Infectious Disease Models

A Project (Phase I) Report submitted in partial fulfillment of the requirements for the award of the degree of

MASTER OF SCIENCE

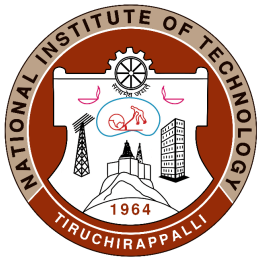
in

MATHEMATICS

By

Himakshi

(216124003)



DEPARTMENT OF MATHEMATICS

NATIONAL INSTITUTE OF TECHNOLOGY

TIRUCHIRAPPALLI – 620 015

DECEMBER 2025

ABSTRACT

This project presents a comprehensive study of COVID-19 dynamics by integrating real-world data analysis with SEIR-based mathematical modeling and parameter estimation. COVID-19 case and vaccination datasets were collected, carefully cleaned, and analyzed to understand temporal trends, identify highly affected regions, and examine variations in transmission and recovery patterns. A deterministic SEIR model with biologically meaningful parameters was formulated to describe the progression of individuals among susceptible, exposed, infectious, and recovered compartments. Using nonlinear least-squares techniques, key epidemiological parameters were estimated from reported data, and numerical simulations were performed to compare model predictions with observed cumulative case trajectories across different regions. During the analysis, discrepancies arising from data revisions and preprocessing differences were identified and addressed to improve consistency prior to model fitting, leading to parameter estimates that differ from some previously reported results. The basic reproduction number (R_0) was computed to assess transmission potential and interpret the impact of control measures. The results demonstrate that the calibrated models successfully capture the overall dynamics of COVID-19 and provide meaningful insights into transmission intensity, infectious period, and epidemic control. As a direction for future work, the proposed modeling and estimation framework can be extended and applied to other epidemic diseases to support comparative analysis and effective public-health planning.

BONAFIDE CERTIFICATE

This is to certify that the Project (Phase I) titled **Infectious Disease Models** is a bonafide record of the work done by **Himakshi (216124003)** in partial fulfillment of the requirements for the award of the degree of **Master of Science** in **Mathematics** of the **NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI**, during the year 2025-2026.

Supervisor **Head of the Department**

Project (Phase I) Viva-Voce held on: _____

Chairperon
Project Evaluation Committee

External Examiner

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project supervisor, Prof. V. Kumaran, for his constant guidance, encouragement, and valuable suggestions throughout the work. His insights on mathematical modeling and data analysis greatly helped me in shaping and improving this project.

I am deeply thankful to the Department of Mathematics, National Institute of Technology Tiruchirappalli (NIT Trichy), for providing academic support, necessary resources, and a motivating learning environment during the course of this work. I also thank the faculty members and staff of the department for their continuous help and cooperation.

My heartfelt thanks go to my friends and classmates for their encouragement and constructive discussions. Finally, I gratefully acknowledge my family for their unconditional support, patience, and inspiration, which motivated me to complete this work successfully.

Date: December 30, 2025 **Himakshi**

List of Figures

2.1 Simulation of the Bhilwara SEIR–type model (Mishra et al., 2021). . . . 7
2.2 Effect of increased transmission and mortality on epidemic dynamics. . . 7
2.3 SEIR simulation for Nigeria using estimated model parameters. 10

3.1 Model fit to Nigeria training data (first 62 days). 15
3.2 Testing the fitted Nigeria SEIR model beyond calibration period. 16
3.3 SEIR fit for India cumulative case data. 17

List of Tables

3.1 Estimated SEIR parameters for Nigeria. 15
3.2 Estimated SEIR parameters for India. 17

Contents

Title Page i

Abstract i

Bonafide Certificate ii

Acknowledgement iii

List of Figures iv

List of Tables iv

1 Introduction **2**

1.1 Literature Review 2

1.2 Problem Description 3

1.3 Objectives 3

2 Simulations **5**

2.1 Model Study and Simulations 5

2.1.1 Bhilwara SEIR-type model (Mishra et al., 2021) 5

2.1.2 Model from Aloke et al. (2023) 8

3 Parameter Estimation **11**

3.1 Data and preprocessing 11

3.1.1 Nigeria data 11

3.1.2 India data 12

3.2 Parameter estimation methodology 12

3.2.1 Objective function: Least–Squares Formulation 12

3.2.2 Initial guesses and constraints 13

3.3 Results — Nigeria 14
3.3.1 Estimated parameters 14
3.3.2 Fit figure (training) 15
3.3.3 Extended testing 16
3.4 Results — India 16
3.4.1 Estimated parameters 16
3.4.2 India fit figure 17
3.4.3 R_0 computation and interpretation 18

4 Conclusion and Future Work **19**
4.1 Conclusion 19
4.2 Future Work 20

Bibliography **21**

Chapter 1

Introduction

The COVID-19 pandemic has had a profound impact on public health, the economy, education, and day-to-day life across the world. Understanding how an infectious disease spreads, how fast it grows, and how it can be controlled is essential for planning effective public-health strategies. Mathematical modeling and data analysis provide a systematic way to study epidemic behaviour, and they have played a significant role in interpreting and predicting COVID-19 dynamics.

