Forecasting Disease Progression Using Time-Varying Parameters

by

S.T. NGOMANE

HONOURS PROJECT
submitted in the fulfilment of the requirements for the degree

Bachelor of Science Honours

in

Applied Mathematics

at the

UNIVERSITY OF JOHANNESBURG

November 2024

SUPERVISOR: Prof F. Chirove

Contents

1 Introduction

1.1 Background information

Over the past 20 years, coronaviruses (CoVs) have been connected to significant illness outbreaks in East Asia and the Middle East [1]. The Middle East respiratory syndrome (MERS) and the severe acute respiratory syndrome (SARS) initially surfaced in 2002 and 2012, respectively [2]. The coronavirus that causes coronavirus disease 2019 (COVID-19), identified as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first arose in late 2019 [3]. It has become a global health concern, producing a pandemic that is currently ongoing in many nations and territories[4], with its usual symptoms of fever, sore throat, exhaustion, cough, or dyspnea[5]. First detected in Wuhan, Hubei Province, China in December 2019, the novel coronavirus, known as SARS-CoV-2, most likely originated from a seafood market that sold live wild animals. It quickly moved from person to person through contact, resulting in COVID-19 [6].

The COVID-19 outbreak was deemed a public health emergency of international concern by the World Health Organization (WHO) on January 30^{th} , 2020 [7]. The pandemic quickly spread, infecting millions of people globally by June 2020. By then, the virus had spread to more than 185 countries, infected over 7,145,800 people, and claimed 407,067 lives[8]. As of April 13^{th}, 2024, the Coronavirus Tracker has recorded 704,753,890 cases and 7,010,681 deaths worldwide.[8] This highlights the critical need for efficient public health interventions and predictive modeling to control and lessen the spread of infectious diseases[9]. Worldometer provided quick and reliable statistics throughout the pandemic, and historical data is still available.[8] However, because most countries have stopped reporting, the data is no longer updated.[8]

Many COVID-19 patients experience a period of relatively modest symptoms followed by a quick escalation, underscoring the need for sophisticated risk classification models. To combat the COVID-19 pandemic and deal with the difficulties during the pandemic, healthcare organizations and physicians around the world used a variety of machine learning and artificial intelligence technologies [10]. Numerous techniques have been offered since the epidemic began, especially for forecasting in-hospital mortality [11], such as advanced machine learning techniques (MLTs) [12]. Being that MLTs have a number of benefits over normal methods, they are being utilized more frequently in clinical settings for outcome prediction [13]. AI is used in the medical field to support physicians' decisions based on their models, not to replace human interactions [14]. Even when there are numerous variables to be examined and only a small number of events have occurred, they aid in understanding intricate correlations between co-variants and outcomes of interest. When more data becomes available, the MLT's prediction capacity might increase by using predictive models.[15] Healthcare providers can use it to identify patients who are more likely to die and offer them support in an effort to minimize deaths as quickly as feasible[15].

1.2 Problem statement

The COVID-19 pandemic has brought attention to the necessity of precise and current disease progression estimates in order to direct public health interventions. Conventional models, such as the susceptible-infectious-recovered (SIR) model, sometimes fall short in explaining dynamic real-world variables including shifting policies, human behavior, and virus mutations. Through the integration of machine learning techniques with the conventional SIR model, this study seeks to increase the accuracy of epidemic forecasting by accounting for time-varying variables.

1.3 Aims

In order to increase the predicting accuracy of newly infected individuals during the COVID-19 pandemic, the main aim of this project is to develop an interpretable model that integrates machine learning techniques and compartmental models. This model will incorporate time-varying characteristics to take into account changes in public opinion and official laws, with a focus on the South African setting.

1.4 Objectives

- To create and examine a set of time-varying, nonlinear first-order differential equations that characterize COVID-19 dynamics in a population.
- To estimate the model's parameters based on the collected data in order to fit the model to the data.
- to determine the impact of policy changes, the number of infections in the daily prior, and the amount of time that has passed since the last policy change on the dynamics of COVID-19 transmission.
- To use the created model to evaluate the impact of different public health initiatives and policy modifications on the COVID-19 outbreak.
- To assess the model's predictive accuracy for both short- and medium-term estimates of new infections.
- to use machine learning methods to find trends and raise the model's prediction accuracy for new infections.

1.5 Significance

Using the work's incorporation of machine learning with the SIR model, disease forecasting is made more adaptable and readily available. The project's output will be extremely beneficial to a wide range of stakeholders, including biological mathematical modelers, health departments, and others interested in the most recent COVID-19 research. It will also improve preparedness and response for epidemics, advancing overall community well-being and adding to the overall findings of epidemiology.

2 Literature review

The necessity of precise forecasting models for forecasting the spread of infectious diseases has been revealed by the COVID-19 pandemic. Standard epidemiological models offer a fundamental framework for interpreting disease dynamics, such as the Susceptible-Infected-Recovered (SIR) model. However, these models frequently attempt to capture the complexity of everyday situations where behavioral shifts, policy adjustments, and other external influences cause parameters to vary over time. Researchers have started incorporating machine learning methods into traditional predictions to increase their responsiveness and forecast accuracy and overcome these limitations.

This review's concept is derived from mathematical and epidemiological research combined with machine learning. Eleven publications are reviewed, five of which are epidemiological studies and six of which concentrate on mathematical and statistical research.

The SIR model is a widely accepted theory that explains the transmission of infectious diseases. The simplicity of the SIR model, while helpful for understanding, becomes an issue in situations that are evolving quickly, such as the COVID-19 pandemic, as noted by [16]. The SIR model assumes fixed parameters, which limits its accuracy in dynamic real-world circumstances where factors such as public health interventions and behavioral changes continuously affect the path of the epidemic. However, the model can be useful in understanding the fundamental dynamics of disease spread.

