Noname manuscript No.
(will be inserted by the editor)

A Time-Dependent SEIRD Model for Forecasting the COVID-19 Transmission Dynamics

Taarak Rapolu [1], · **Brahmani Nutakki**[2], ·
T.Sobha Rani[3], · **S. Durga Bhavani**[4]

08 June 2020

Abstract The spread of a disease caused by a virus can happen through human to human contact or could be from the environment. A mathematical model could be used to capture the dynamics of the disease spread to estimate the infections, recoveries, and Deceased that may result from the disease. An estimation is crucial to make policy decisions and for the alerts for the medical emergencies that may arise. Many epidemiological models are being used to make such an estimation. One major factor that is important in the forecasts using the models is the dynamic nature of the disease spread. Unless we can come up with a way of estimating the parameters that guide this dynamic spread, the models may not give accurate forecasts. In this work, using the SEIRD model, attempts are made to forecast Infected, Recovered and Deceased rates of COVID-19 up to a week using an *incremental* approach. A method of estimating the parameters of the model is also discussed thoroughly in this work. The model is evaluated using the data taken from COVID-19 India tracker [2], a crowd-sourced platform for India. The model is tested with the whole country as well as all the states and districts. The results of all the states and districts obtained from our model can be seen in [12]. The forecasts are reasonable which can help the governments in planning for emergencies such as ICU requirements, PPEs, hospitalizations, and so on as the infection is going to be prevalent for some time to come.

Keywords Time-Dependent · SEIRD · data-driven

1 Introduction

Novel Coronavirus has become a pandemic within no time from the time of its detection in Wuhan, a province of China. This has been declared as a pandemic by WHO resulting in around 6,876,647 cases worldwide, by 6^{th} of June[3]. Around 237,754 were affected in India alone. With 6,650 reported Deaths, the cases are rapidly rising,

[1,2,3,4]School of Computer and Information Sciences, University of Hyderabad,Hyderabad, India

where Maharashtra is leading the tally. Its rapid progress has necessitated the need to come up with models to model the spread of the virus under different conditions like lockdown, hotspots, and migration of people across the places, and so on. The outbreak of novel coronavirus Covid19 and the ensuing utter chaos and the utter uncertainty caused by the pandemic in the entire world is unprecedented. More than ever before, it emphasizes the need for robust mathematical models that can guide policies to control the spread of infection and help in planning the hospital requirements such as PPEs, ventilators, etc [10].

In the literature several epidemiological models such as Susceptible, Infected, Recovered(SIR), Susceptible, Exposed, Infected, Recovered(SEIR) and Susceptible, Exposed, Infected, Recovered and Deceased (SEIRD), etc have been proposed to model the virus spreads like H1N1, SARS, Ebola, and others. *EpiModel* is a very useful software package, developed in 'R' language, that allows simulation of compartmental models, stochastic individual contact models, and the more recent network models [6].

2 Existing Models

A few existing pandemic models, from which the current model is derived are discussed here. The first model used to model the pandemic virus spread is the SIR model.

2.1 SIR

One of the prediction models available is the SIR or the Susceptible, Infected, Recovered Model. This popular epidemic model considers a closed population. It initially considers a small part of the population as infected. This small percentage is considered to infect R_0 others, where R_0 is the Basic Reproduction Rate[1]. The SIR model can be described as

$$\frac{\partial S}{\partial t} = -\beta \frac{SI}{N}$$

$$\frac{\partial I}{\partial t} = \beta \frac{SI}{N} - \gamma I$$

$$\frac{\partial R}{\partial t} = \gamma I$$

Here S, I, R stand for Susceptible, Infected and Recovered respectively. β is the Transmission rate and γ is the Recovery rate.

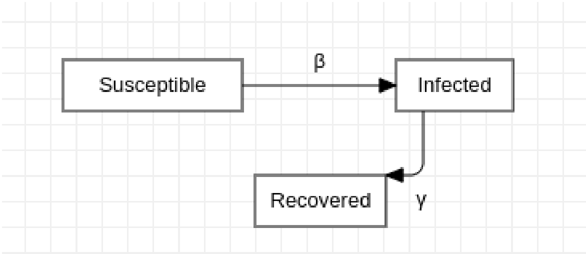

Fig. 1: **Fig.1** SIR Model

2.2 SEIR

The SIR model discussed here does not consider the percentage of the population who are exposed to the disease, but do not show any symptoms. When the incubation time i.e, the time elapsed before developing symptoms is significant, the SIR model will not be able to capture it. This leads to the SEIR model- Susceptible, Exposed, Infected, Recovered. The model is similar to SIR except that there is a transition from S to E instead of S to I. And the exposed percentage can also infect the Susceptibles. In a closed population, the SEIR model can be represented as

$$\frac{\partial S}{\partial t} = -\beta \frac{SI}{N}$$

$$\frac{\partial E}{\partial t} = \beta \frac{SI}{N} - \alpha E$$

$$\frac{\partial I}{\partial t} = \alpha E - \gamma I$$

$$\frac{\partial R}{\partial t} = \gamma I$$

Here, β is the Transmission rate. α is the Incubation rate (Transition rate from E to I) while γ is the Recovery rate.