In this work, real COVID-19 data are combined with SEIR-based epidemiological models in order to obtain meaningful insights into transmission, recovery, and control measures. The study integrates dataset analysis, theoretical modeling, parameter estimation, and simulation to develop a comprehensive framework for studying infectious disease spread.

1.1 Literature Review

Compartmental models such as SIR and SEIR have been widely used in mathematical epidemiology to describe the flow of individuals through disease states. The classical SEIR model introduces an exposed class to represent the incubation period, making it particularly suitable for diseases such as COVID-19. Many researchers have extended SEIR-type models to incorporate quarantine, vaccination, isolation, and loss of

immunity, making them more realistic and adaptable to real situations.

Alongside theoretical development, recent studies emphasise parameter estimation using real datasets so that models reflect actual transmission behaviour. Least-squares fitting, Bayesian inference, and numerical simulations are commonly used to calibrate epidemiological models and to compute key indicators such as the basic reproduction number.

1.2 Problem Description

Although case counts and vaccination data provide useful information, they do not directly reveal how strongly a disease spreads or how effective interventions are. Important epidemiological quantities such as transmission rate, incubation period, infectious rate, and reproduction number must be inferred using mathematical techniques.

Therefore, the main problem addressed in this work is:

> To model the spread of COVID-19 using an SEIR framework, estimate its parameters from real data, and interpret the resulting dynamics using the basic reproduction number.

1.3 Objectives

The specific objectives of this study are:

1. To collect and analyze real COVID-19 datasets.
2. To review SEIR-based mathematical models.
3. To estimate biologically meaningful parameters using nonlinear least-squares fitting.
4. To simulate model behaviour and compare it with observed data.
5. To compute and interpret the basic reproduction number R_0.

6. To discuss how the modeling framework can support analysis of other epidemic diseases in future work.

This integrated approach connects theoretical mathematics with real-world epidemiological data and aims to provide a deeper understanding of infectious-disease dynamics.

Chapter 2

Simulations

2.1 Model Study and Simulations

Mathematical simulations allow us to observe how an epidemic evolves under different transmission and control conditions without conducting real-world experiments. Once a model has been formulated, parameter values can be varied to study their influence on the number of susceptible, exposed, infectious, quarantined, and recovered individuals.

In this chapter, simulations are used to gain intuition about the behaviour of both the extended SEIR model and its variants. The goal is not only to reproduce observed data, but also to understand how changes in parameters alter the epidemic trajectory.

2.1.1 Bhilwara SEIR-type model (Mishra et al., 2021)

Mishra *et al.* [2] extend the classical SEIR model to include suspected home-quarantine and medical-quarantine compartments. Using the notation from the paper, the compartments are:

$S(t)$ (susceptible), $E(t)$ (exposed), $Q_1(t)$ (suspected / home quarantine), $I(t)$ (infectious),

$Q_2(t)$ (medical quarantine), $R(t)$ (recovered).

The model ordinary differential equations (system (1) in the paper) are:

$$\frac{dS}{dt} = A - \beta SI - \mu_1 S + \gamma_2 Q_1, \tag{2.1}$$

$$\frac{dE}{dt} = \beta SI - (\alpha_1 + \alpha_2 + \mu_1)\, E, \tag{2.2}$$

$$\frac{dQ_1}{dt} = \alpha_1 E - (\gamma_1 + \gamma_2 + \mu_1)\, Q_1, \tag{2.3}$$

$$\frac{dI}{dt} = \alpha_2 E + \gamma_1 Q_1 - (\delta_1 + \delta_2 + \mu_1 + \mu_2)\, I, \tag{2.4}$$

$$\frac{dQ_2}{dt} = \delta_1 I - (\mu_1 + \mu_3 + \varepsilon)\, Q_2, \tag{2.5}$$

$$\frac{dR}{dt} = \varepsilon Q_2 + \delta_2 I - \mu_1 R. \tag{2.6}$$

Here the parameters mean:

- A: influx (birth) rate,
- μ_1: natural death rate,
- μ_2, μ_3: disease death rates in infectious and medical-quarantine classes,
- α_1, α_2: rates moving from exposed to suspected/home-quarantine and to infectious,
- γ_1, γ_2: rates transferring Q_1 to I or back to S (after negative test),
- δ_1, δ_2: rates from I to medical quarantine and to recovered,
- ε: recovery rate from medical quarantine,
- β: effective transmission/contact rate (per susceptible–infectious contact).