To overcome the limitations of the SIR model, several research have suggested incorporating machine learning techniques into the model. In [16], for example, SIMLR—a machine learning-enhanced SIR model—was presented. Through the use of machine learning algorithms, the model dynamically modifies parameters based on real-time data, making it more accurate and responsive in its ability to predict the progression of disease.
A comprehensive evaluation of the several uses of AI and machine learning in COVID-19 pandemic management is provided by [17]. These applications include forecasting the spread of disease, allocating resources as efficiently as possible, and creating individual treatments. [18] illustrates the wide-ranging potential of these technologies in pandemic management by emphasizing further how artificial intelligence may improve diagnostic tools and increase the effectiveness of public health responses.

Using symptom data, [19] showed how well machine learning models could predict the diagnosis of COVID-19. Their research demonstrates how these models can help with early identification and intervention efforts, which are critical to stopping the virus's spread. Similarly, [20] used a variety of machine learning techniques to examine COVID-19 data, demonstrating how well these techniques could identify complex trends and boost prediction accuracy. These findings highlight how crucial it is to improve the ability to predict of epidemiological models by integrating complex data analysis methods. [8]

Evaluation using testing is essential to proving that improved models work. The impact of social distance on the COVID-19 epidemic in South Africa was modeled by [21], offering insights into how policy actions can change disease dynamics. Their results highlight the significance of incorporating time-varying characteristics into models to correctly capture the practical effects of public health policies. [22] demonstrated the promise of sophisticated mathematical techniques in epidemic modeling by using optimal control theory to predict COVID-19 dynamics in South Africa. Their research demonstrates how combining machine learning with conventional models can result in more successful intervention tactics. With an emphasis on vaccination models, [23] confirmed the suitability of upgraded SIR models in various settings. Their research highlights the value of dynamic models in the planning and evaluation of vaccination programs.

There are important policy ramifications when machine learning is included in standard epidemiological models. Better forecasting models have the potential to improve public health outcomes by assisting policymakers in making timely and well-informed decisions. [24] highlighted the necessity for ongoing innovation in this discipline as they explored the contributions of machine learning techniques in reacting to SARS-CoV-2 infection.

One major development in epidemic forecasting is the integration of machine learning methods with the standard SIR model. These updated models allow more successful public health interventions by providing greater predicted accuracy and response for dynamically adapting model parameters to reflect real-time data. Future studies need to concentrate on enhancing these models and investigating their utilization in different epidemiology situations. Our capacity to regulate and reduce the effects of infec-

tious diseases is expected to improve with the continued integration of machine learning into standard epidemiology techniques, which will ultimately lead to the development of better public health systems.

3 Main Model

3.1 Model formulation

A compartmental model called the SIR model is used to explain how a disease spreads through a population. The population is split into three different classes: Recovered R, who have recovered from the disease and are presumed to have immunity, and cannot contract the disease again; Susceptible S, who are healthy but may contract the disease; and Infected I, who have already contracted the disease and are currently infectious. The sizes of these classes change over time, represented as functions of time t, $S(t)$: Number of susceptible individuals at a certain time, $I(t)$: Number of infected individuals at a certain time $R(t)$: Number of recovered individuals at a certain time. The overall population size, denoted as $N = S(t) + I(t) + R(t)$, stays constant. The term "transmission rate" $\beta(t)$ refers to the speed at which vulnerable persons become ill after coming into contact with infected individuals. The recovery rate, $\gamma(t)$, represents the probability of recovery for an infected individual. It also represents the probability of disease transmission per contact between a susceptible and an infected individual. In addition, it represents the rate at which infected individuals recover from the disease and move to the recovered class.

The SIR model equations with the assumption of constant population:

$$\begin{cases} \dfrac{dS}{dt} = -\beta S \dfrac{I}{N} \\ \dfrac{dI}{dt} = -\beta S \dfrac{I}{N} + \gamma I \\ \dfrac{dR}{dt} = \gamma I \end{cases} \tag{1}$$

An overview of variables and parameters

Variables	**Description**	**Units**
S	Population of the susceptibles	Individuals
I	Population of the infected	Individuals
R	Population of the recovered	Individuals
t	Time	Days

Table 1: Variables

Parameters	**Description**	**Units**
β	Transmission rate	Per day
γ	Recovery rate	Per day

Table 2: Parameters and their Description

Firure 1: Shows the flow-diagram of the SIR model:

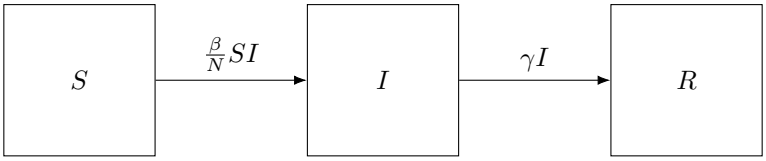

Figure 1: Model of SIR Diagram

3.2 Model Analysis

3.2.1 Feasible region

The total population remains constant over time, where $N = S + I + R$ which is constant over time.

$$\begin{aligned}
\frac{d}{dt}\left[N = R + S + I\right] &= (\gamma I) + \left(-\beta S\frac{I}{N}\right) + \left(-\beta S\frac{I}{N} - \gamma I\right) \\
\frac{d}{dt}N &= 0 \\
\int_0^t N(s)d &= \int_0^t 0dt \\
N(t) - N(0) &= 0 \\
N(t) &= N(0) \\
\implies N(t) &= N_0
\end{aligned}$$

$N(t) = N_0$ for all t in the interval $0 < N_0 < t$, where N_0 is the initial total population at $t = 0$.

