Time Series Analysis of COVID – 19 in India

Venkatakrishnan R
venkatkrishnan2003@gmail.com
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham, Chennai

Siva Jyothi Nath Reddy B
ss2619339@gmail.com
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham, Chennai

Sarthak Yadav
sarthaknyadav@gmail.com
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham, Chennai

Prasanna Kumar R
r_prasannakumar@ch.amrita.edu
Associate Professor,
Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham, Chennai

Abstract— **The goal of the study is to use mathematical concepts like Ordinary Differential Equations to model Epidemiology (i.e. disease control). The paper's main goal is to assist the nation's healthcare business in surviving and assisting the citizens of the country in fighting the pandemic. The CovsirPhy[1] python module was used to analyse the modelling. The SIR model is a mathematical model of disease. An outbreak occurs when the number of people infected with a disease grows in a population. Using various parameters such as rate of recovery, time elapsed, total population, effective contact rate, and so on, to conduct a pandemic analysis and provide acceptable results and future projections. We created a comparison graph for the First Phase of the COVID-19 pandemic, which runs from March 19 to March 30, 2020. The SIRF model was then drawn for the Second Phase of the pandemic, which runs from May 12 to May 24, 2021.**

I. Introduction

The extreme acute respiratory coronavirus 2 condition, often known as SARS-COV-2, causes Coronavirus Disease 2019 (COVID-19). Corona virus gets its name from the Latin word corona, which means "crown." The name comes from the electron microscopic analysis of the virion's structure, which has a distinctive appearance. The virus's surface is covered in viral spike protein. Droplets of saliva or nasal discharge from an infected person are the agents that disseminate the disease. The SIR model is a mathematical model of disease. An outbreak occurs when the number of people infected with a disease grows in a population. Obtaining an analysis of the pandemic and providing acceptable results and future projections using various parameters such as rate of recovery, time elapsed, total population, effective contact rate, and so on. We created a comparison graph for the COVID-19 pandemic's First Phase, which runs from March 19 to March 30, 2020. Then, for the Second Phase of the pandemic, from May 12 to May 24, 2021, we drew the SIRF model.

II. Prediction Models

COVID –19 looks to be more virulent than other types of corona viruses, hence forecasting the number of cases appears to be crucial for the country's health sector. Many methods for predicting the pandemic's spread have been proposed. The various models proposed can be grouped and categorized as 1) mechanistic models based on SIR (Susceptible, infected, and recovered state), 2) time series models like Grey Model, Markov chain model, ARIMA, and 3) agent type models. In each category, there are various types of techniques. With respect to the SIR model, we primarily focus on deterministic models involving differential equations in this study.

III. Literature Survey

In 2021, Khandaker Tabin Hasan et al introduced the 4P model ([4]), which is based on four probabilities (4P) that have been found to be true for all countries included. These include the probability of the state's control over the situation P(G), the probability of citizens' compliance with imposed laws and health regulations P(P), the probability of a person becoming infected after being exposed P(I), and the moving average of the probability of getting positive cases out of total samples tested P. (T).

Mashael Khayyat et al proposed the Prophet Model [5] for COVID-19 Outbreak Prediction. It utilizes Time Series Analysis with Facebook Prophet Model to predict Coronavirus illness (COIVD-19). Facebook Prophet Model is an open-source toolkit built for creating forecasts for time series datasets and using forecasting tools accessible in Python. On a daily or weekly basis, we can monitor and predict the spread of the coronavirus pandemic.

The PDE ([6])based modelling of COVID-19 infection was given by Sashikumaar Ganesan and Deepak Subramani. This model is based on a population balance equation with a high number of dimensions. , the proposed model predicts the distribution of infected people across the region, their age, the day they were infected, and the severity of their infection with time. Furthermore, the newly developed model takes into account immunity, pre-medical history, efficient treatment, infected population point-to-point movement (e.g., by air, train, etc.), interactivity (community spread), hygiene, and population social distance.