Simulation with Different parameters Values

Simulation parameters:

$$A = 0.02, \quad \mu_1 = 0.01, \quad \mu_2 = 0.018, \quad \mu_3 = 0.015, \quad \beta = 0.1, \quad \alpha_1 = 0.3, \quad \alpha_2 = 0.05,$$

$$\gamma_1 = 0.3, \quad \gamma_2 = 0.02, \quad \delta_1 = 0.07, \quad \delta_2 = 0.1, \quad \varepsilon = 0.02.$$

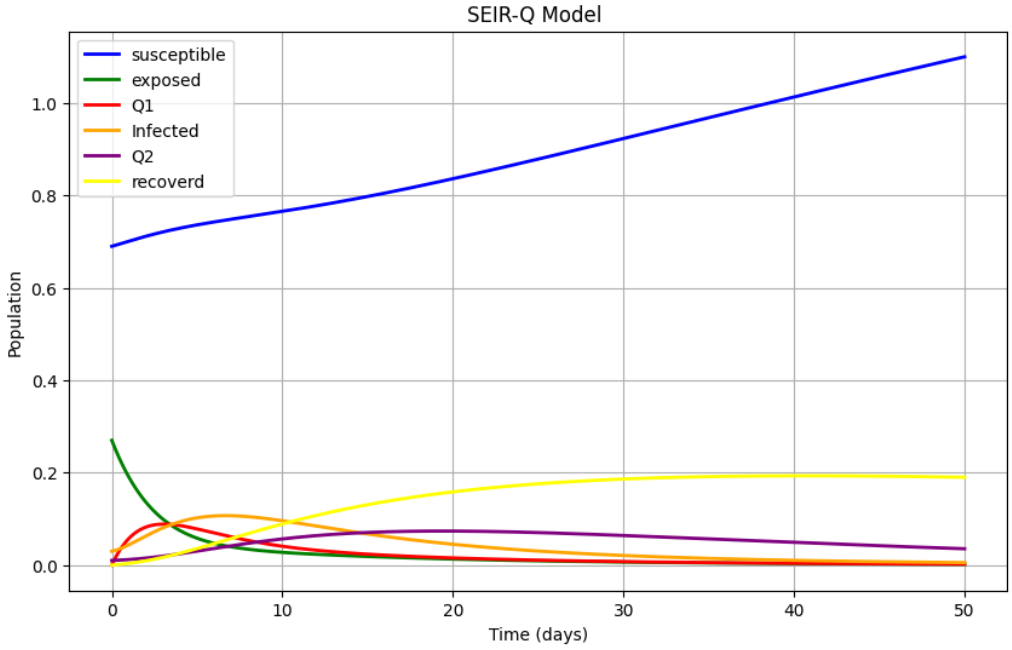


Figure 2.1: Simulation of the Bhilwara SEIR–type model (Mishra et al., 2021).

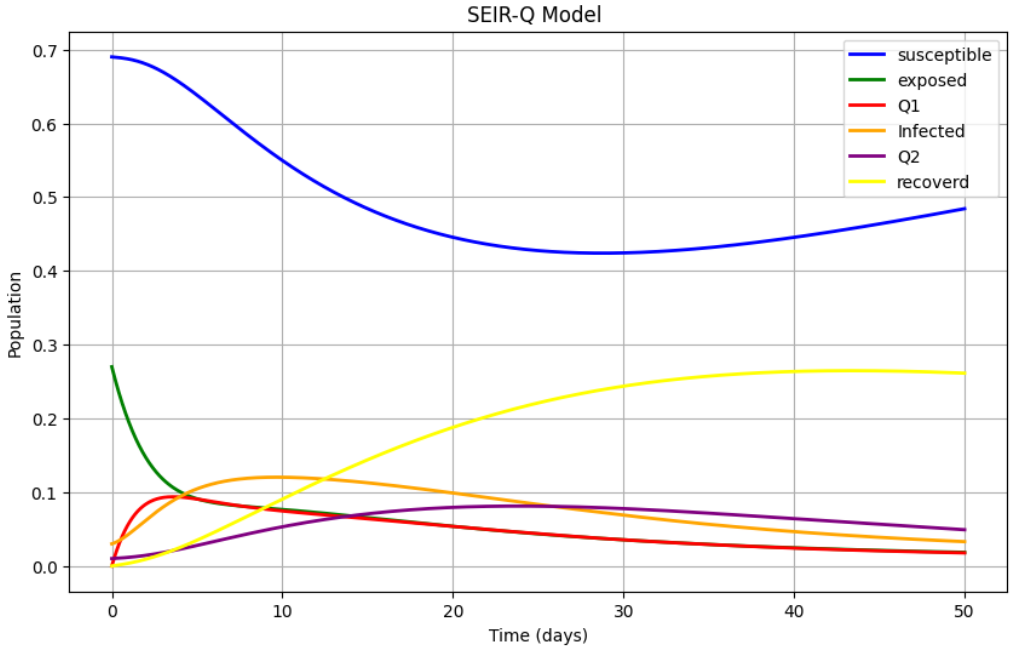


Figure 2.2: Effect of increased transmission and mortality on epidemic dynamics.