The collection of all possible states (values of S, I, and R) that the system can reach under the model's restrictions and initial conditions is represented by the feasible region as:

$$\Omega = \{(S, I, R) \in \mathbb{R}^3_+ \mid 0 < N(t) \leq N_0\}$$

3.2.2 Existence and uniqueness

- **Positivity of boundary of solutions:**

Considering the equation for the susceptible class, we have:

$$\begin{aligned}
\frac{d}{dt}S &= -\beta S\frac{I}{N} \\
\frac{dS}{dt} + \beta S\frac{I}{N} &= 0 \\
IF &= e^{\int_0^t \frac{\beta}{N}I(s)ds} = e^{\frac{\beta}{N}\int_0^t I(s)ds} \\
\int_0^t d(S(s)IF(s)) &= 0\int_0^t dt \\
S(t)e^{\frac{\beta}{N}\int_0^t I(s)ds} - S(0)e^0 &= 0 \\
S(t)e^{\frac{\beta}{N}\int_0^t I(s)ds} &= S(0) \\
S(t) &= S(0)e^{-\frac{\beta}{N}\int_0^t I(s)ds} \\
\implies S &= S_0e^{-\frac{\beta}{N}\int_0^t I(s)ds}
\end{aligned}$$

The starting number of susceptible individuals is $S(0) = S_0$, and the exponential function is always positive.
$\implies S(t) > 0, \quad \forall t \geq 0$

Considering the equation for the infected class, we have:

$$\begin{aligned}
\frac{dI}{dt} &= -\beta S\frac{I}{N} + \gamma I \\
\frac{dS}{dt} + \beta S\frac{I}{N} - \gamma I &= 0 \\
\frac{dS}{dt} + (\gamma - \frac{\beta}{N}SI &= 0 \\
IF &= e^{\int_0^t (\gamma - \frac{\beta}{N}Sdp} \\
\int_0^t d(I(p)IF(p)) &= 0\int_0^t dt \\
Ie^{\int_0^t (\gamma - \frac{\beta}{N}Idp} - I(0)e^0 &= 0 \\
Ie^{\int_0^t (\gamma - \frac{\beta}{N}Idp} &= I(0) \\
I &= I(0)e^{\int_0^t (\frac{\beta}{N}S(p) - \gamma)dp} \\
\implies I &= I_0 e^{\int_0^t (\frac{\beta}{N}S(p) - \gamma)dp}
\end{aligned}$$

The starting number of Infected individuals is $I(0) = I_0$, and the exponential function is always non-negative.
$\implies I(t) \geq 0, \quad \forall t \geq 0$

Considering the Recovered class, we have:

$$\begin{aligned}
\frac{dR}{dt} &= \gamma I \\
\int_0^t dR &= \int_0^t \gamma I(s)ds \\
R - R(0) &= \gamma \int_0^t I(s)ds \\
R &= R(0) + \gamma \int_0^t I(s)ds
\end{aligned}$$

For every t in $(0, t)$, the integrand $\gamma I(s)$ is non-negative, therefore the integral will also be non-negative. So R(t) $\geq 0, \quad \forall t \geq 0$

The solution of all the classes will always be non-negative for all $t \geq 0$.

3.2.3 Boundedness of solutions

As shown above,

- The Susceptible:
  $S = S_0 e^{-\frac{\beta}{N}\int_0^t I(s)ds}$, $0 \leq S(t) \leq N_0$
- Infectious:
  $I = I_0 e^{\int_0^t (\frac{\beta}{N}S(s) - \gamma)ds}$, $0 \leq I(t) \leq N_0$
- Recovered:
  $R = R(0) + \gamma \int_0^t I(s)ds$, $0 \leq R(t) \leq N_0$

The Populations that satisfy $0 \leq S(t) \leq N_0$, $0 \leq I(t) \leq N_0$, and $0 \leq R(t) \leq N_0$ are respectively the Susceptible $S(t)$, Infectious $I(t)$, and Recovered $R(t)$. All three populations are non-negative. Within the closed interval $0 \leq S(t), I(t), R(t) \leq N_0$.

3.2.4 Equilibrium Points

The SIR equations:

$$\frac{dS}{dt} = -\beta S\frac{I}{N}, \;\; \frac{dI}{dt} = -\beta S\frac{I}{N} + \gamma I, \;\; \frac{dR}{dt} = \gamma I$$

At equilibrium:

$$\frac{dS}{dt} = -\beta S\frac{I}{N} = 0 \tag{2}$$

$$\frac{dI}{dt} = -\beta S\frac{I}{N} + \gamma I = 0 \tag{3}$$

$$\frac{dR}{dt} = \gamma I = 0 \tag{4}$$

Equation 6: $\gamma I(t) = 0$, we get:

$$I(t) = 0.$$

Substituting $I(t) = 0$ into the equation 4:

$$\frac{dS}{dt} = -\beta S\frac{0}{N} = 0,$$

For equation 5:

$$\frac{dI(t)}{dt} = -\beta S\frac{0}{N} + \gamma \cdot 0 = 0,$$

We have $I(t) = 0$, $\frac{dS(t)}{dt} = 0$ and $\frac{dI(t)}{dt} = 0$
This gives us one equilibrium point:

$$E^* = (S(t), 0, R(t)) \text{ where } S(t) + R(t) = N.$$

As may be observed, the following are the SIR model's equilibrium points: $E^* = (N, 0, 0)$ in addition to $(0, 0, 0)$. Yes, there is an equilibrium point at $(S, I, R) = (0, 0, 0)$. In the context of an epidemic model, however, it suggests that there are no susceptible, infected, or recovered individuals in the community. However, mathematically speaking, it is an equilibrium point. As a result, the equilibrium point free is $E^* = (S^*, I^*, R^*) = (N, 0, 0)$

3.2.5 Basic Reproduction Number

- **Next Generation Matrix**

The system is linearized in disease-free equilibrium (DFE) and the matrix corresponding to the anticipated number of secondary infections in every class is represented. The disease-free equilibrium (DFE), $I = 0$, $R = 0$, and $S = N$ (the population as a whole is nearly sensitive)
Let the generation matrix be $K = FV^{-1}$, where the transmission matrix F represents new infections, transition matrix V represents the movement of individuals between classes:

The differential equation for the infectious class:I

$$\frac{dI}{dt} = \beta \frac{S}{N} I - \gamma I$$

Which corresponds to:

$$\frac{dI}{dt} = \mathcal{F}(I) - \mathcal{V}(I)$$

where $\mathcal{F}(I)$ represents the rate of new infections.
Hence:

$$\mathcal{F}(I) = \beta \frac{S}{N} I$$

For the new infection:

$$F = \frac{d\mathcal{F}}{dI} = \beta \frac{S}{N}$$

Now $\mathcal{V}(I)$ represents the rate of transition out of the infectious class.

$$\mathcal{V}(I) = \gamma I$$

For the transition terms out of the infected state:

$$\frac{d\mathcal{V}}{dI} = \gamma$$

The inverse of V:

$$V^{-1} = \frac{1}{\gamma}$$

The next generation matrix which is $1x1$:

$$\begin{aligned} K &= FV^{-1} \\ &= \beta \frac{S}{N} \cdot \frac{1}{\gamma} \\ &= \frac{\beta S}{\gamma N} \end{aligned}$$

At the disease-free equilibrium, $S \approx N$ (since almost everyone is susceptible), so:

$$K = \frac{\beta}{\gamma} = R_0$$

Which is known as the basic reproduction number (R_0)

- **Basic Reproduction Number R_0**

The basic reproduction number R_0 is defined as the expected number of secondary cases produced by a single infected individual in a completely susceptible population. We consider the new infections

generated by one infectious individual in a susceptible population:

$$R_0 = \frac{\beta}{\gamma}$$

Which is the rate at which an infectious individual generates new infections (β) compared to the rate at which individuals recover (γ).

3.2.6 Stability Analysis

From the SIR model equations The Jacobian matrix J of the model is given by:

$$J = \begin{pmatrix} \frac{\partial}{\partial S}\left(-\frac{\beta}{N}SI\right) & \frac{\partial}{\partial I}\left(-\frac{\beta}{N}SI\right) & \frac{\partial}{\partial R}\left(-\frac{\beta}{N}SI\right) \\ \frac{\partial}{\partial S}\left(\frac{\beta}{N}SI - \gamma I\right) & \frac{\partial}{\partial I}\left(\frac{\beta}{N}SI - \gamma I\right) & \frac{\partial}{\partial R}\left(\frac{\beta}{N}SI - \gamma I\right) \\ \frac{\partial}{\partial S}\left(\gamma I\right) & \frac{\partial}{\partial I}\left(\gamma I\right) & \frac{\partial}{\partial R}\left(\gamma I\right) \end{pmatrix}$$

The partial derivatives are:

$$J = \begin{pmatrix} -\frac{\beta}{N}I & -\frac{\beta}{N}S & 0 \\ \frac{\beta}{N}I & \frac{\beta}{N}S - \gamma & 0 \\ 0 & \gamma & 0 \end{pmatrix}$$

Disease-free equilibrium: $E^* = (S^*, I^*, R^*) = (N, 0, 0)$.

$$J|_{(N,0,0)} = \begin{pmatrix} 0 & -\beta & 0 \\ 0 & \beta - \gamma & 0 \\ 0 & \gamma & 0 \end{pmatrix}$$

The characteristic equation is:

$$J - \lambda I = \begin{pmatrix} -\lambda & -\beta & 0 \\ 0 & \beta - \gamma - \lambda & 0 \\ 0 & \gamma & -\lambda \end{pmatrix}$$

The determinant is:

$$\begin{aligned} |J - \lambda I| &= 0 \\ (-\lambda)\left((\beta - \gamma - \lambda)(-\lambda)\right) - (-\beta)(\gamma)(0) + 0 &= 0 \\ \lambda^2(\beta - \gamma - \lambda) &= 0 \end{aligned}$$

Thus, the eigenvalues are:

$$\begin{aligned} \lambda_1 &= 0, \\ \lambda_2 &= 0, \\ \lambda_3 &= \beta - \gamma \end{aligned}$$

Now,

$$\begin{aligned} |J|_(N, 0, 0)| &= \lambda_1 \cdot \lambda_2 \cdot \lambda_3 \\ &= 0 \cdot 0 \cdot (\beta - \gamma) \\ &= 0 \end{aligned}$$

Therefore, the matrix $|J|_{(N,0,0)}$ is not invertible.

- **Threshold Analysis**

We use a different method, which is threshold analysis or threshold theorem: We begin analyzing the infectious class:

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I$$

the initial rate of change of the infected individuals $I(0) = I_0$:

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I = \gamma\left(\frac{\beta}{N}S - \gamma\right)$$
$$\left.\frac{dI}{dt}\right|_{t=0} = I_0\left(\frac{\beta}{N}S_0 - \gamma\right)$$

With the basic reproduction number $R_0 = \frac{\beta}{\gamma}$ and $S(0) = S_0$, $I(0) = I_0$ then

$$I_0\left(\frac{\beta}{N}S_0 - \gamma\right) > 0$$
$$\frac{\beta}{N}S_0 I_0 > \gamma I_0$$
$$\frac{\beta}{N}S_0 I_0 \cdot \frac{N}{\beta I_0} > \gamma I_0 \cdot \frac{N}{\beta I_0}$$

$\implies S_0 > \frac{\gamma N}{\beta}$, so the infection will grow.

$$I_0\left(\frac{\beta}{N}S_0 - \gamma\right) < 0$$
$$\frac{\beta}{N}S_0 I_0 < \gamma I_0$$
$$\frac{\beta}{N}S_0 I_0 \cdot \frac{N}{\beta I_0} < \gamma I_0 \cdot \frac{N}{\beta I_0}$$
$$\implies S_0 < \frac{\gamma N}{\beta}$$

$\implies S_0 < \frac{\beta}{\gamma}N$, so the infection will die out.