[7] Mohd Zufaezal Che Azemin used a deep learning prediction model supported the ResNet-101 convolutional neural specification that was trained to discover anomalies in chest X-ray footage before being retrained to discriminate things from 1,000,000 photos. In terms of space underneath the receiver operational curve, sensitivity, specificity, and accuracy, the model scored zero.82, 77.3 percent, 71.8 percent, and 71.9 percent, severally. [8] Janik Schuttler et al utilized a Gauss model (GM) as an easy, analytically tractable model to develop coronavirus epidemic forecasts. A gramme may be a map from time to a bulging mathematician perform to explain deaths per day and nation. They investigated the model's
predictions victimization existing information from twenty five countries throughout the primary corona pandemic wave, that ar even by the sigmoid nature of an epidemic, i.e. initial exponential unfold to ultimate saturation, associated an agent-based model. The exponent daily mortality iatrogenic by the coronavirus infection is well delineated by a quadratic perform in time (Covid-19).
[9] The Kalman Filter, by Koushlendra Kumar Singh et al, is used to estimate SARS-future Cov-2's spread, and it produces good results on the data evaluated. The Random Forest inferences emphasise importance, while Pearson Correlation yields many similarities and few differences, indicating that these techniques are successful in identifying the many contributing components.

[10] The method developed by Hongwei Zhao et al. for foretelling future COVID-19 cases entails: 1) modelling discovered incidence cases employing a statistical distribution for the daily incidence variety and a gamma distribution for the series interval; 2) estimating the effective replica variety underneath the idea that its price remains constant over a brief time interval; and 3) drawing future incidence cases from their posterior distributions, guaranteeing that the effective replica variety r remains constant.

[11] The ARIMA and NAR predictions were projected by Farhan Mahound Khan and Rajiv Gupta. The ARIMA approach is additionally referred to as the Box-Jenkins methodology. to suit a mixed ARIMA model to a given set of information, the Box- Jenkins methodology is employed. ARIMA (Auto-Regressive Integrated Moving Average) may be a category of models accustomed forecast future values by explaining a given statistic supported its past values, i.e., its own lags and lagged forecast errors. associate ARIMA model is outlined by 3 terms: p, d, and q.

IV. Methodology

A. *Computing Growth Factor*

Here, C denotes the number of confirmed cases. Using the growth factor, we can determine whether a particular country has an outbreak or a halt to the pandemic:

$$\text{Growth factor} = \frac{\Delta C(t)}{\Delta C(t-1)}$$

- In Out breaking case: growth factor will be <1 for the last 7 days
- In Stopping case: growth factor will be >1 for the last 7 days

The SIR model is a fundamental mathematical model of disease. An outbreak occurs when the number of people infected with a disease increases in a population.

B. *SIR Modelling*

S.I.R. is an abbreviation for susceptible. These are people who have not yet been infected with the virus. They are not immune to it, however, and may become infected in the future. Infected or infectious is denoted by the letter I. These are people who have the disease and can pass it on to others. The letter R stands for retrieved. These are people who have recovered from the disease and have developed immunity, so they can no longer be infected with it.

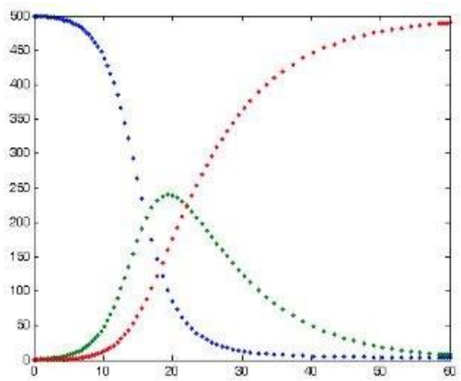

Fig 1. SIR modelling. The three dotted lines in blue, green, and red represent the S, I, and R, respectively. The color blue represents the number of vulnerable people. At the start of the outbreak, the number of susceptible (blue) people decreases while the number of infected (green) people increases. Over time, the number of recovered (red) people grows. In this model, after the outbreak has passed, everyone has been infected and healed. This is not

always true; some people who are susceptible to infection can remain uninfected.

$$\frac{dS}{dT} = -N^{-1}\beta\ SI \qquad (1)$$

$$\frac{dI}{dT} = N^{-1}\beta SI - (\gamma)I \qquad (2)$$

$$\frac{dR}{dT} = \gamma I \quad (3)$$

The term ds/dt refers to the change in the rate of susceptible, which is directly proportional to the product of (the effective contact rate [1/min]), the number of susceptible (S), N(total population), T(elapsed time), and the number of infected (I)

The term dR/dt corresponds to the rate of recovery(R) i.e which is directly proportional to the product of γ (*The effective fatality rate*[1/*min*]).