Simulation parameters:

$$A = 0.02, \quad \mu_1 = 0.02, \quad \mu_2 = 0.03, \quad \mu_3 = 0.035, \quad \beta = 0.4, \quad \alpha_1 = 0.3, \quad \alpha_2 = 0.05,$$

$$\gamma_1 = 0.3, \quad \gamma_2 = 0.02, \quad \delta_1 = 0.07, \quad \delta_2 = 0.1, \quad \varepsilon = 0.02.$$

The Bhilwara model includes two quarantine compartments, which makes it particularly suitable for COVID-19 situations where suspected individuals may be isolated before confirmation of infection.

From the simulations, it is observed that when the transmission rate β is relatively small, the infectious population grows slowly and the quarantine mechanism helps reduce spread. However, when β is increased, a rapid increase in infections occurs and the epidemic peak becomes higher and earlier.

This behaviour highlights the importance of reducing contact rates through measures such as social distancing, mask usage and quarantine — all of which effectively act to reduce β.

2.1.2 Model from Aloke et al. (2023)

Model compartments and assumptions

The population is partitioned into four compartments: Susceptible $S(t)$, Exposed $E(t)$, Infectious $I(t)$ and Recovered $R(t)$, with total population $N(t) = S + E + I + R$. Assumptions (as in the paper) include homogeneous mixing, non-negative states, and a small probability of loss of immunity.

Mathematical model (7-parameter form)

Following Aloke et al. [1], the extended SEIR ODEs are (parameters renamed to α_i to match the paper):

$$\frac{dS}{dt} = \alpha_6 N - \alpha_1 SI + \alpha_5 R - \alpha_7 S, \tag{2.7}$$

$$\frac{dE}{dt} = \alpha_1 SI - \alpha_2 E - \alpha_7 E, \tag{2.8}$$

$$\frac{dI}{dt} = \alpha_2 E - \alpha_3 I - \alpha_4 I - \alpha_7 I, \tag{2.9}$$

$$\frac{dR}{dt} = \alpha_3 I - \alpha_5 R - \alpha_7 R. \tag{2.10}$$

The parameters are:

- α_1: effective transmission rate (contacts per unit time $\times$ transmission probability),

- α_2: progression rate from exposed to infectious (inverse latent period),
- α_3: recovery rate (inverse infectious period),
- α_4: COVID-19 specific death rate,
- α_5: loss-of-immunity rate (recovered to susceptible),
- α_6: birth (recruitment) rate,
- α_7: natural death rate.

Aloke et al. treat N as approximately constant for the fitting window; however the model above allows for demography when required.
Assumption :α_6=α_7

Basic reproduction number R_0 (next-generation)

Applying the next-generation method (infectious compartments E, I), the reproduction number derived in the paper is:

$$R_0 = \frac{\alpha_1 \alpha_2}{(\alpha_2 + \alpha_7)(\alpha_3 + \alpha_4 + \alpha_7)}. \tag{2.11}$$

This quantity measures the expected secondary infections produced by a single infected individual introduced into a fully susceptible population under the model's assumptions. Aloke et al. reported $R_0 \approx 2.201866$ for the Nigerian fit. [1]

Note: Parameter values are taken from the fitted dataset reported in Aloke et al. [1]. In this simulation, the SEIR model captures the gradual transition from susceptible individuals to exposed, then infectious, and finally recovered. The curve clearly demonstrates the role of incubation delay through the exposed compartment.

Although the SEIR model is simpler than the Bhilwara model, it is still capable of reproducing key qualitative trends. Differences between model output and real data arise because parameters are assumed constant and spatial or behavioural factors are not explicitly modeled.

Nevertheless, the simulation confirms that the estimated parameters are epidemiologically reasonable and consistent with the observed COVID-19 dynamics.