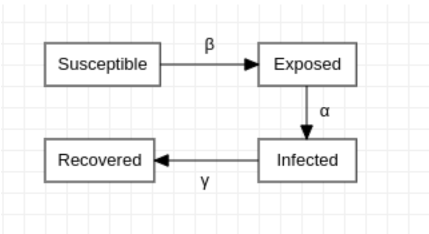

Fig. 2: **Fig.2** SEIR Model

Many extensions of the compartmental model have been proposed. These include extra compartments to denote, for example, the contaminated environmental reservoir [7], the eight compartment model of Tang et al.[9] to include quarantined individuals and hospitalization, etc. Most of the papers in the literature consider the SEIR model with a deterministic approach by fixing the parameters to model the spread of infection.[8][11].

Yang and Wang consider the dynamic nature of the tuning parameters themselves. They consider the time-dependent parameters to model the spread of COVID 19 virus in Wuhan extending the SEIR model [7]. They have concluded that the disease is an endemic process and requires a long term plan to spread of the virus.

The model of B.Tang et al. [4] is one of the few which considers the parameters including the rate of transmission, contact rate, recovery rate as functions of time and simulate the model in order to predict the size of the infected population. They use the Markov Chain Monte Carlo (MCMC) procedure to fit the model to the data.

We observe that one of the main challenges in adopting the compartmental models lies in tuning the number of parameters involved in the model. The work in the literature fixes the parameters based on the indicators given by epidemiological experts in the scenario. The emphasis of the current work is to estimate the parameters in a dynamic manner.

3 Model formulation and Analysis

- Basic SEIR Model
  We initially ran our data against a basic SEIR model. It was observed that the results are not as accurate as expected. We were also not able to fit the Recovered rates as expected. So we extended our model to include parameter estimation- an optimized concept to estimate the parameters as per the data rather than assuming them.
- Approaches to Parameter Estimation
  - Grid Search
    In this model, a Grid Search is used to estimate parameters. A broad range is assigned to each of the parameters. The model then tunes the parameters to get possible values that fit the data.
    This model is computationally expensive. It takes about an hour and a half to run it on Google Colab. Once the range of parameters is narrowed down, it forecasts the rates which are more accurate than the previous model.
  - Walk forward with Grid Search
    After working on different models, it is evident that the parameters are non-stationary i.e, they change constantly. This model implements the Walk forward approach. Until the last model, the parameters are estimated for the training set as a whole. In this model, they are estimated incrementally one day at a time. The parameters obtained for the previous day are used to estimate the current day parameters.
    Though this model is relatively more accurate than the previous versions, it is extremely expensive in terms of computation. Efforts are made to extend

this model using Parallel computation to no avail. It took about 3-4 hours on Google Colab to run this model with no results.

4 Time-Dependent SEIRD Model

In order to optimize the model, different approaches to the SEIR model are considered. This leads to the implementation of a time-dependent SEIRD model [4].
This approach resulted in accurate forecasts of Infected, Recovered, and Deceased rates for a week. Run time is exceptionally low, one minute at the most. This model can be represented as follows:

$$\Delta S = -\frac{\beta(t)S(t)I(t)}{N} \tag{1}$$

$$\Delta E = \frac{\beta(t)S(t)I(t)}{N} - \alpha(t)E(t) \tag{2}$$

$$\Delta I = \alpha(t)E(t) - \gamma(t)I(t) - \delta(t)I(t) \tag{3}$$

$$\Delta R = \gamma(t)I(t) \tag{4}$$

$$\Delta D = \delta(t)I(t) \tag{5}$$

From (4), we have

$$\gamma(t) = \frac{\Delta R}{I(t)} \tag{6}$$

From (5), we have

$$\delta(t) = \frac{\Delta D}{I(t)} \tag{7}$$

Using (6) and (7) in (3) yields

$$\alpha(t) = \frac{\Delta I + \Delta R + \Delta D}{E(t)} \tag{8}$$

Using (8) in (2) yields

$$\beta(t) = \frac{(\Delta E + \Delta I + \Delta R + \Delta D)N}{S(t)I(t)} \tag{9}$$

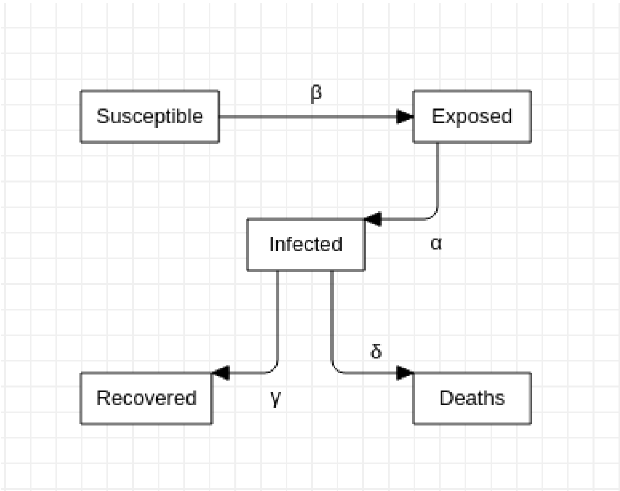

Fig. 3: **Fig.3** SEIRD Model

4.1 Data requirements and format to run the model

This model takes a *.csv* file with Cumulative Confirmed, Recovered, and Deceased values. It forecasts the results based on the parameters of the last five days of the test data. It takes the data file and 'N', the population of the area as input. Shown in Figure 4, is the format of data that is being used.