The term dI/dt corresponds to the change in the rate of infection which is proportional to the difference of rate of susceptible and rate of recovery.

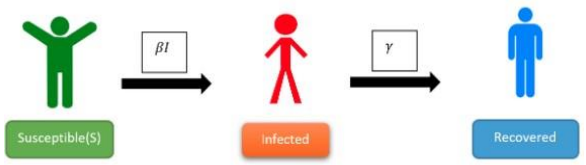

Fig. 2 :Pictorial representation of SIR modelling with factors such as β(*Effective contact rate*[1/*min*]), *I* (*infected*)*and* γ(*Recovery rate*[1/*min*])

C. *SIR-F Modelling*

With a slight modification in the SIR model we have SIR-F model. In this model S represents Susceptible, S* represents Confirmed and Uncategorized, I represents Confirmed and Categorized as Infected, R represents Confirmed and Categorized as Recovered, and F represents Confirmed and Categorized as Fatal.

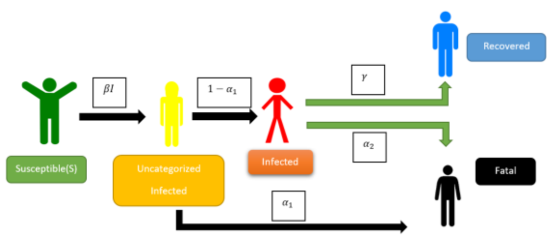

Fig. 3 :Pictorial representation of SIRF modelling with factors such as β(*Effective contact rate*[1/*min*]), *I* (*infected*) *and* γ (*Recovery rate*[1/*min*]), α_1 represents the direct fatality probability of S*, α_2 represents the mortality rate of infected cases,

$$\frac{dS}{dT} = -N^{-1}\beta\ SI \qquad (4)$$

$$\frac{dI}{dT} = N^{-1}(1-\alpha_1)\beta SI - (\gamma + \alpha_2)I \qquad (5)$$

$$\frac{dR}{dT} = \gamma I \qquad (6)$$

The term ds/dt corresponds to the change in the rate of susceptible which is directly proportional to the product of β (The effective contact rate) , number of susceptible (S), N(total population), T(time elapsed) and the number of infected(I).

The term dI/dt corresponds to the change in the rate of infection which is proportional to the difference of rate of susceptible with α_1 represents the direct fatality probability of S and rate of recovery with α_2 represents the mortality rate of infected cases,

The term dR/dt corresponds to the rate of recovery(R) i.e which is directly proportional to the product of γ (*The effective fatality rate*[1/*min*]).

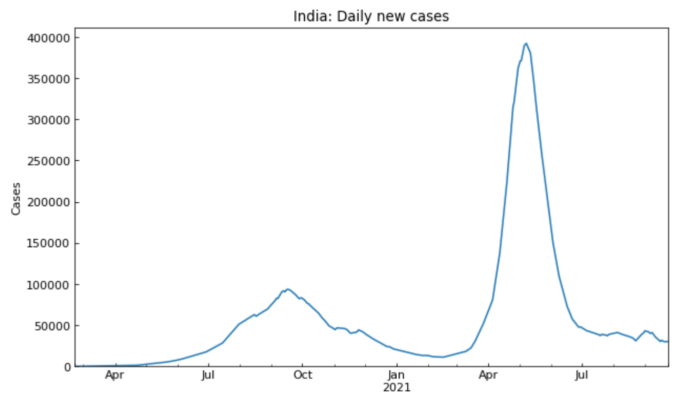

Fig. 4 The Blue line shows the no of cases of infected people of covid-19 .we can observe two peaks in the curve which represents the 1st wave and 2nd wave over time.