Analyzing the Susceptible and Infectious class:

$$\frac{dS}{dt} = -\frac{\beta}{N}SI, \quad \frac{dI}{dt} = \frac{\beta}{N}SI - \gamma I.$$

$$\implies \frac{\left(\frac{dI}{dt}\right)}{\left(\frac{dS}{dt}\right)} = \frac{dI}{dS} = \frac{\left(\frac{\beta}{N}SI - \gamma I\right)}{-\frac{\beta}{N}SI}$$
$$= -1 + \frac{\gamma I}{\frac{\beta}{N}SI}$$
$$= -1 + \frac{N\left(\frac{\gamma}{\beta}\right)}{S}$$
$$= -1 + \frac{N}{SR_0}$$

$$\frac{dI}{dS} = -1 + \frac{N}{SR_0}$$
$$dI = \left(-1 + \frac{N}{SR_0}\right) dS$$
$$\int_0^t dI = \int_0^t \left(-1 + \frac{N}{SR_0}\right) dS$$
$$I(t) - I(0) = \left(-S(s) + \frac{N}{R_0} \ln S(s)\right)\Big|_0^t$$
$$I(t) - I(0) = -S(t) + \frac{N}{R_0} \ln S(t) + S(0) - \frac{N}{R_0} \ln S(0)$$
$$I(t) + S(t) - \frac{N}{R_0} \ln S(t) = I(0) + S(0) - \frac{N}{R_0} \ln S(0)$$

With the initial conditions: $I(t) = I, S(t) = S \ \ and \ \ S(0) = S_0, I(0) = I_0$

$$I + S - \frac{N}{R_0} \ln S = I_0 + S_0 - \frac{N}{R_0} \ln S_0.$$

Maximum number of infectives occurs when $\frac{dI}{dt} = 0$ from the model, which leads to $S = \frac{\gamma N}{\beta} = \frac{N}{R_0}$:

$$I_{max} + \frac{N}{R_0} - \frac{N}{R_0} \ln\left(\frac{N}{R_0}\right) = I_0 + S_0 - \frac{N}{R_0} \ln S_0.$$
$$I_{max} = I_0 + S_0 - \frac{N}{R_0} + \frac{N}{R_0} \ln\left(\frac{N}{R_0}\right) - \frac{N}{R_0} \ln S_0$$
$$I_{max} = I_0 + S_0 - \frac{N}{R_0} + \frac{N}{R_0} \ln\left(\frac{N}{R_0 S_0}\right)$$

From the boundness of solutions, we had that $S_0 > 0, I_0 > 0$ and since $R(0) = 0$ for $t > 0$, $0 \leq S + I < N$.

$$I_{max} = N - \frac{N}{R_0} + \frac{N}{R_0} \ln\left(\frac{N}{R_0 S_0}\right)$$

Analyzing the Susceptible and Recovery Class:

$$\frac{\frac{dS}{dt}}{\frac{dR}{dt}} = \frac{dS}{dR} = -\frac{\frac{\beta}{N} SI}{\gamma I} = -\frac{R_0 S}{N}$$

$$\frac{dS}{dR} = -\frac{R_0 S}{N}$$
$$\int \frac{1}{S} dS = -\int \frac{R_0 S}{N} dR$$
$$\ln S(t) = -\frac{R_0}{N} R(t) + C$$
$$S(t) = K e^{-\frac{R_0}{N} R(t)}$$
$$S(0) = S_0 = K e^{-\frac{R_0}{N} R(0)} = K$$
$$\implies S(t) = S_0 e^{-\frac{R_0}{N} R}$$
$$S(t) = S_0 e^{-\frac{R_0}{N} R} \geq S_0 e^{-\frac{R_0}{N} N} = S_0 e^{-R_0} > 0$$

Analyzing the Recovery only now:

$$\frac{dR}{dt} = \gamma I = \gamma \left(N - R - S_0 e^{-\frac{R_0}{N} R}\right)$$

Tylor series expansions of $e^{-\frac{R_0}{N}R}$:

$$e^{-\frac{R_0}{N}R} \approx 1 - \frac{R_0}{N}R + \frac{R_0^2}{2N^2}R + O(R_0R^3)$$

$$\begin{aligned}
\frac{dR}{dt} &= \gamma\left[N - R - S_0\left(1 - \frac{R_0}{N}R + \frac{R_0^2}{2N^2}R\right)\right] \\
\frac{dR}{dt} &= \gamma\left[N - R - S_0 + \frac{R_0S_0}{N}R - \frac{R_0^2S_0}{2N^2}R^2\right] \\
\frac{dR}{dt} &= \gamma\left[N - S_0 + \left(\frac{R_0S_0}{N} - 1\right)R - \frac{R_0^2S_0}{2N^2}R^2\right]
\end{aligned}$$

Let $A = N - S_0$, $B = \frac{R_0S_0}{N} - 1$ and $C = \frac{R_0^2S_0}{2N^2}$.

$$\begin{aligned}
\implies \frac{dR}{dt} &= \gamma\left(A + BR - CR^2\right) \\
&= -\gamma C\left(R^2 - \frac{B}{C}R - \frac{A}{C}\right) \\
&= -\gamma C\left[\left(R - \frac{B}{2C}^2\right) - \frac{B^2}{4C^2} - \frac{A}{C}\right] \\
&= -\gamma C\left[\left(R - \frac{B}{2C}^2\right) - \frac{B^2 + 4AC}{4C^2}\right] \\
&= -\gamma C\left[-\frac{B^2 + 4AC}{4C^2} + \left(R - \frac{B}{2C}^2\right)\right] \\
&= \gamma C\left[\left(\frac{\sqrt{B^2 + 4AC}}{2C}\right)^2 - \left(R - \frac{B}{2C}\right)^2\right]
\end{aligned}$$