Day	Date	Confirmed	Recovered	Deaths
1	2020-03-24	571	40	10
2	2020-03-25	657	43	11
3	2020-03-26	730	50	16
4	2020-03-27	883	75	19
5	2020-03-28	1019	85	24
6	2020-03-29	1139	102	27

Fig. 4: **Fig.4** Data format

4.2 Model

In our model, the population is classified into 5 categories: the Susceptible, the Exposed, the Infected, the Recovered, and the Deceased. Parameters are estimated on a day to day basis. The data until the previous day and the current day is used to calculate the parameters of the previous day. Then the value for the Exposed population that is calculated is passed on to calculating the parameters for the next day.

This model is run on the data collected from 24^{th} of March onward for states when the first lockdown is implemented in India [2] and from 9^{th} March for Lombardia, Italy [14] and 9^{th} March for Moscow, Russia [15]. For districts in India, the data is available from 24^{th} April in [2]. This model requires a minimum training of 18 days from the start of lockdown. It is stated that the mean of Incubation Period is 5-7 days [7]. We consider it as 6 days and use the last 6 days of the data to get 5 parameters. We then use these 5 parameters to forecast the Infected and Deceased values for the upcoming 7 days. Out of these 5 sets of 7 days forecasts each, the median set based on the Infected cases of each set is selected as the final forecats. The reason to consider the median (to avoid outlayer) based on Infected cases rather than Deceased cases is that the count and the magnitude of change of Infected cases is huge than Deceased cases. It is also observed that when Deceased cases are used as median, the accuracy of the forecasted Infected cases is quite less.

4.3 Algorithm

The code for the model is put up at the link given in [5]. Here, s[], e[], i[], r[], d[] are arrays to store calculated susceptible, exposed, infected, recovered and Deceased values
alpha[], beta[], gamma[], delta[] –arrays to store calculated alpha, beta, gamma and delta values
preds are all the predictions stored in an array pred_values are stored in a stack that contains s, e, i, r, d array values in seird function.
start_date is starting date of the data taken from data.csv
T number of days in the training data taken from data.csv
incub_period is the Incubation Period

Algorithm 1 Incremental SEIRD Model

Input: N (population), data.csv
Output: Forecasts for Infected and Deceased cases for 5 days from the date training ends
$N \leftarrow population$
$Train \leftarrow data.csv$
$start\_date \leftarrow Train.date$
$T \leftarrow Train.days$
$incub\_period = 5$
for x = 1 to T **do**
  $i[x] = Train.c[x] - Train.r[x] - Train.d[x]$
  $r[x] = Train.r[x]$
  $d[x] = Train.d[x]$
end for
for t = 1 to T **do**
  parameter_estimation(t)
end for
prediction()
end

---

parameter_estimation(k)
if k = 1 **then**
  $alpha[k] = 1/incub\_period$
  $gamma[k] = (r[k+1] - r[k])/i[k]$
  $gamma[k+1] = (r[k+2] - r[k+1])/i[k+1]$
  $delta[k] = (d[k+1] - d[k])/i[k]$
  $delta[k+1] = (d[k+2] - d[k+1])/i[k+1]$
  $e[k] = (i[k+1] - ((1 - gamma[k] - delta[k]) * i[k]))/alpha[k]$
  $e[k+1] = (i[k+2] - ((1 - gamma[k+1] - delta[k+1]) * i[k+1]))/alpha[k]$
  $s[k] = N - e[k] - i[k] - r[k] - d[k]$
else
  $alpha[k] = ((i[k+1] - i[k]) + (r[k+1] - r[k]) + (d[k+1] - d[k]))/e[k]$
end if
$beta[k] = (((e[k+1] - e[k]) + (i[k+1] - i[k]) + (r[k+1] - r[k]) + (d[k+1] - d[k])) * N)/(s[k] * i[k]))$
$gamma[k] = (r[k+1] - r[k])/i[k]$
$delta[k] = (d[k+1] - d[k])/i[k]$
seird(alpha[k], beta[k], gamma[k], delta[k], k, k+1)
end