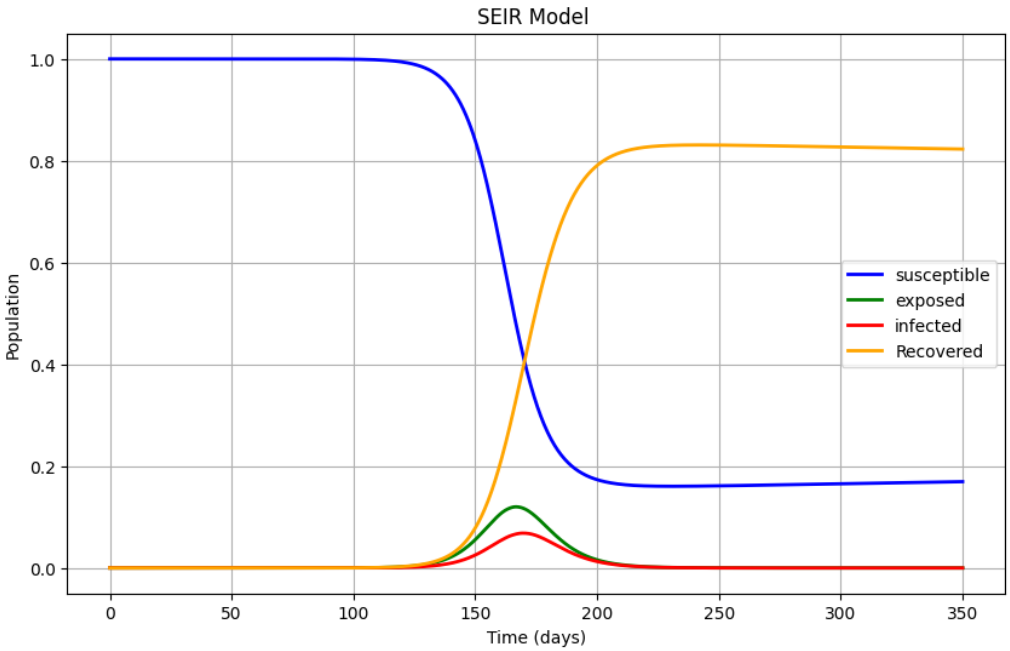


Figure 2.3: SEIR simulation for Nigeria using estimated model parameters.

Chapter 3

Parameter Estimation

3.1 Data and preprocessing

3.1.1 Nigeria data

The Nigerian COVID-19 cumulative case data used for parameter estimation (first 62 days) were obtained from publicly available global repositories [5]. The dataset corresponds to the same period described by Aloke et al. [1], where the first 62 daily cumulative infected observations prior to the epidemic peak were used for training.

In this work we embed the graphs reproduced from our implementation of their model. These data serve as the calibration dataset against which the SEIR model parameters are estimated.

In addition to extracting the cumulative infected series, basic preprocessing steps were applied. Missing entries were smoothed using local interpolation, while obvious reporting spikes were handled by replacing extreme outliers with short moving averages. Daily values were then converted into cumulative counts to reduce noise and reporting irregularities.

The use of cumulative data is particularly helpful in epidemiological modeling because it smooths fluctuations caused by delayed testing or reporting corrections and makes the fitting process numerically more stable. However, it must also be interpreted carefully,

because cumulative values never decrease even when real transmission declines.

3.1.2 India data

For India, the cumulative confirmed infected data were processed and fitted using the same SEIR structure. The dataset was obtained from national COVID-19 reporting platforms [3], while Paul et al. (2021) [4] provide both a model formulation and case study for the Indian epidemic.

Using their study as reference, we preprocess the Indian cumulative case series and fit the cumulative confirmed cases using the same model structure described by Aloke et al. [1].

Similar preprocessing was carried out for the Indian dataset. Data were inspected for inconsistencies, duplicates, and sudden unrealistic jumps caused by backlog releases. Wherever necessary, mild smoothing was applied while ensuring that the original trend of the epidemic was preserved.

Although India experienced multiple epidemic waves, this study focuses on a single continuous period so that parameter estimation remains mathematically well-posed. The goal is not to perfectly reproduce every fluctuation, but to capture the dominant transmission behavior represented in the data.

3.2 Parameter estimation methodology

3.2.1 Objective function: Least–Squares Formulation

Parameter estimation in this study is carried out using the classical non–linear least–squares approach. The goal is to identify parameters such that the model output follows the observed epidemic data as closely as possible.

Let y_i denote the observed cumulative number of infected individuals at time t_i, and let

$$I_m(t_i; \theta)$$

denote the model-predicted value at the same time, where

$$\theta = (\alpha_1, \alpha_2, \ldots, \alpha_7)$$

is the unknown parameter vector.

The residual at each observation is

$$r_i(\theta) = I_m(t_i; \theta) - y_i.$$

The least–squares cost function is

$$J(\theta) = \sum_{i=1}^{n} \left[I_m(t_i; \theta) - y_i \right]^2 .$$

Squaring the errors avoids negative cancellations and penalizes large mismatches more heavily.

Because the SEIR model is nonlinear, numerical optimization algorithms such as the Levenberg–Marquardt (LM) method (implemented in `lmfit` or `scipy.optimize.least_squares`) are required. The algorithm iteratively updates θ until improvement becomes negligible. It is important to emphasize that several different parameter sets may produce visually similar curves. Therefore, the objective is not only to minimize the numerical error, but also to obtain biologically meaningful estimates. Parameters must remain consistent with known clinical characteristics of COVID-19, such as reasonable incubation periods, infectious duration, and recovery times.

To verify reliability, fitted parameters were cross-checked against values reported in the literature, and sensitivity checks were carried out by slightly perturbing the initial guesses and confirming that the optimized solution did not change drastically.

3.2.2 Initial guesses and constraints

To prevent unrealistic parameter collapse or divergence, bounds are enforced:

$$10^{-8} \leq \alpha_i \leq 10, \qquad i = 1, \ldots, 7.$$

Initial guesses were guided by values reported in Aloke et al. [1] and other SEIR literature. Using realistic starting values improves convergence speed and reliability.

After estimation, residual patterns and epidemiological plausibility are examined to ensure credible results.