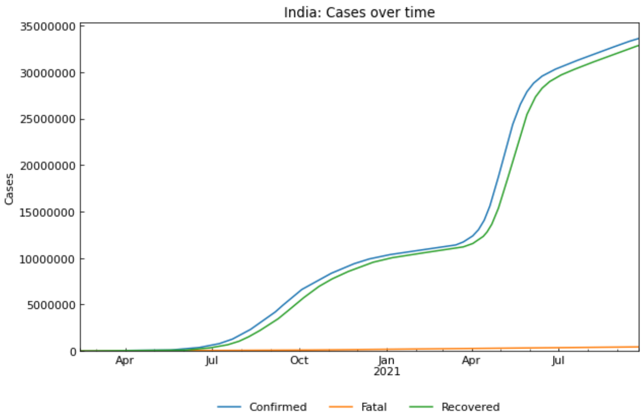

Fig.5 The Blue line represents the infected cases , orange line represents the fatal cases , green line represents the recovered cases from COVID-19 over waves.

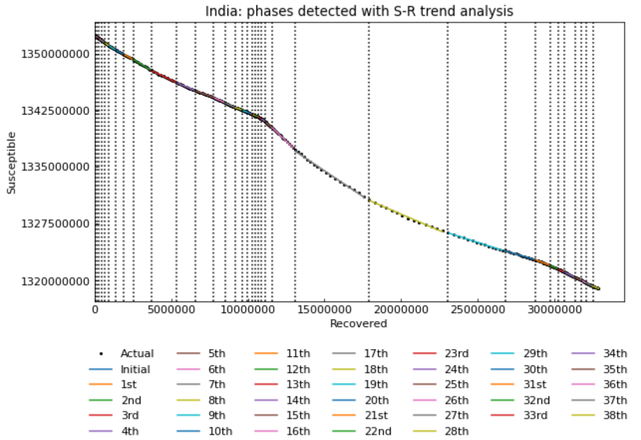

Fig. 6.A S-R trend analysis with respect to various phases in Figure 6.B

The inference made from the above plot of Covid Cases in India, the first wave started from 13 Feb 2020 and reached its peak at end of September. The Second wave started from end of march and peak was reached at May mid. The primary comparison is such that the intensity of the virus is more in the second wave that can be said because the number of days in order to reach the peak was very less in comparison. With this conclusion we can estimate that the third wave would be even more intense.

D. SR Trend Analysis

With the sirf model proposed, we plot the susceptible versus the recovered ratio. The relationship can between S and R can be formulated as below,

$$\frac{dS}{dT} = -\frac{\beta}{N\gamma} * R \qquad (7)$$

By solving the corresponding equation we get a logarithmic relating expression between susceptible and the recovered

$$logS(R) = -\frac{\beta}{N\gamma} * R + logN \quad (8)$$

The term dS/dR refers to the change in the rate of susceptible versus the rate of recovered, which is directly proportional to the product of (the effective contact rate), N(total population), (recovery rate), and R. (number of recovered cases)

0th	13-Feb-20	30-Mar-20
1st	31-Mar-20	16-Apr-20
2nd	17-Apr-20	03-May-20
3rd	04-May-20	20-May-20
4th	21-May-20	08-Jun-20
5th	09-Jun-20	26-Jun-20
21st	20-Jan-21	01-Feb-21
22nd	02-Feb-21	18-Feb-21
23rd	19-Feb-21	07-Mar-21
34rd	16-Jul-21	25-Jul-21
35th	26-Jul-21	12-Aug-21
36th	13-Aug-21	22-Aug-21
37th	23-Aug-21	01-Sep-21
38th	02-Sep-21	24-Sep-21

Fig. 6.B Different Phase calculation of the COVID-19 wave

Considering a case of lockdown, first we have to estimate the number of people who are susceptible meeting the infected people. That can be determined by considering the population pyramid of the country. We framed the following results after making a tabular form