---

seird(alpha, beta, gamma, delta, k, t)
$s[t] = s[k] - beta * s[k] * i[k]/N$
$e[t] = e[k] - beta * s[k] * i[k]/N - alpha * e[k]$
$e[t+1] = e[k+1] - beta * s[k+1] * i[k+1]/N - alpha * e[k+1]$
$i[t] = i[k] + alpha * e[k] - gamma * i[k] - delta * i[k]$
$r[t] = r[k] + gamma * i[k]$
$d[t] = d[k] + delta * i[k]$
pred_values = (s[t], e[t], i[t], r[t], d[t])
RETURN pred_values
end

5 Results and Analysis

Forecasted infected and deceased values and plots for India 1, few districts of Maharashtra 2, Tamilnadu 3, Gujarat 4 and the cities Lombardia, Italy 5 and Moscow,

```
prediction()
for x = T−4 to T do
    for y = x+1 to T+7 do
        pred[x][y] ← seird(alpha[x], beta[x], gamma[x], delta[x], x, y)
    end for
end for
for y = T+1 to T+7 do
    index = retrieve-first-index-of( median(pred.i[T-4][y],pred.i[T-3][y], pred.i[T-2][y],pred.i[T-1][y],pred.i[T][y] ) )
    Final_pred_i = pred.i[index][y]
    Final_pred_r = pred.r[index][y]
    Final_pred_d = pred.d[index][y]
end for
end
```

Russia 6 are shown here. Test data from 31^{th} of May to 6^{th} of June is used for the districts in India and 30^{th} of May to 5^{th} June for others.

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-31	89740	97950	5185	5155
2020-06-01	93379	102412	5407	5381
2020-06-02	97019	107096	5608	5618
2020-06-03	101081	112014	5830	5866
2020-06-04	106722	117173	6089	6125
2020-06-05	111905	122585	6363	6396
2020-06-06	116302	128261	6649	6679

Table 1: Forecasts of Infected and Deceased cases of *India*

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-31	3658	3366	320	321
2020-06-01	3914	3445	338	329
2020-06-02	3531	3537	348	338
2020-06-03	3511	3641	367	347
2020-06-04	3675	3756	376	356
2020-06-05	3768	3881	390	365
2020-06-06	3768	4016	390	374

Table 2: Forecasts of Infected and Deceased cases of *Pune, Maharashtra*

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-31	6776	6750	132	128
2020-06-01	7444	6961	141	134
2020-06-02	7868	7186	153	141
2020-06-03	8400	7424	153	147
2020-06-04	9067	7676	167	154
2020-06-05	9420	7940	179	162
2020-06-06	9420	8218	179	169

Table 3: Forecasts of Infected and Deceased cases of *Chennai, Tamil Nadu*

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-31	4420	4809	842	849
2020-06-01	3922	4735	864	868
2020-06-02	3158	4676	888	888
2020-06-03	3221	4631	910	907
2020-06-04	3188	4598	938	925
2020-06-05	3231	4576	968	944
2020-06-06	3231	4565	968	962

Table 4: Forecasts of Infected and Deceased cases of *Ahmedabad, Gujarat*

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-30	21809	22462	16079	16049
2020-06-31	20996	22250	16112	16049
2020-06-01	20861	22047	16131	16086
2020-06-02	20255	21852	16143	16123
2020-06-03	20224	21665	16172	16160
2020-06-04	20224	21486	16201	16196
2020-06-05	19853	21315	16222	16268

Table 5: Forecasts of Infected and Deceased cases of *Lombardia, Italy*

Date	Infected		Deceased	
	Actual	Forecasted	Actual	Forecasted
2020-05-30	97464	96737	2408	2392
2020-06-31	98135	95281	2477	2459
2020-06-01	98296	93931	2553	2526
2020-06-02	94481	92681	2624	2591
2020-06-03	92877	91527	2685	2656
2020-06-04	91750	90464	2749	2720
2020-06-05	90905	89487	2806	2783

Table 6: Forecasts of Infected and Deceased cases of *Moscow, Russia*

5.1 Analysis

Actual vs. Forecasted plots of Infected and Deceased cases for India are shown in Figure 5 and 6 repectively. For Pune, figures 7, 8, Chennai 9, 10, Ahmedabad 11, 12, Lombardia 13, 14 and Moscow 15, 16 are presented here for Infected and Deceased cases respectively.