It is also important to note that slight discrepancies were observed between the datasets used in this study and those reported in Aloke et al. [1]. While the overall structure and trend of the epidemic data are consistent, differences arise due to variations in data sources, reporting updates, and preprocessing procedures. In particular, cumulative case counts may change over time as health authorities revise previously reported figures or correct inconsistencies.

As a result, the parameter estimates and reproduction numbers obtained in this work do not exactly match those reported in Aloke et al. These differences do not indicate errors in the modeling approach, but rather reflect the sensitivity of parameter estimation to the underlying data. Such variability is common in real-data-driven epidemiological studies and highlights the importance of transparent data selection and preprocessing.

3.3 Results — Nigeria

3.3.1 Estimated parameters

Table 3.1 lists the parameters obtained from the fitting procedure.

The estimated values indicate relatively high progression and recovery rates, consistent with the early stages of an emerging epidemic. The extremely small death and natural mortality parameters reflect the fact that these processes occur on a much longer time scale than the model window.

Overall, the parameter set suggests rapid transmission dynamics during the selected training period, which is further supported by the computed reproduction number discussed later.

Parameter	Symbol	Estimated value
Transmission rate	α_1	0.6507310441
Progression rate	α_2	0.5435791335
Recovery rate	α_3	0.4616872431
COVID death rate	α_4	0.0000003294
Loss of immunity	α_5	0.0058295201
Birth rate	α_6	0.0000067631
Natural death rate	α_7	0.0000067631

Table 3.1: Estimated SEIR parameters for Nigeria.

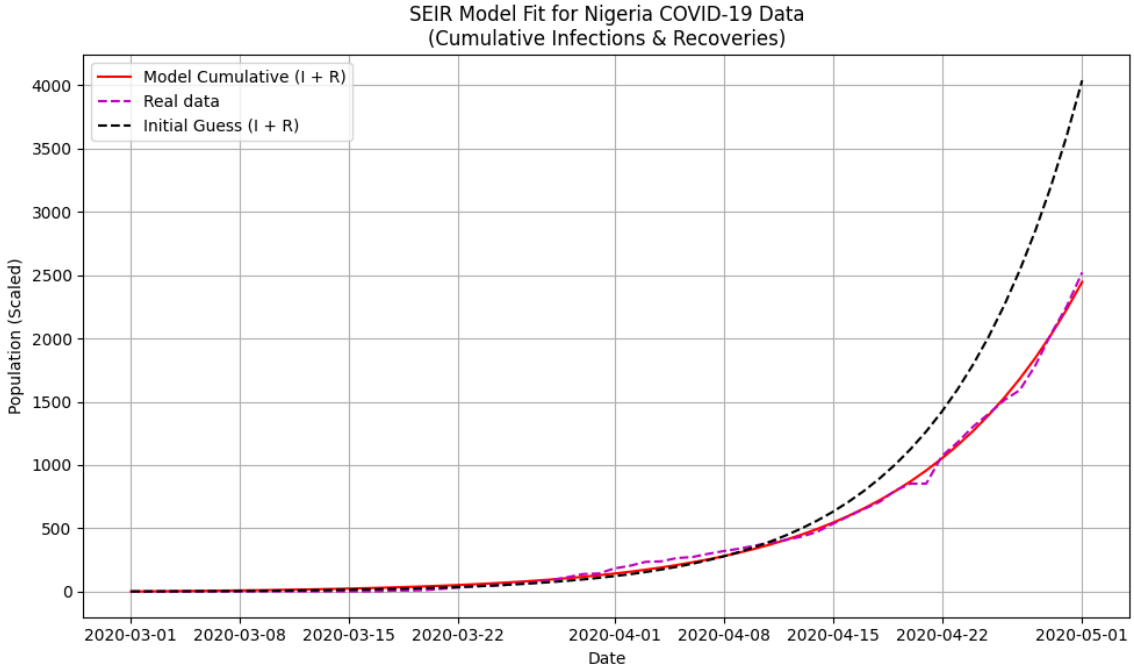


Figure 3.1: Model fit to Nigeria training data (first 62 days).

3.3.2 Fit figure (training)

The fitted curve follows the observed cumulative data closely across the training window, showing that the estimated parameters successfully capture the main pattern of growth. Minor deviations can be attributed to irregular reporting and unmodeled external influences such as policy changes and behavioral responses.

3.3.3 Extended testing

After calibration, the fitted model was simulated beyond the training window to test predictive capability.

