# Evaluating growth pattern and assessing future scenario of COVID-19 epidemic of India

Nandan L. Patil<sup>1\*</sup> and Lakshmi R. Gangavati<sup>1</sup>

<sup>1</sup>Department of Genetics and Plant Breeding,

College of Agriculture,

University of Agricultural Sciences,

Dharwad- 580 005.

\* Corresponding author.

Email address: patilnads@gmail.com

## 1 Abstract

2 COVID-19 the modern pandemic has spread across the world at a rapid pace. SARS-CoV 3 2 is highly transmissible and the rate of infection is exponential for heavily infected countries. 4 Asymptotic carriers and longer incubation period have been key towards such a large-scale distribution of disease. Data released by official authorities on COVID-19 cases is significantly 5 6 affected by various factors such as size of sample, incubation period of disease and time taken to 7 test the sample. These factors mask the useful pattern (signal) of disease spread. Thus, an ingenious method to group data into cycles of five and seven days, for studying pattern of 8 9 disease spread is undertaken. Occurrence of recurrent peaks as indicated by Adjusted Rate of 10 infection per day indicated the spread of disease has been non-uniform. Currently, India is yet to reach the critical point (peak of epidemic) with adjusted daily cases more than 1000. Increasing 11 12 testing capacity along with random sampling and sample pooling can help in preventing formation of these peaks in future. The proposed method helps in assessing the current state and 13 for predicting future scenarios epidemics. 14

15 Key Words: COVID-19, SARS-CoV 2, India, epidemic, rate of Infection, daily cases

16

## 17 **1. Introduction**

18 Novel coronavirus has swept the world at rapid pace with more than 3 million cases 19 across 185 countries since its first report on November, 2019 in Wuhan, China. The disease was 20 termed as Public Health Emergency by World Health Organization (WHO)on January 30, 2020 21 and further on 11 March was announced as pandemic. The pandemic has resulted in huge 22 economic cost, loss of jobs and resulting in largest global recession since The Great Depression 23 (Gopinath, 2020). The International Virus Classification Commission (ICTV) classified this 24 novel coronavirus as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) on February 11, 2020 and on the same day WHO named the disease as COVID-19. Virus is 25 26 transmitted mainly through respiratory droplets and has high transmissibility rate but low mortality unlike Middle Easter Respiratory Syndrome (MERS) and SARS-CoV 1 which have 27 high mortality rate. The mean incubation period of disease is6.6 with range of 2.1 to 11.1 28 29 days(Backer et al., 2020, Huang et al., 2020). COVID-19 symptoms include aches and pains, nasal congestion, runny nose, sore throat or diarrhea and needs technical tests such as RT-PCR, 30 31 Anti body testing or CRISPAR (Liu et al., 2020) to detect the virus. Many cases are reported were no symptoms are visible but the person is tested positive for the virus. These asymptotic 32 people have been key for such a large scale spread of virus. 33

34 India reported its first case on 30 January, 2020 from Southern state of Kerala, who was a returning student from Wuhan, China. Only three cases were reported up till 1 March and since 35 then the cases have increased exponentially and as of April 23, 27000 cases have to identified 36 37 with Maharashtra recording highest positive cases. The virus has entered the country through 38 foreign travelers coming mainly from China, Europe US and Gulf countries. Foreign flights were stopped on 22 March and nation-wide lockdown was announced on 24 March for 21 days i.e. up 39 till 14 April which was further extended to 3 May. But large population of migrant workers had 40 returned to their native place during initial days of lockdown. Currently, in India, asymptotic 41 42 people with international travel history and all contacts of laboratory confirmed positive cases are being tested for the virus mainly through RT-PCR. As of now there have been no effective 43 44 drugs or vaccine against the disease, thus identifying positive cases and isolating/treating them is 45 effective option.

Compartmental models namely SIR (Susceptible, Infectious and Recovered) models and their modification (Giordano et al., 2020, Wu et al., 2020, Prem et al., 2020) are widely used in analyzing the epidemiological data. Furthermore, statistical models such as ARIMA and neural networks (Wei et al., 2016, Anne and Jeeva, 2020, Ceylan, 2020) are used in forecasting the COVID-19 cases. These models become complex when a high number of parameters are involved. Thus, here we formulate a simple statistical method to study the growth of epidemics and predict future scenarios based on only the positive cases reported.

