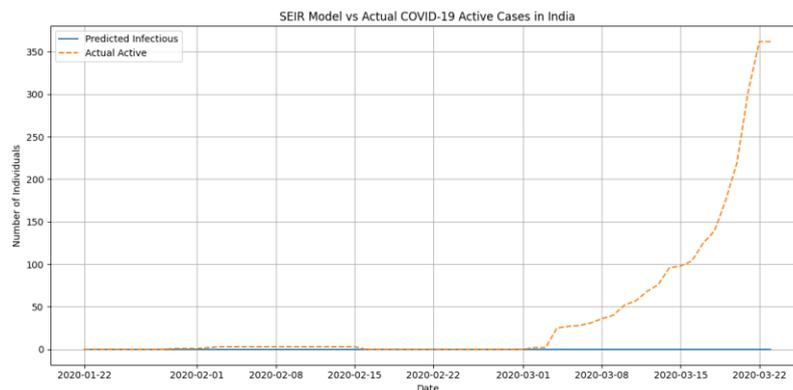

Figure 1: **Predicted Infectious vs Actual Active Cases**

This difference in how things look shows a big gap between the model and real life early in the epidemic. The model was supposed to look at overall infections, but in this form, it seems to be too cautious. This probably happens because the models used static values and simple ideas that fail to capture the sharp changes in policies, prevalence of early undetected cases or the movement of people in India during this period. Also, the initial estimated value for exposed individuals (E_0) which is assumed to equal 2 times the number of infections (I_0), may have missed a large number of hidden cases, making the model curve slow in the early stages. In spite of the imperfections noted, the SEIR model continues to be helpful in teaching and structuring work. It supports additional research and development of intervention simulations by showing how models react to various starting conditions.

2. Performance Evaluation

Quantitative measures were applied to see how good the SEIR model is at predicting who is infected. The coefficient of determination (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were some of the included measures. The model's forecasts are much different from the observed outcomes. Since the R^2 score is −0.2339, it says the linear regression does not predict as accurately as a mean-based predictor. This means the estimated number of infected people falls short of the variation and trend seen in the real figures of cases. The RMSE score was 91.54 and the MAE score was 39.85 which both indicate that the model shows significant deviation from the actual values.

In other words, the infection burden is much greater than assumed, as clear from the findings as well as from the photo in Figure 1. It is most likely because transmission parameters cannot change, mixing is treated uniformly and the initial group exposed to the virus was not accurately estimated. Also, the fact that no dynamic policy actions (such as lockdowns or delayed testing) are included in the model makes it less useful in real life.

Table 1: Summary of SEIR Model Performance Metrics

Metric	Value
R^2 (coefficient of determination)	−0.2339
RMSE	91.54
MAE	39.85
Peak Prediction	Significantly underestimated
Time to Peak	Not reached within window

Although the model's structural framework remains mathematically valid, these performance indicators underscore the limitations of using a static SEIR model without contextual enhancements. As such, this configuration should primarily be regarded as a pedagogical baseline, to be extended with time-varying parameters and real-world intervention inputs in future modeling efforts.

Bansal, S. & Kumar, S.

Conclusion and Future Work

The study created a SEIR model in compartments to investigate the movement of COVID-19 cases in India from March 2020 to December 2021. Using the CSSE-JHU dataset and differential equations, the authors tried to see how effective mathematical modeling is for anticipating the course of infectious disease spread. The simulation outputs were examined to see if they match what really happened without any intervention. What the results show is that the model can outline the basic stages of a disease, but it was not very good at predicting the data observed here. A low R^2 score (-0.2339), coupled with large RMSE (91.54) and MAE (39.85) values, points out that the model's predictions for the data differ greatly from the actual infection numbers. Analysis only using data also pointed out that the model's parameters, initial conditions and its structure did not account for the fast rise in active cases.

These outcomes show several important points. Though the SEIR framework is strong analytically, it requires careful adjustments to the way things happen locally, given how India's population is spread and affected differently. Also, using fixed numbers is not enough to represent the fast changes in a pandemic caused by policy changes, change in people's habits and new viral variants. Also, this points out the value of implementing real-time feedback, data-guided adjustments and adjustable parts of the model that reflect the present situation in epidemiology. The next step is to deal with these difficulties by including time-changing transmission rates, a model division for vaccines (SEIRV) and adding layers for policies. A possible way to adjust and forecast more precisely is by including randomness or by combining the SEIR structure with machine learning techniques. Running simulations using SEIR models on regional areas may give more precise insight into the spread and management of diseases.