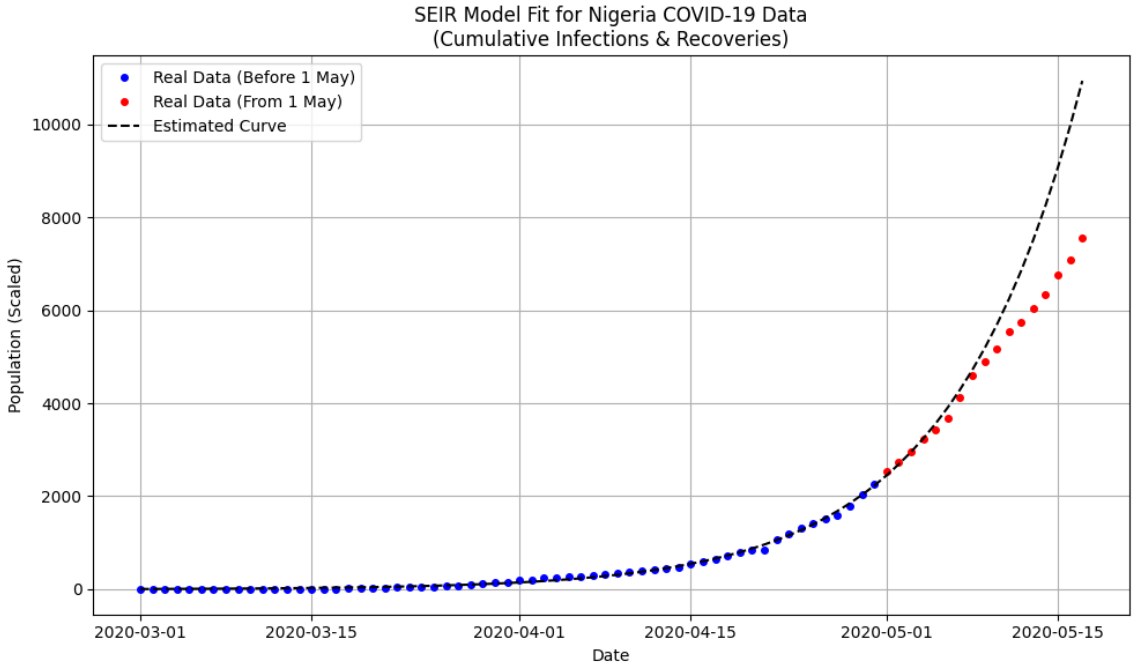


Figure 3.2: Testing the fitted Nigeria SEIR model beyond calibration period.

When the model is extended beyond the calibration window, the prediction begins to diverge slightly from the observed curve. This behavior highlights an important limitation: constant-parameter SEIR models cannot fully represent evolving control strategies, seasonal changes, and vaccination effects. Nevertheless, the extension still reproduces the qualitative trend and provides meaningful insights into the epidemic trajectory.

3.4 Results — India

3.4.1 Estimated parameters

Compared with Nigeria, the transmission and progression rates for India are lower, indicating slower initial spread. This difference is likely influenced by demographic variation, testing policy, and early intervention measures.

Parameter	Symbol	Estimated value
Transmission rate	α_1	0.37605099
Progression rate	α_2	0.16675411
Recovery rate	α_3	0.27354472
COVID death rate	α_4	0.00000012
Loss of immunity	α_5	0.00000010
Birth rate	α_6	0.00000010
Natural death rate	α_7	0.00000010

Table 3.2: Estimated SEIR parameters for India.

Very small mortality and immunity-loss terms again suggest that these effects play only a minor role within the limited time window considered. Despite this, the fitted model successfully reproduces the overall epidemic pattern.