Let $u = R - \frac{B}{2C}$, $\frac{du}{dt} = \frac{dR}{dt}$ and $D = \frac{\sqrt{B^2 + 4AC}}{2C}$

$$\frac{du}{dt} = \gamma C\left(D^2 - u^2\right)$$

$$\frac{du}{D^2 - u^2} = \gamma C$$

$$\int \frac{du}{D^2 - u^2} = \gamma C \int dt$$

Let $u = D\tanh(\theta) \implies du = D\mathrm{sech}^2(\theta)\,d\theta$

$$\int \frac{D\mathrm{sech}^2(\theta)}{D^2 - D^2\tanh^2(\theta)}d\theta = \int \gamma C dt$$

$$\frac{1}{D}\int \frac{\mathrm{sech}^2(\theta)}{1 - \tanh^2(\theta)}d\theta = \int \gamma C dt$$

$$\frac{1}{D}\int \frac{\mathrm{sech}^2(\theta)}{sech^2(\theta)}d\theta = \int \gamma C dt$$

$$\frac{1}{D}\int d\theta = \int \gamma C t$$

$$\frac{\theta}{D} = \gamma C t + k, \text{ Where } k \text{ is a constant}$$

$$\text{Since } u = D\tanh(\theta)$$

$$\tanh(\theta) = \frac{u}{D} \implies \theta = \tanh^{-1}\left(\frac{u}{D}\right)$$

$$\frac{1}{D}\tanh^{-1}\left(\frac{u}{D}\right) = \gamma C t + k$$

$$\tanh^{-1}\left(\frac{R - \frac{B}{2C}}{D}\right) = D\left(\gamma C t + k\right)$$

$$\frac{R - \frac{B}{2C}}{D} = \tanh\left(D\left(\gamma C t + k\right)\right)$$

$$R - \frac{B}{2C} = D\tanh\left(D(\gamma C t + k)\right)$$

$$R(t) = \frac{B}{2C} + D\tanh\left(D(\gamma C t + k)\right)$$

for $t = 0$:

$$R(0) = \frac{B}{2C} + D\tanh\left(D(\gamma C(0) + k)\right)$$

$$R(0) = \frac{B}{2C} + D\tanh\left(Dk\right)$$

$$R(0) - \frac{B}{2C} = D\tanh\left(Dk\right)$$

$$\tanh\left(Dk\right) = \frac{R(0) - \frac{B}{2C}}{D}$$

$$Dk = \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)$$

$$k = \frac{1}{D}\tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)$$

Now

$$R(t) = \frac{B}{2C} + D\tanh\left(D\left(\gamma C t + \frac{1}{D}\tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)\right)$$

$$= \frac{B}{2C} + D\tanh\left(\gamma DC t + \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)$$

$$\text{With } DC = \left(\frac{\sqrt{B^2 + 4AC}}{2C}\right) \cdot C = \frac{1}{2}\sqrt{B^2 + 4AC}$$

$$= \frac{1}{2}\sqrt{\left(\frac{R_0 S_0}{N} - 1\right)^2 + \frac{2R_0^2 S_0(N - S_0)}{N^2}},$$

$$\text{and let } \sqrt{\left(\frac{R_0 S_0}{N} - 1\right)^2 + \frac{2R_0^2 S_0(N - S_0)}{N^2}} = \alpha$$

$$R(t) = \frac{B}{2C} + D\tanh\left(\frac{\gamma\alpha t}{2} + \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)$$

$$R(t) = \frac{B}{2C} + \frac{\sqrt{B^2+4AC}}{2C}\tanh\left(\frac{\gamma\alpha t}{2} + \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)$$

$$R(t) = \frac{1}{2\left(\frac{R_0^2 S_0}{2N^2}\right)}\left[\frac{R_0 S_0}{N} - 1 + \sqrt{B^2+4AC}\tanh\left(\frac{\gamma\alpha t}{2} + \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)\right]$$

$$R(t) = \frac{N^2}{R_0^2 S_0}\left[\left(\frac{R_0 S_0}{N} - 1\right) + \alpha\tanh\left(\frac{\gamma\alpha t}{2} + \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right)\right)\right]$$

$$\begin{aligned}
\text{Let } \phi &= \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right), \text{ with } R(0) = 0 \\
&= \tanh^{-1}\left(\frac{R(0) - \frac{B}{2C}}{D}\right) \\
&= \tanh^{-1}\left(-\frac{B}{2C} \times \frac{1}{D}\right) \\
&= \tanh^{-1}\left(-\frac{B}{2C} \times \frac{2C}{\sqrt{B^2+4AC}}\right) \\
&= -\tanh^{-1}\left(\frac{\frac{R_0 S_0}{N} - 1}{\sqrt{B^2+4AC}}\right) \\
\phi &= -\tanh^{-1}\left(\frac{\frac{R_0 S_0}{N} - 1}{\alpha}\right)
\end{aligned}$$

$$R(t) = \frac{N^2}{R_0^2 S_0}\left[\left(\frac{R_0 S_0}{N} - 1\right) + \alpha\tanh\left(\frac{\gamma\alpha t}{2} + \phi\right)\right]$$

$$\implies \frac{dR}{dt} = \frac{N^2}{R_0^2 S_0}\left[\frac{\alpha^2\gamma}{2}\text{sech}^2\left(\frac{\gamma\alpha t}{2} + \phi\right)\right]$$

Thus the Removal Rate:

$$\frac{dR}{dt} = \frac{N^2\alpha^2\gamma}{2R_0^2 S_0}\text{sech}^2\left(\frac{\gamma\alpha t}{2} + \phi\right)$$

4 Data Analysis

We looked at Gauteng's COVID-19 Provincial Cumulative Timeline Confirmed Cases data in our research, covering every wave of the pandemic from the earliest to the latest. With a focus on the fifth wave, which ran from November 9, 2021, to April 15, 2022, the dataset covers the period from March 5, 2020, to July 25, 2022. From a larger dataset that contained all provinces, we took out the pertinent information for Gauteng and filled in the blanks by averaging the values of the days that followed. To see how each wave progressed, we showed the total confirmed instances. We also made a new column to examine the daily frequency of new cases. The frequency plot offers a comprehensive picture of the epidemic by clearly displaying its different waves. 1