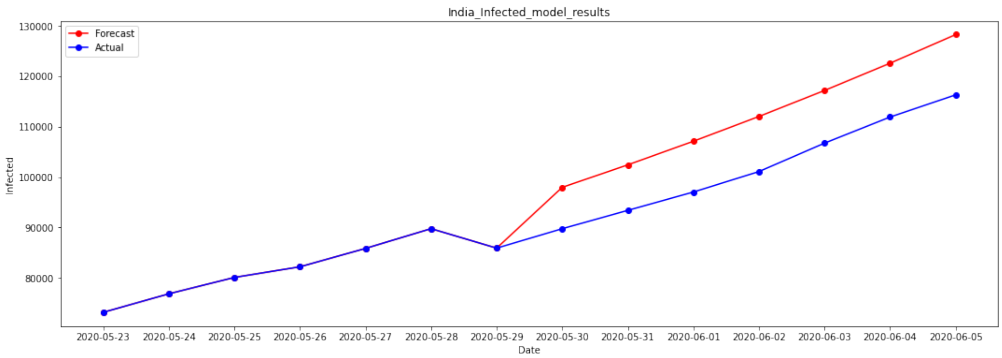

Fig. 5: Forecasts of Infected Cases of *India* from 23^{rd} May to 5^{th}*June*

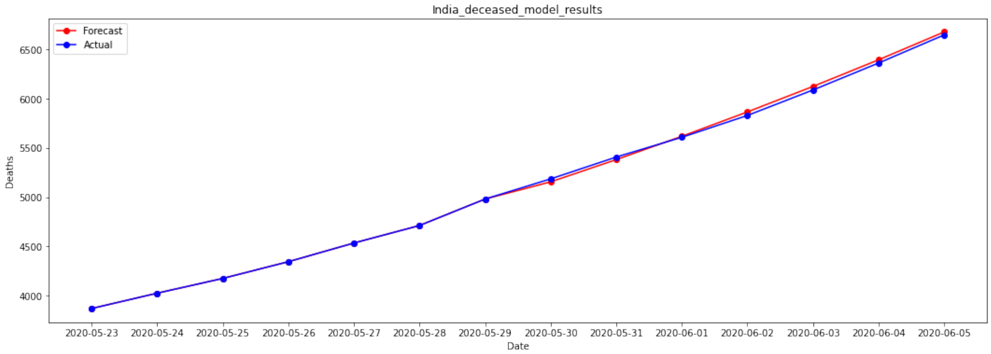

Fig. 6: Forecasts of Deceased Cases of *India* from 23^{rd} May to 5^{th} *June*

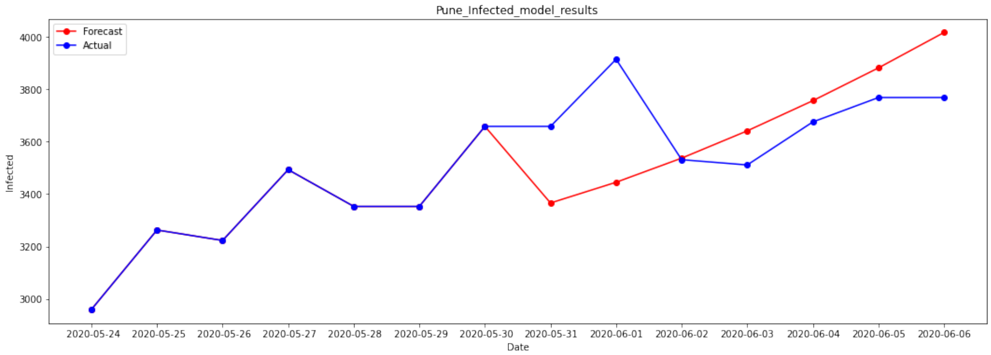

Fig. 7: Forecasts of Infected Cases of *Pune* from 24^{th} May to 6^{th} *June*

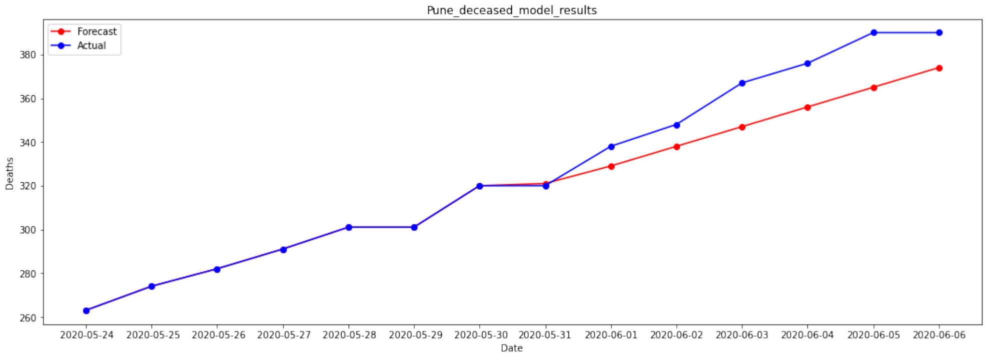

Fig. 8: Forecasts of Deceased Cases of *Pune* from 24^{th} May to 6^{th} *June*

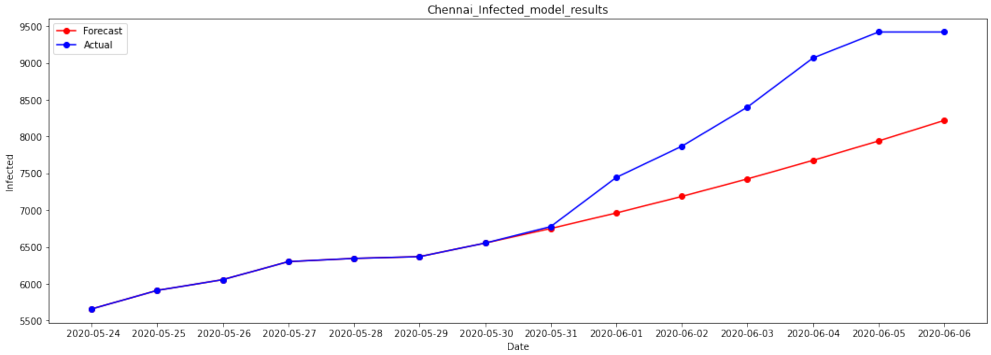

Fig. 9: Forecasts of Infected Cases of *Chennai* from 24^{th} May to 6^{th} *June*