3.4.2 India fit figure

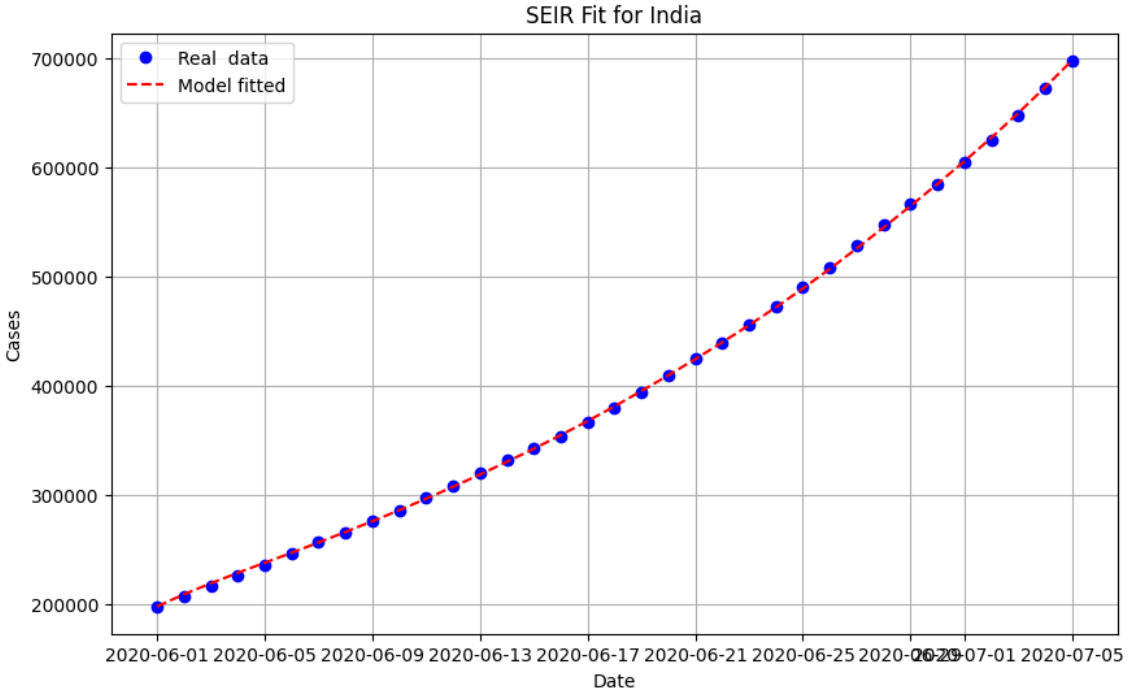


Figure 3.3: SEIR fit for India cumulative case data.

Interpretation. The India trajectory shows a different epidemic shape compared to Nigeria, with slower growth initially and different timing of acceleration. This difference is reflected in the estimated parameter values.

3.4.3 R_0 computation and interpretation

The basic reproduction number R_0 measures the expected number of secondary cases from one infectious case in a fully susceptible population:

$$R_0 = \begin{cases} < 1, & \text{disease dies out,} \\ > 1, & \text{disease spreads.} \end{cases}$$

For this SEIR model,

$$R_0 = \frac{\alpha_1 \alpha_2}{(\alpha_2 + \alpha_7)(\alpha_3 + \alpha_4 + \alpha_7)}.$$

Using the fitted parameters:

$$R_0^{(\text{Nigeria})} = 1.4094, \qquad R_0^{(\text{India})} = 1.3747.$$

These results indicate sustained transmission in both countries during the study window, though Paul et al. reported higher R_0 values due to different datasets and assumptions.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

This study developed a comprehensive framework for analyzing infectious-disease dynamics by integrating real-world data analysis, SEIR-based mathematical modeling, parameter estimation, and reproduction number analysis. COVID-19 datasets were carefully cleaned and examined to understand temporal behaviour, and a deterministic SEIR model was formulated to describe the transition of individuals among susceptible, exposed, infectious, and recovered compartments. Using nonlinear least-squares techniques, key epidemiological parameters were estimated, and the resulting model outputs were compared with observed cumulative case data.

The results demonstrate that the SEIR framework is capable of capturing the qualitative behaviour of COVID-19 transmission and provides meaningful epidemiological indicators, including the basic reproduction number, infectious period, and the potential impact of control measures. During this analysis, discrepancies were identified between the epidemic datasets used in earlier studies and the updated publicly available data considered in the present work. These differences arise from data revisions, reporting updates, and preprocessing choices, which can significantly influence parameter estimation outcomes.

By carefully re-examining and correcting the input datasets prior to model fitting, this

study provides parameter estimates that more accurately reflect the revised data trends. Consequently, the numerical results differ from those reported in previous work, not due to changes in the underlying modeling framework, but because of improved data consistency and preprocessing. Overall, this work confirms that mathematical modeling, when combined with reliable data handling and systematic estimation techniques, offers a powerful tool for interpreting epidemic trends and supporting informed decision-making in public-health contexts. The modeling framework developed here also provides a solid foundation for future extensions and applications to other infectious diseases.

4.2 Future Work

The present study can be extended in several meaningful directions. First, the use of more detailed and disaggregated datasets—including age-structured populations, spatial heterogeneity, and vaccination status—would improve the accuracy and realism of the model. In addition, time-varying parameters can be introduced to capture changes in human behaviour, government interventions, seasonal effects, and public-health responses over time.

A particularly significant direction for future work is the application of the same modeling and parameter estimation framework to **other epidemic diseases**. In future studies, a mathematical model for the disease *Leptospirosis* can be developed, and the techniques used in this project—including SEIR-type modeling, numerical simulation, parameter estimation, and reproduction number analysis—can be applied to study its transmission dynamics. Such an extension would enable comparisons between diseases with different transmission mechanisms and environmental influences, further demonstrating the flexibility and usefulness of mathematical models in epidemiological research.

Bibliography

[1] S. N. Aloke et al. Parameter estimation of covid-19 seir model. *Asian Journal of Pure and Applied Mathematics*, 5(1):229–241, 2023.

[2] B. K. Mishra et al. Mathematical model, forecast and analysis on the spread of covid-19. *Chaos, Solitons & Fractals*, 147:110995, 2021.

[3] C. I. Organization. Covid-19 tracker: India. `https://www.covid19india.org/`, 2020. Accessed: 2025-01-01.

[4] S. Paul et al. Study of seir epidemic model and scenario analysis of covid-19 pandemic. *Ecological Genetics and Genomics*, 2021.

[5] H. Ritchie, E. Ortiz-Ospina, and M. Roser. Coronavirus pandemic (covid-19). `https://ourworldindata.org/coronavirus`, 2020. Accessed: 2025-01-01.