4.1 Simulation

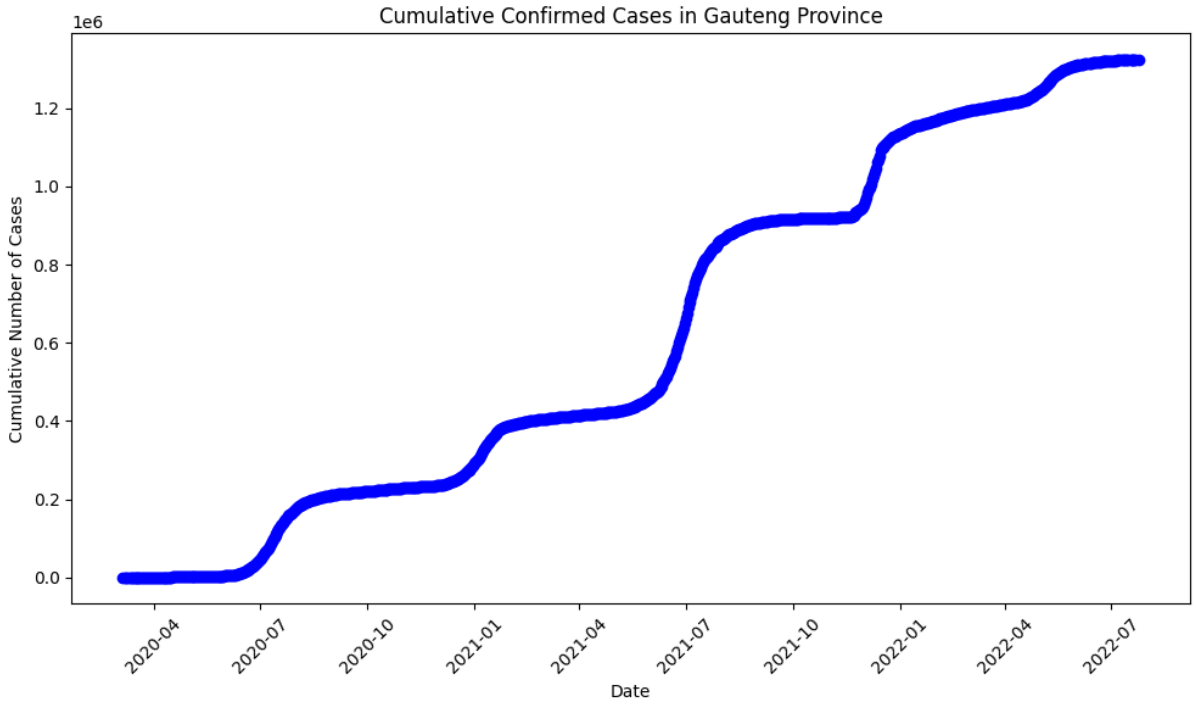

Figure 2: Cumulative Confirmed Cases in Gauteng Province

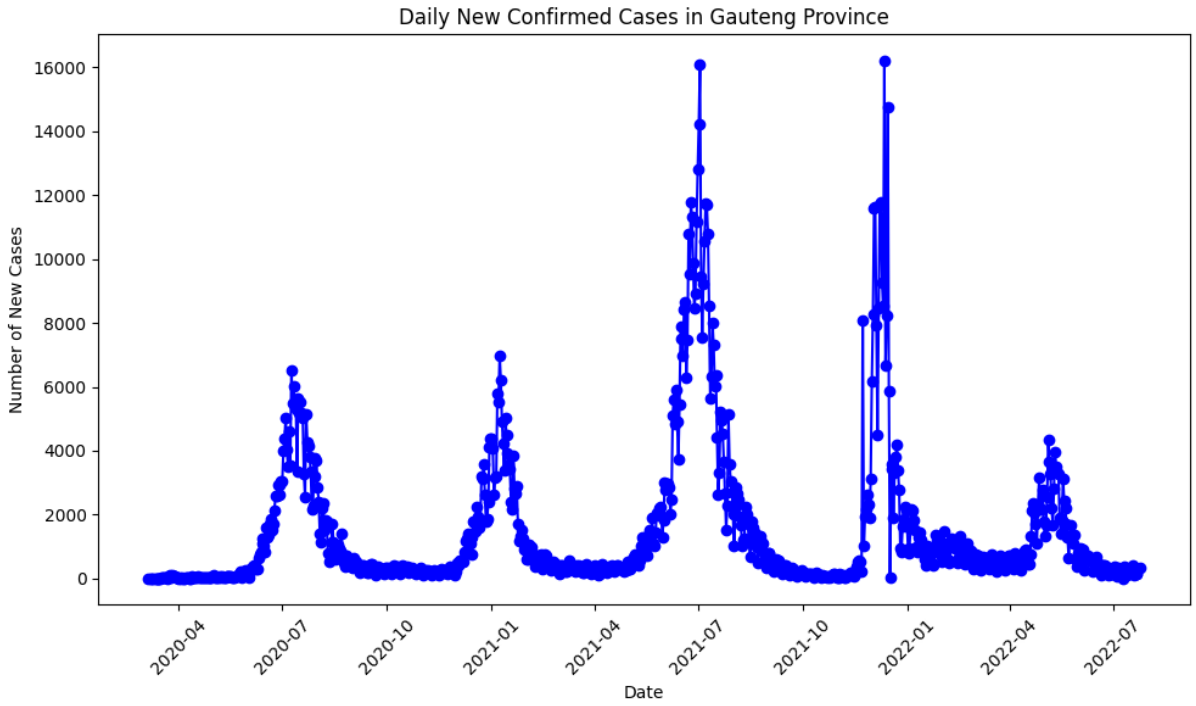

Figure 3: Daily New Confirmed Cases in Gauteng Province

4.2 Prediction

5 Discussions and Conclusion

A Code listings

- Data Cleaning.

References

[1] Jasper FW Chan, Susanna KP Lau, Kelvin KW To, Vincent CC Cheng, Patrick CY Woo, and Kwok-Yung Yuen. Middle east respiratory syndrome coronavirus: another zoonotic betacoronavirus causing sars-like disease. *Clinical microbiology reviews*, 28(2):465–522, 2015.