	Age_first	Age_last	Period_of_life	School	Office	Others	Age	Population	Portion
0	0	2	nursery	5	0	0	2	70069092	0.052472
1	3	5	nursery school	5	0	1	5	70702729	0.052946
2	6	10	elementary school	6	0	1	10	121356985	0.090879
3	11	13	middle school	6	0	1	13	76190943	0.057056
4	14	18	high school	6	0	1	18	125948517	0.094318
5	19	25	university/work	3	3	0	25	170272038	0.127510
6	26	35	work	0	6	1	35	224676012	0.168251
7	36	45	work	0	5	1	45	184775152	0.138371
8	46	55	work	0	5	1	55	141199301	0.105738
9	56	65	work	0	5	1	65	101392985	0.075929
10	66	75	retired	0	0	4	75	32191471	0.024107
11	76	85	retired	0	0	3	85	13549164	0.010146
12	86	95	retired	0	0	2	95	3040240	0.002277

Fig. 7 Population pyramid of India

The number of susceptible people are calculated by considering the activity of the people and the portion of the population. The result was 6.27 which is huge number so considering lockdown was important. The restriction in lockdown for office were considered in two scenario One with 30% restriction and 50% restriction. The schools were closed in both scenario.

	Age_first	Age_last	Period_of_life	School	Office	Others	Age	Population	Portio
0	0	2	nursery	0	0.0	1	2	70069092	0.05247
1	3	5	nursery school	0	0.0	2	5	70702729	0.05294
2	6	10	elementary school	0	0.0	2	10	121356985	0.09087
3	11	13	middle school	0	0.0	2	13	76190943	0.05705
4	14	18	high school	0	0.0	2	18	125948517	0.09431
5	19	25	university/work	0	0.9	1	25	170272038	0.12751
6	26	35	work	0	1.8	2	35	224676012	0.16825
7	36	45	work	0	1.5	2	45	184775152	0.13837
8	46	55	work	0	1.5	2	55	141199301	0.10573
9	56	65	work	0	1.5	2	65	101392985	0.07592
10	66	75	retired	0	0.0	4	75	32191471	0.02410
11	76	85	retired	0	0.0	3	85	13549164	0.01014
12	86	95	retired	0	0.0	2	95	3040240	0.00227

Fig 8: Population pyramid of the country with 70% restrictions

The result of the first scenario is 2.77 people meet per day with each other.

	Age_first	Age_last	Period_of_life	School	Office	Others	Age	Population	Portion
0	0	2	nursery	0	0.0	1	2	70069092	0.052472
1	3	5	nursery school	0	0.0	2	5	70702729	0.052946
2	6	10	elementary school	0	0.0	2	10	121356985	0.090879
3	11	13	middle school	0	0.0	2	13	76190943	0.057056
4	14	18	high school	0	0.0	2	18	125948517	0.094318
5	19	25	university/work	0	1.5	1	25	170272038	0.127510
6	26	35	work	0	3.0	2	35	224676012	0.168251
7	36	45	work	0	2.5	2	45	184775152	0.138371
8	46	55	work	0	2.5	2	55	141199301	0.105738
9	56	65	work	0	2.5	2	65	101392985	0.075929
10	66	75	retired	0	0.0	4	75	32191471	0.024107
11	76	85	retired	0	0.0	3	85	13549164	0.010146
12	86	95	retired	0	0.0	2	95	3040240	0.002277