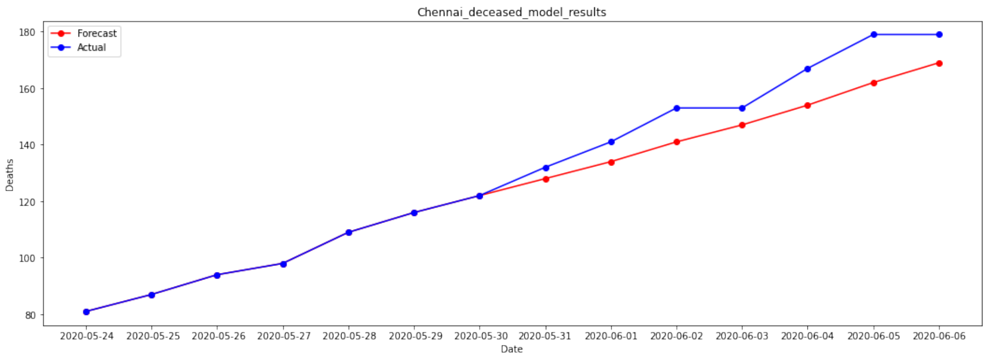

Fig. 10: Forecasts of Deceased Cases of *Chennai* from 24^{th} May to 6^{th} *June*

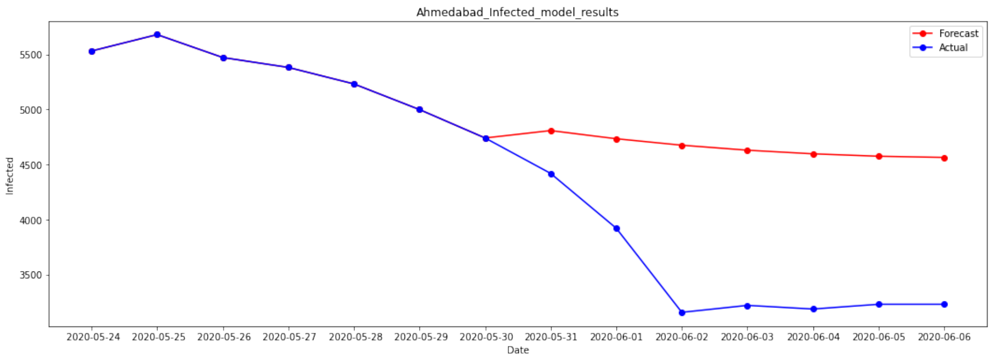

Fig. 11: Forecasts of Infected Cases of *Ahmedabad* from 24^{th} May to 6^{th} *June*

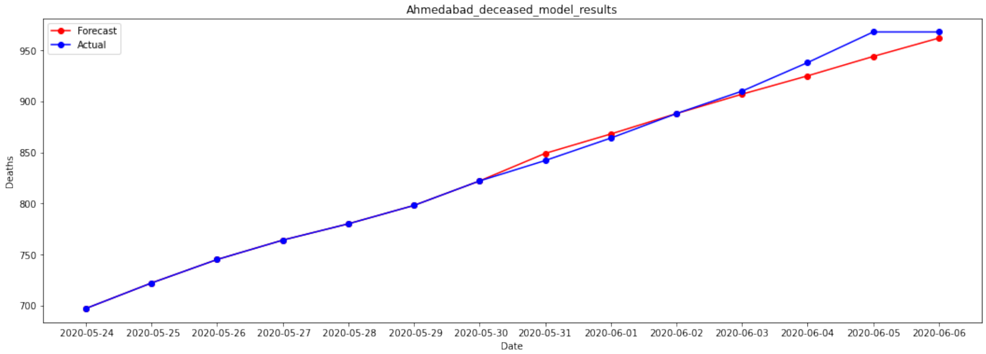

Fig. 12: Forecasts of Deceased Cases of *Ahmedabad* from 24^{th} May to 6^{th} *June*