[2] David S Hui. Epidemic and emerging coronaviruses (severe acute respiratory syndrome and middle east respiratory syndrome). *Clinics in chest medicine*, 38(1):71–86, 2017.

[3] Mahendra Pal, Gemechu Berhanu, Chaltu Desalegn, and Venkataramana Kandi. Severe acute respiratory syndrome coronavirus-2 (sars-cov-2): an update. *Cureus*, 12(3), 2020.

[4] Alfonso J Rodriguez-Morales, D Katterine Bonilla-Aldana, Graciela Josefina Balbin-Ramon, Ali A Rabaan, Ranjit Sah, Alberto Paniz-Mondolfi, Pasquale Pagliano, Silvano Esposito, et al. History is repeating itself: Probable zoonotic spillover as the cause of the 2019 novel coronavirus epidemic. *Infez Med*, 28(1):3–5, 2020.

[5] Tomasz Dzieciatkowski, Lukasz Szarpak, Krzysztof J Filipiak, Milosz Jaguszewski, Jerzy R Ladny, and Jacek Smereka. Covid-19 challenge for modern medicine. *Cardiology journal*, 27(2):175–183, 2020.

[6] Noah C Peeri, Nistha Shrestha, Md Siddikur Rahman, Rafdzah Zaki, Zhengqi Tan, Saana Bibi, Mahdi Baghbanzadeh, Nasrin Aghamohammadi, Wenyi Zhang, and Ubydul Haque. The sars, mers and novel coronavirus (covid-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *International journal of epidemiology*, 49(3):717–726, 2020.

[7] Catrin Sohrabi, Zaid Alsafi, Niamh O'neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). *International journal of surgery*, 76:71–76, 2020.

[8] Michael Worobey, Joshua I Levy, Lorena Malpica Serrano, Alexander Crits-Christoph, Jonathan E Pekar, Stephen A Goldstein, Angela L Rasmussen, Moritz UG Kraemer, Chris Newman, Marion PG Koopmans, et al. The huanan seafood wholesale market in wuhan was the early epicenter of the covid-19 pandemic. *Science*, 377(6609):951–959, 2022.

[9] Worldometer. Covid-19 coronavirus pandemic. https://www.worldometers.info/coronavirus/. Accessed: 2024-04-13.

[10] Mihaela Van der Schaar, Ahmed M Alaa, Andres Floto, Alexander Gimson, Stefan Scholtes, Angela Wood, Eoin McKinney, Daniel Jarrett, Pietro Lio, and Ari Ercole. How artificial intelligence and machine learning can help healthcare systems respond to covid-19. *Machine Learning*, 110:1–14, 2021.

[11] Laure Wynants, Ben Van Calster, Gary S Collins, Richard D Riley, Georg Heinze, Ewoud Schuit, Elena Albu, Banafsheh Arshi, Vanesa Bellou, Marc MJ Bonten, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*, 369, 2020.

[12] Chansik An, Hyunsun Lim, Dong-Wook Kim, Jung Hyun Chang, Yoon Jung Choi, and Seong Woo Kim. Machine learning prediction for mortality of patients diagnosed with covid-19: a nationwide korean cohort study. *Scientific reports*, 10(1):18716, 2020.

[13] Kee Yuan Ngiam and Wei Khor. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5):e262–e273, 2019.

[14] Gloria Phillips-Wren and Nikhil Ichalkaranje. *Intelligent decision making: An AI-based approach*, volume 97. Springer Science & Business Media, 2008.

[15] Zhihai Chen, Wen Xie, Ziruo Ge, Yajie Wang, Hong Zhao, Jingjing Wang, Yanli Xu, Wei Zhang, Meihua Song, Shuping Cui, et al. Reactivation of sars-cov-2 infection following recovery from covid-19. *Journal of infection and public health*, 14(5):620–627, 2021.

[16] Roberto Vega, Leonardo Flores, and Russell Greiner. Simlr: Machine learning inside the sir model for covid-19 forecasting. *Forecasting*, 4(1):72–94, 2022.

[17] Samuel Lalmuanawma, Jamal Hussain, and Lalrinfela Chhakchhuak. Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review. *Chaos, Solitons & Fractals*, 139:110059, 2020.

[18] Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B Munroe, Bina Joe, and Xi Cheng. Artificial intelligence and machine learning to fight covid-19, 2020.

[19] Yazeed Zoabi, Shira Deri-Rozov, and Noam Shomron. Machine learning-based prediction of covid-19 diagnosis based on symptoms. *npj digital medicine*, 4(1):1–5, 2021.

[20] Milind Yadav, Murukessan Perumal, and M Srinivas. Analysis on novel coronavirus (covid-19) using machine learning methods. *Chaos, Solitons & Fractals*, 139:110050, 2020.

[21] Farai Nyabadza, Faraimunashe Chirove, CW Chukwu, and Maria Vivien Visaya. Modelling the potential impact of social distancing on the covid-19 epidemic in south africa. *Computational and mathematical methods in medicine*, 2020(1):5379278, 2020.

[22] SP Gatyeni, CW Chukwu, Faraimunashe Chirove, F Nyabadza, et al. Application of optimal control to the dynamics of covid-19 disease in south africa. *Scientific African*, 16:e01268, 2022.

[23] Christina J Edholm, Benjamin Levy, Lee Spence, Folashade B Agusto, Faraimunashe Chirove, C Williams Chukwu, David Goldsman, Moatlhodi Kgosimore, Innocent Maposa, KA Jane White, et al. A vaccination model for covid-19 in gauteng, south africa. *Infectious Disease Modelling*, 7(3):333–345, 2022.

[24] Mohammad Sadeq Mottaqi, Fatemeh Mohammadipanah, and Hedieh Sajedi. Contribution of machine learning approaches in response to sars-cov-2 infection. *Informatics in Medicine Unlocked*, 23:100526, 2021.