Fig 9: Population pyramid of the country with 50% restrictions

The result of the second scenario is 3.377 people meet per day with each other

V. Simulating and Predicting the End of Second Wave

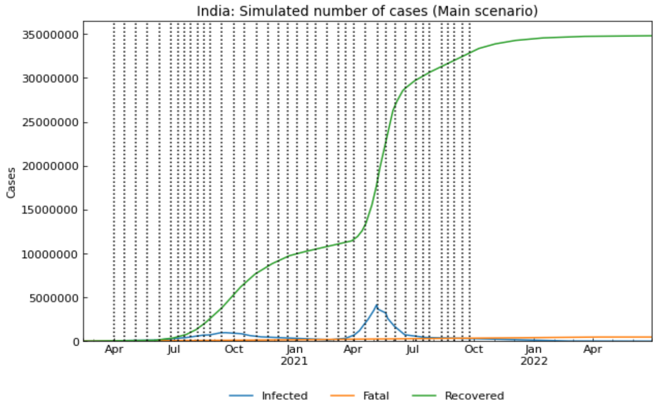

Fig.10 :Predicting the coming up days. The Blue line represents the infected cases ,orange line represents the fatal cases , green line represents the recovered cases from COVID-19 over waves.

Date	Infected	Fatal	Recovered
25/09/2021	290036	448182	32953608
06/10/2021	239132	452110	33282570
17/11/2021	114098	461739	34088945
11/01/2022	43378	467171	34543924
12/01/2022	42618	467229	34548810
11/04/2022	8896	469817	34765526
12/04/2022	8740	469829	34766523
13/04/2022	8588	469840	34767503
29/05/2022	3814	470206	34798169
30/05/2022	3748	470212	34798594
31/05/2022	3683	470217	34799013
02/06/2022	3556	470226	34799827
03/06/2022	3494	470231	34800224
15/07/2022	1666	470371	34811969
16/07/2022	1637	470373	34812156

Fig.11 Tabular form of comparison and prediction of Infected, Fatal and Recovered.