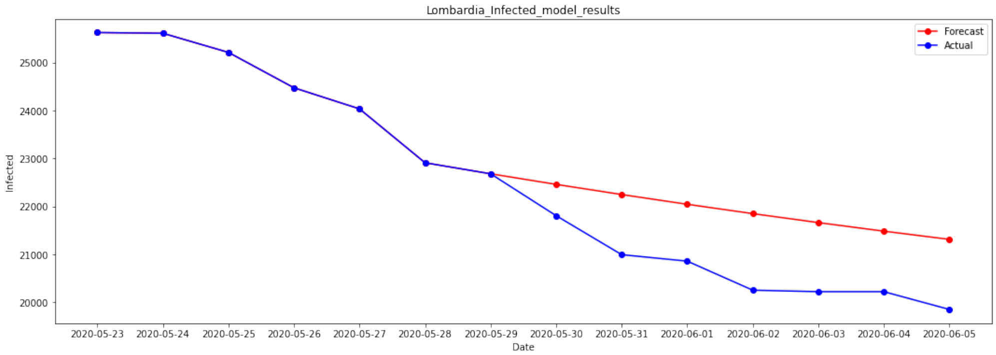

Fig. 13: Forecasts of Infected Cases of *Lombardia, Italy* from 23^{rd} May to 5^{th} *June*

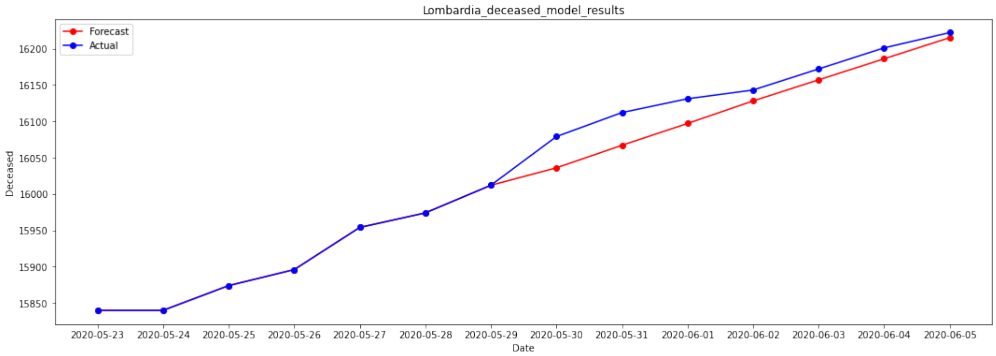

Fig. 14: Forecasts of Deceased Cases of *Lombardia, Itlay* from 23^{rd} May to 5^{th} *June*

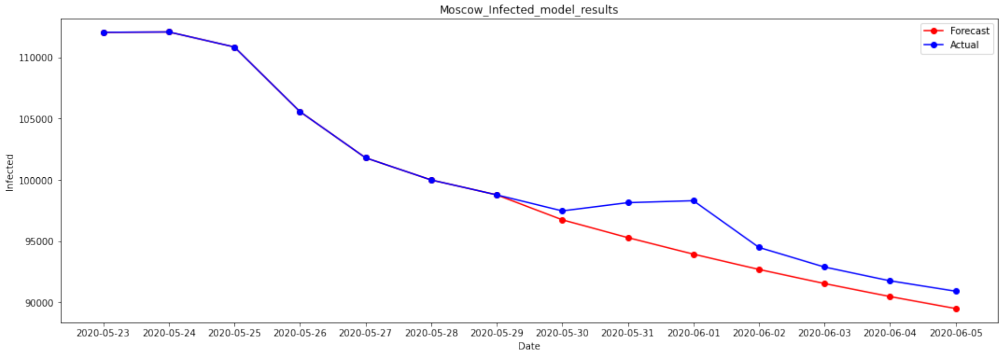

Fig. 15: Forecasts of Infected Cases of *Moscow* from 23^{rd} May to 5^{th}*June*

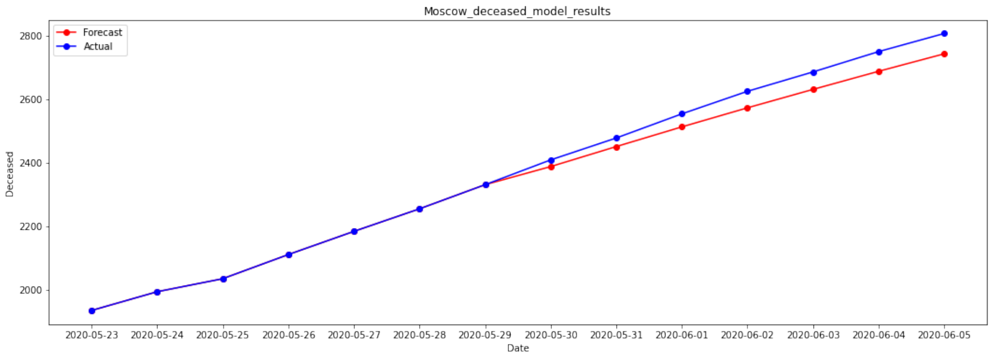

Fig. 16: Forecasts of Deceased Cases of *Moscow* from 23^{rd} May to 5^{th}*June*

From the results obtained for the above mentioned areas, it can be summarized that the model can capture the current-trend properly. Its forecasts are based on the growth rate of the actual curve. If there is a sudden increase or decrease in the growth rate, the forecasts will not be so accurate until the model stabilizes.