## 53 **2. Methodology**

54 Data of positive cases for COVID-19 epidemic was collected from Johns Hopkins 55 University Center for Systems Science and Engineering (JHU CSSE, 2020) starting from 22 56 February up to 23 April.

57 Let the first day *i.e.* February 22 be  $d_1$ , February 23 $d_2$  and so on with  $d_n$  being n<sup>th</sup> day from 58 the first day.

59  $cc_1, cc_2, cc_3, \dots, cc_n$  is number of cumulative cases (cc) for the  $d_1, d_2, d_3, \dots, d_n$  day and 60  $dc_1, dc_2, dc_3, \dots, dc_n$  is number of daily cases (dc) for  $d_1, d_2, d_3, \dots, d_n$  day.

61

62 i) Estimating Unadjusted Rate of Infection $(UR_d)$ :

63

$$UR_d = \frac{dc_n}{dc_{(n-1)}} \times 100$$

64 where,

65  $UR_d$  is Unadjusted **R**ate of Infection for d<sup>th</sup> day estimated from raw data,

66  $dc_n$  is number of positive cases for n<sup>th</sup>day,

67 
$$dc_{(n-1)}$$
 is number of positive cases for  $(n-1)^{\text{th}}$ day

68 Cumulative cases (cc) or daily cases (dc) reported by official authorities is a mixture of 69 both useful signal and noise. A clear pattern of epidemic growth cannot be ascertained from the 70 raw data itself. Thus, in order to obtain a clear picture of epidemic growth and to minimize the 71 noise (sampling effect) in the data we group the data based on five day and seven day cycle to 72 estimate new coefficients  $R_{cmi}$  and  $Adc_n$  as described below.

Let 'c' be the number of days in a given cycle, then number of possible combinations the data can be grouped is equal to 'm' which are repeated over 'i' intervals where 'm' is equal to 'c'. These combinations are referred to as Models from henceforth. In the current study we group the data in cycle of five days (5D) and seven days (7D).

77 ii) Total positive cases for m<sup>th</sup> model during i<sup>th</sup> interval for cycle of seven days (c=7) is
78 calculated as follows,

79		<i>PCI</i> <sub>711</sub> =	$d_{1+}d_{2+}$	+d <sub>7</sub>
80		<i>PCI</i> <sub>721</sub> =	$d_2 + d_3 + .$	+d <sub>8</sub>
81			•	•
82			•	•
83			•	•
84		<i>PCI</i> <sub>771</sub> =	$d_7 + d_3 + $ .	+d <sub>13</sub>
85				
86	For 'c' day cycle,			

$$PCI_{cmi} = \sum_{n=c(i-1)+m}^{m+ci-1} dc_n$$

87 where, 88 c = 1,2, 3,..., c and m = 1,2, 3,... and i = 1,2, 3,...89  $PCI_{cmi}$  is **P**ositive Cases of m<sup>th</sup> model during i<sup>th</sup> Interval for 'c' day cycle 90  $dc_n$  is number of **d**aily **c**ases for n<sup>th</sup> day 91 iii) Estimating Rate of infection per 'c' day cycle ( $R_{cmi}$ ):  $R_{cmi} = \frac{PCI_{cmi}}{PCI_{cm(i-1)}} \times 100$ 

98  $AR_{c1} = \frac{R_{c11}}{c}$ 

99 
$$AR_{c2} = \frac{R_{c11}}{c} + \frac{R_{c21}}{c}$$

100 
$$AR_{c3} = \frac{R_{c11}}{c} + \frac{R_{c21}}{c} + \frac{R_{c31}}{c}$$
 and so on

101 v) Calculation of Adjusted daily cases  $(Adc_n)$ :

102 The Adjusted daily cases for 'c' day cycle for  $n^{th}$  day  $(Adc_n)$  is calculated as,

103

104 
$$Adc_1 = \frac{PCI_{c11}}{c}$$

105 
$$Adc_2 = \frac{PCI_{c21}}{c} + \frac{PCI_{c11}}{c}$$

106 
$$Adc_3 = \frac{PCI_{c31}}{c} + \frac{PCI_{c21}}{c} + \frac{PCI_{c11}}{c}$$
 and so on.