VI. Results

Model comparison of first phase and last phase

Phase	Start	End
1st	19-Mar-20	30-Mar-20

Phase	Start	End
38th	14-Sep-21	24-Sep-21

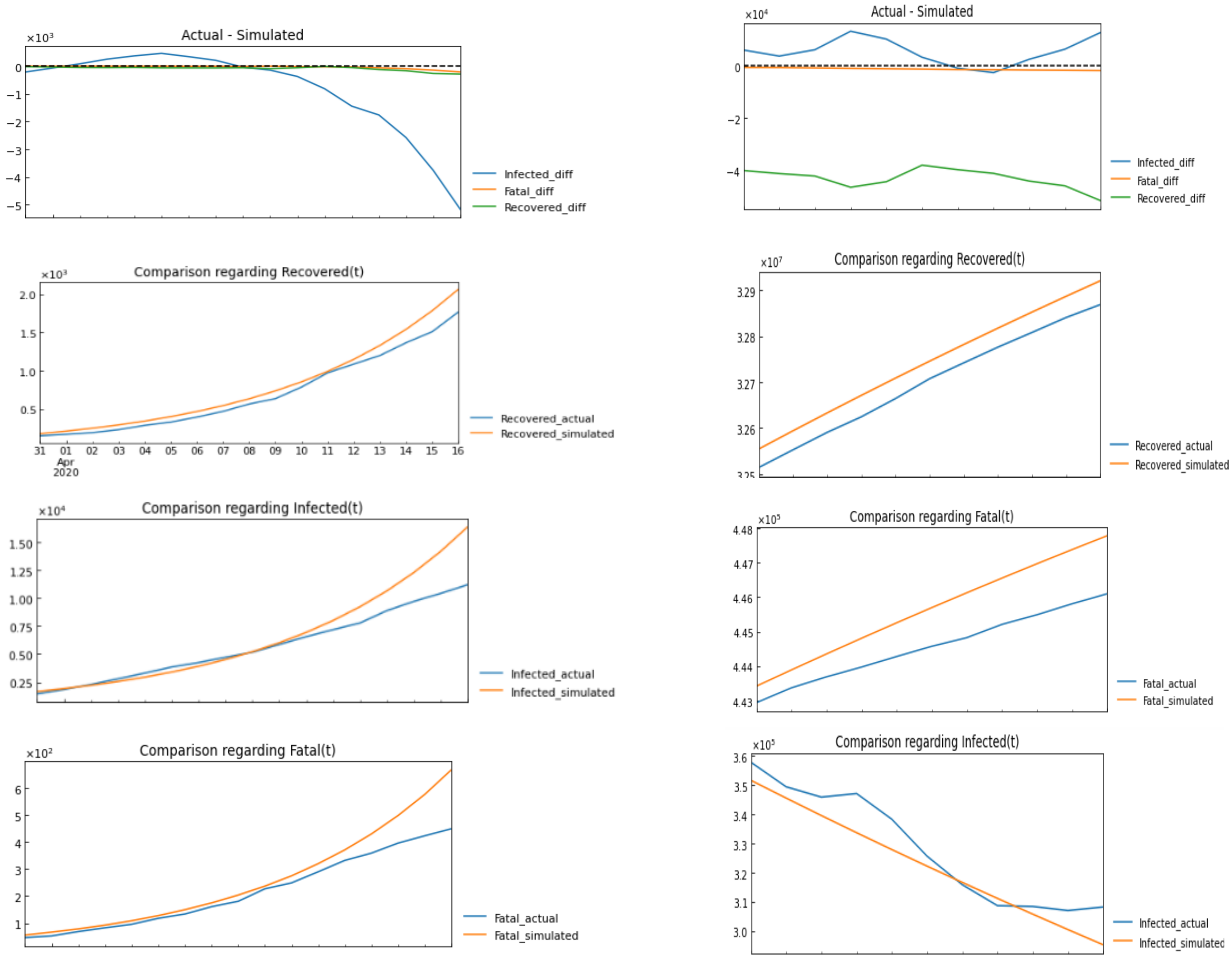

Fig 12 SIR-F Model Analysis comparison of two phases such as the 1st phase and the 38th phase and prediction of Infected, Fatal and Recovered

VIII. CONCLUSION

The need of the disease modelling is crucial in the future. The SIRF model provides a strong premise in decision making by demonstrating the causality, recovery and infectious of the epidemic over which the state alongside its citizens has control.

.

References

[1] Covid-19 Data with SIR Model. https://www.kaggle.com/lisphilar/covid-19-data-with-sir-model.

[2] Covsirphy Repository. https://github.com/lisphilar/covid19-sir.

[3] https://github.com/tdrvlad/covidsirmodelling.git

[4] Hasan, K.T., Rahman, M.M., Ahmmed, M.M. et al. 4P Model for Dynamic Prediction of COVID-19: a Statistical and Machine Learning Approach. Cogn Comput (2021). https://doi.org/10.1007/s12559-020-09786-6

[5] M. Khayyat, K. Laabidi, N. Almalki, and M. Al-zahrani, Time Series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction. Computers, Materials Continua, 67 (2021) 3781–3793. https://doi.org/10.32604/cmc.2021.014918.

[6] Ganesan, S., Subramani, D. Spatio-temporal predictive modeling framework for infectious disease spread. Sci Rep 11, 6741 (2021). https://doi.org/10.1038/s41598-021-86084-7

[7] Mohd Zulfaezal Che Azemin , Radhiana Hassan , Mohd Izzuddin Mohd Tamrin , Mohd Adli Md Ali . COVID-19 Deep Learning Prediction Model Using Publicly Available Radiologist-Adjudicated Chest X-Ray Images as Training Data: Preliminary Findings

[8] Schu¨ttler, J.; Schlickeiser, R.; Schlickeiser, F.; Kr¨oger, M. Covid-19 Predictions Using a Gauss Model, Based on Data from April 2. Physics 2020, 2, 197-212. https://doi.org/10.3390/physics2020013

[9] Singh, K.K., Kumar, S., Dixit, P. et al. Kalman filter based short term prediction model for COVID-19 spread. Appl Intell 51, 2714–2726 (2021).

[10] Hongwei Zhao ,Naveed N. Merchant ,Alyssa McNulty ,Tiffany A. Radcliff ,Murray J. Cote ,Rebecca S. B. Fischer ,Huiyan Sang ,Marcia G. Ory. COVID-19: Short term prediction model using daily incidence data.

[11] Farhan Mohammad Khan, Rajiv Gupta, ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India, Journal of Safety Science and Resilience, Volume 1, Issue 1, 2020