5.2 Results

Shown below are the forecasts of Infected and Deceased cases from 6^{th} June to 12^{th} June for India 7, Lombardia 8, and Moscow 9.

Date	Infected	Deceased
2020-06-06	120913	6946
2020-06-07	125745	7255
2020-06-08	130804	7576
2020-06-09	136098	7910
2020-06-10	141634	8258
2020-06-11	147420	8620
2020-06-12	153466	8997

Table 7: Forecasts for *India* from 6^{th} June to 12^{th} June

Date	Infected	Deceased
2020-06-06	20144	16258
2020-06-07	20122	16287
2020-06-08	20102	16316
2020-06-09	20084	16345
2020-06-10	20068	16373
2020-06-11	20054	16402
2020-06-12	20042	16431

Table 8: Forecasts for *Lombardia, Italy* from 6^{th} June to 12^{th} June

Date	Infected	Deceased
2020-06-06	89721	2874
2020-06-07	88813	2936
2020-06-08	87972	2997
2020-06-09	87194	3058
2020-06-10	86477	3118
2020-06-11	85818	3178
2020-06-12	85214	3237

Table 9: Forecasts for *Moscow, Russia* from 6^{th} June to 12^{th} June

5.3 Extensions of the present model

As mentioned earlier, the main goals of using this model are the accuracy of the forecasts and minimum assumptions. We have taken care to avoid assumptions while building the model. As we try to add more compartments to the model, the number of parameters involved also increase. In order to estimate the parameters, either we make an educated guess or derive them from the data. This model is developed keeping India in mind. Since there is no proper data available regarding the tests, the quarantined and others for India, we did not compartmentalize our model. If any other country has the data required for compartmentalization, the model can be extended further.

6 Limitations

This model considers a closed population. Birth, Mortality rates, and others are not considered. This model is limited to short term forecasts as the parameters keep changing and they can not be approximated to long term. The transmission rate for the exposed is not considered due to the uncertainty in calculating the exposed.

7 Future Scope

Considering population density instead of a homogeneous population to forecast accurate results. The forecasts of Infected cases are affected by the sudden change in the data. The model has to be improved so that it need not wait until the model is stabilized. Transmission rate for exposed has to be tuned properly. Quarantine factor and others can be included to get a more detailed analysis of the situation.

8 Conclusions

Several papers are published using SEIR to predict the results. Initial parameters are assumed to be constant in these papers. The parameters were assumed based on input from hospitals and other sources. In this model, parameters are calculated from the data rather than making an educated guess. The goal was to forecast these results so that we can estimate and plan for Hospital equipment and Personal Protective Equipment in advance.

References

1. Sourish Das, Prediction of Covid-19 Disease Progression in India, arXiv:2004.03147 [q-bio.PE], 2020.
2. *Covid19India* `https://api.covid19india.org/`
3. Worldometer, `https://www.worldometers.info/coronavirus/`
4. Yi-Cheng Chen, Ping-En Lu, Cheng-Shang Chang, Tzu-Husan Liu, A Time-dependent SIR Model for COVID-19 with Undetected Infected Persons, `http://gibbs1.ee.nthu.edu.tw/A_TIME_DEPENDENT_SIR_MODEL_FOR_COVID_19.PDF` , 2020.
5. *Github Link* `https://github.com/Taarak9/COVID-19_Predictive_Model_SEIRD/`
6. Samuel M. Jenness, Steven M. Goodreau, Martina Morris EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks, J Stat Softw. doi:10.18637/jss.v084.i08, (2018).
7. Chayu Yang and Jin Wang , A mathematical model for the novel coronavirus epidemic in Wuhan, China, Mathematical Biosciences and Engineering, Volume 17, Issue 3, 2708–2724, (2020).
8. Zifeng Yang, Zhiqi Zeng, Ke Wang et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J Thorac Dis, 12(3):165-174, (2020).
9. Biao Tang, Nicola Luigi Bragazzi, Qian Li et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov), Infectious Disease Modelling 5, 248-255, (2020).
10. Meher K. Prakash, Shaurya Kaushal, Soumyadeep Bhattacharya, Akshay Chandran, Aloke Kumar and Santosh Ansumali, A minimal and adaptive prediction strategy for critical resource planning in a pandemic, medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057414, April, (2020).
11. Gaurav Pandey, Poonam Chaudhary, Rajan Gupta, Saibal Pal, SEIR and Regression Model based COVID-19 outbreak predictions in India, arXiv:2004.00958 [q-bio.PE], (2020).
12. `http://c19.dakshas.org:5000/`

13. https://forms.gle/sCvsLzopqTfFmoog9
14. https://github.com/pcm-dpc/COVID-19/blob/master/dati-regioni/dpc-covid19-ita-regioni.csv
15. https://www.kaggle.com/kapral42/covid19-russia-regions-cases?select=covid19-russia-cases-scrf.csv