The unadjusted daily cases (Udc<sub>n</sub>) is defined as daily cases calculated from official record i.e.
dc<sub>1</sub>, dc<sub>2</sub>, dc<sub>3</sub>.....

# 109 **2.1 Outliers**

110 If the magnitude of  $PCI_{cm(i-1)}$  is less than 10 it leads to greater  $R_{cmi}$  which effects the 111  $AR_{cn}$  and masks pattern in the growth of epidemic. Thus, if  $PCI_{cm(i-1)} < 10$ , then respective 112  $R_{cmi}$  are removed from analysis.

#### 113 2.2 Critical Point

114 Critical point is defined as a stage where the  $R_{cmi}$  or adjusted daily cases  $(AR_{cn})$  are equal 115 to 100 and  $\frac{100}{c}$ , respectively. In other words, critical point refers to period past the peak of the

epidemic curve. The  $AR_{cn}$  critical point for 5D and 7D cycle is 25 and 14.29, respectively.

# 117 2.3 Classifying the growth of epidemic period

118 Based on the critical point, the epidemic period can be classified into two phases,

a. Growth phase: It is the initial phase were epidemic has not yet reached the critical point.
 During this period of epidemic during which the number of daily cases is continuously
 rising and epidemic is yet to reach its peak stage.

b. Safe phase: It is the phase after passing the critical point or peak of epidemic when thedaily cases are continuously decreasing.

124 **2.4 Prediction of** *AR*<sub>cn</sub> **values** 

ARIMA is a statistical method to forecast time series data which was proposed by Box 125 126 and Jenkins in 1970. These models have been commonly used to forecast COVID-19 positive cases (Benvenuto et. al., 2020, Ceylan, 2020, Anne and Jeeva, 2020). ARIMA model was used to 127 forecast the  $AR_{cn}$  value of 5D and 7D cycle. The three best fitting parameters of ARIMA model 128 p (order of autoregression), d (degree of difference) and q (order of moving average) were (2, 0, 129 0) and (3, 2, 1) for 5D and 7D cycle selected based on root mean square error (RMSE). 130 Furthermore, univariate Neural Network Models (NNM) were used in predicting the  $AR_{cn}$  of 5D 131 and 7D cycle by using *forecast* package in R. 132

## 133 3. Result and Discussion

134 COVID-19 pandemic has caused large scale disruption of normal life with nationwide 135 lockdown being imposed to control spread of virus. We have analyzed the data in a novel way to 136 understand growth of virus and anticipate future conditions. Unadjusted rate of infection per day 137  $(UR_d)$  plot (Figure 1) estimated from raw data shows continuous fluctuation throughout the 138 epidemic period. No growth pattern of disease can be delineated from the plot. The reason for 139 such fluctuation is that raw data is a mixture of both useful signal which represents growth of 140 epidemic along with redundant noises which mask the signal. The noise in the data arises due to 141 various reasons such as sampling size, time taken to test the sample, time taken to trace the 142 contacts and incubation period for the patient. Thus, making it impossible to assess present 143 situation and predict future conditions using  $UR_d$  plots.