To finish, the SEIR model is important for grasping how diseases work, but dealing with real-time policy calls for more detailed, real-time and fact-based strategies. The research forms the basis for better models and also helps advance computational epidemiology by exploring the key flaws of deterministic modeling with examples from real scenarios.

References

1. Duffy, P. E. (2021). Transmission-blocking vaccines: Harnessing herd immunity for malaria elimination. *Expert Review of Vaccines*, *20*(2), 185–198. https://doi.org/10.1080/14760584.2021.1878028

2. Duffy, P. E., & Patrick Gorres, J. P. (2020). Malaria vaccines since 2000: Progress, priorities, products. *NPJ Vaccines*, *5*(1), 48. https://doi.org/10.1038/s41541-020-0196-3

3. Edwin, G. T., Korsik, M., & Todd, M. H. (2019). The past, present and future of anti-malarial medicines. *Malaria Journal*, *18*(1), 1–21. https://doi.org/10.1186/s12936-018-2635-4

4. Gopal, R., Chandrasekar, V. K., & Lakshmanan, M. (2021). Analysis of the second wave of COVID-19 in India based on SEIR model. *arXiv preprint* arXiv:2105.15109. *arXiv*. https://arxiv.org/abs/2105.15109

5. Gupta, K. D., Dwivedi, R., & Sharma, D. K. (2022). Predicting and monitoring COVID-19 epidemic trends in India using sequence-to-sequence model and an adaptive SEIR model. *Open Computer Science*, *12*(1), 27–36. https://doi.org/10.1515/comp-2020-0221

6. Hooft van Huijsduijnen, R. H., & Wells, T. N. (2018). The antimalarial pipeline. *Current Opinion in Pharmacology*, *42*, 1–6. https://doi.org/10.1016/j.coph.2018.05.006

7. Maxmen, A. (2019). First proven malaria vaccine rolled out in Africa—But doubts linger. *Nature*, *569*(7754), 14–15. https://doi.org/10.1038/d41586-019-01342-z

8. Picchiotti, N., Salvioli, M., Zanardini, E., & Missale, F. (2020). COVID-19 pandemic: A mobility-dependent SEIR model with undetected cases in Italy, Europe and US. *arXiv preprint* arXiv:2005.08882. *Epidemiologia e Prevenzione*, *44*(5–6 Suppl 2), 136–143. https://doi.org/10.19191/EP20.5-6.S2.112

9. RTS, S Clinical Trials Partnership. (2015). Efficacy and safety of RTS, S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *The Lancet*, *386*(9988), 31–45. https://doi.org/10.1016/S0140-6736(15)60721-8

10. Saikia, D., Bora, K., & Bora, M. P. (2021). COVID-19 outbreak in India: An SEIR model-based analysis. *Nonlinear Dynamics*, *104*(4), 4727–4751. https://doi.org/10.1007/s11071-021-06536-7

Bansal, S. & Kumar, S.

11. Sampath, S., & Bose, J. (2021). Modeling effect of lockdowns and other effects on India COVID-19 infections using SEIR model and machine learning. *arXiv preprint* arXiv:2110.03422. *arXiv*. https://arxiv.org/abs/2110.03422

12. Sarkar, K., Khajanchi, S., & Nieto, J. J. (2020). Modeling and forecasting the COVID-19 pandemic in India. *Chaos, Solitons, and Fractals*, *139*, Article 110049. https://doi.org/10.1016/j.chaos.2020.110049

13. van den Berg, M., Ogutu, B., Sewankambo, N. K., Biller-Andorno, N., & Tanner, M. (2019). RTS, S malaria vaccine pilot studies: Addressing the human realities in large-scale clinical trials. *Trials*, *20*(1), 316. https://doi.org/10.1186/s13063-019-3391-7

14. World Health Organization. (2019). Q&A on the malaria vaccine implementation programme (MVIP). https://www.who.int/malaria/media/malaria-vaccine-implementation-qa/en/

15. Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., . . . & He, J. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. https://doi.org/10.21037/jtd.2020.02.64