144

# 3.1 Analyzing Adjusted Rate of Infection per day (AR<sub>cn</sub>)

In order to reduce the noise embedded in raw data, we grouped raw data into 7D and 5D 145 cycle and estimated the adjusted rate of infection per day.  $AR_{cn}$  plot (Fig. 2 and Fig. 3) for 7D 146 147 and 5D cycle reveals the pattern of spread of Epidemic. The Epidemic period can be classified 148 into four and three phases (Table 1) as per  $AR_{cn}$  of 5D and 7D cycle respectively. Phase 1 (5D) was from 5 March to 16 March, with peak on 14 March and Phase 2 (5D) from 17 March to 28 149 150 March with peak on 21 March. The first two phases of 5D combine to form Phase 1 of 7D cycle 151 which was from 5 March to 29 March with peak on 22 March. During these initial stages of epidemic, the virus is difficult to detect due to high incubation period, asymptotic carriers and 152 153 technical tests to detect the virus. Thus, the virus spreads can spread unnoticed if proper 154 preventive measures are not undertaken. With increase in testing capacity a greater number of 155 COVID-19 cases are detected in short time leading to formation of peak on 22 April.

156 Phase 3 (5D) from 29 March to 8 April with peak on 2 April and Phase 2 (7D) 30 March to 13 April with peak on 3 April can be attributed to Tablighi Jamaat congregation which was 157 158 held in New Delhi during mid-March. Many of the attendees have been tested positive. Finally, we observe another peak on 13 April under Phase 4 (5D) starting from 10 April. In a  $AR_{cn}$  plot 159 160 the new peaks formed are more severe than that of early ones. Meaning peak of Phase 4 (5D) is more severe than peak of Phase 3 (5D) and Phase 2 (7D) which is further more severe than peak 161 of Phase 1 (7D) or Phase 2 (5D). The 5D cycle detected a greater number of peaks than 7D cycle 162 as it is more sensitive to sudden increase in cases within short period of time. The recurrent 163

164 peaks which we observe in  $AR_{cn}$  plots may be attributed to limited coverage of testing facility 165 which allows the virus to spread through asymptotic carriers.

## 166 **3.2 Impact of Nationwide lockdown**

167 Nationwide shutdown started from 24 March and is scheduled to be lifted on May 3. As 168 seen from the  $AR_{cn}$  plot the rate of new infection starts to decrease after its peak on March 22. 169 The nationwide lockdown has significantly slowed the spread of virus, but it should be followed 170 by wide scale testing in order to identify the infected people. Delay in action or lack of testing 171 would lead to increase in  $AR_{cn}$  or under more severe conditions new peaks are formed.

172 **3.3 Analyzing** *R*<sub>cmi</sub>**plots** 

173  $R_{cmi}$  represents change in rate of infection for a given days of cycle. When the spread is 174 constant the $R_{cmi}$  of all models for any days of cycle start converging and continue in same 175 direction. But when growth is non-uniform the models start bisecting each other.

For seven-day cycle  $R_{cmi}$  plot (Fig. 4) highest value is 521 of Model 7for interval March 20 to March 26 which coincides with peak of Phase 1 (7D). The models are continuously bisecting each other indicating the spread or detection of virus has not been uniform. Similarly,  $R_{cmi}$  plot of five-day cycle indicates same trend of bisecting models with highest value of 455.263 for interval March 20 to March 24 coinciding with peak of Phase 2 (5D). As seen from the analysis of other countries (not shown) the Models tend to converge (or are close to each other) before reaching the critical point.

183 **3.4 Adjusted daily cases**  $(Adc_n)$ 

184  $Adc_n$  accommodates the fluctuation in the original data to produce a smooth curve 185 representing increase in daily cases. India passed  $Adc_n$ (Fig. 5) of 1,000 on 14 April and as of 18 186 April, with current  $Adc_n$  being 1173 and 1164 as calculated from 5D and 7D cycle respectively. 187  $Adc_n$  is continuously increasing during growth phase whereas starts decreasing once in safe 188 phase. Thus, delay in reaching the critical point enhances the severity of Epidemic.

### 189 **3.5 Future course of Epidemic**

190 ARIMA model forecasted (Fig. 6) formation of peak on 28 April for  $AR_{cn}$  of 5D cycle 191 whereas for 7D cycle it forecasted a gradual decrease. Similarly, neural network model (NNM) 192 forecasted (Fig. 7) towards formation of peak on 30 April for 5D cycle and gradual increase in 193  $AR_{cn}$  for 7D cycle.

194 As seen from the  $R_{cmi}$  plot, though the rate of infection decreasing it the rate at which it is decreasing is slowing down. The last  $R_{cmi}$  value for 7D cycleincreased from 128.22 ( $R_{769}$ ) to 195 129.69 ( $R_{779}$ ), similar trend was observed for 5D cycle which indicates a possible rise COVID-196 197 19 positive cases. Presently India is in growth stage where  $Adc_n$  is continuously rising. With recurrent peaks occurring, India may not reach critical point within May 3, the scheduled date to 198 end the nationwide lockdown. Reducing  $R_{cmi}$  and  $AR_{cn}$  can be achieved through rigorous 199 200 contact tracing and alternate testing methods such as Rapid antibody testing, CRISPAR (Liu et al., 2020) and sample pooling method (Hogan et al., 2020) can be useful for a large nation as 201 India. 202

## 203 **3.6 Limitations of the method**

As of other statistical methods there are few limitations for the present method. When the spread of disease is slow the  $AR_{cn}$  and  $R_{cmi}$  plot can't depict the pattern of disease spread. Thus, the method is useful where local transmission or community spread of disease is prevalent.
Pattern of disease spread is delineated by the method is based on reported data and doesn't
consider the underlaying unreported cases.

209 4. Conclusion

AR<sub>cn</sub> and R<sub>cmi</sub> plots help us in understanding the pattern of Epidemic and predicting the future scenario by minimizing the noise using in the raw data. The growth of epidemic in India has been non-uniform and is currently growing in linear fashion with  $Adc_n$  crossing 1,000. Lockdown combined with greater testing capacity is crucial to contain the spread of novel coronavirus. In the current study, we employed ingenious method to classify data into5D and 7D cycle to assess spread of virus and predict the future progression of epidemic. The method is simple and robust which uses publicly and most commonly available data.

### 217 Acknowledgements

The authors are thankful to all resource persons for their useful suggestions. No funding to declare.

## 220 **Conflict of Interest**

221 The authors declare no conflict of interest.

### 222 **References:**

- Anne, R., and Jeeva, Carolin S., 2020. ARIMA modelling of predicting COVID-19 infections.
- 224 *medRxiv*. <u>https://doi.org/10.1101/2020.04.18.20070631</u>.

- Backer, J. A., Klinkenberg, D. and Wallinga, J., 2020. Incubation period of 2019 novel
  coronavirus (2019-nCoV) infections among travelers from Wuhan, China, 20–28 January
  2020. *EuroSurveillance*, 25(5), 2000062. <u>https://doi.org/10.2807/1560-</u>
  917.ES.2020.25.5.2000062.
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M., 2020. Data in brief
  application of the ARIMA model on the COVID- 2019 epidemic dataset. Data Br
  29,105340. https://doi.org/10.1016/j.dib.2020.105340.
- Ceylan, Z., 2020. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total Environ*.729,138817.https://doi.org/10.1016/j.scitotenv.2020.138817.
- Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., DiFilippo, A., DiMatteo, A. and Colaneri,
   M., 2020. Modelling the COVID-19 epidemic and implementation of population
   wideinterventionsinItaly.*NatureMedicine*.1-6.https://doi .org /10.<u>1038/s41591-020-0883-</u>
   <u>7</u>.
- Gopinath,G.,2020.TheGreatlockdown:Worsteconomicdownturnsincethegreatdepression,Internatio
   nalMonetaryFund(IMF),Insightsandanalysisoneconomicsandfinance.https://blogs.imf.org/2
   020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression.
- Hogan,C.A.,Sahoo,M.K.andPinsky,B.A.,2020.Samplepoolingasastrategyto detect community
   transmission of SARS-CoV-2. *JAMA*. http://doi.org/10.1001/JAMA.2020.5445.

243	Huang, C., Wang, Y., L1, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X. and Cheng, Z., 2020.
244	Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The
245	Lancet.395(10223),497-506. https://doi.org/101016/S0140-6736(20)30183-5.

7 2020

JHU CSSE, 2020. Covid-19 Dashboard, COVID-19 Map - Johns Hopkins Corona virus
 Resource Center. https://coronavirus.jhu.edu/map.html

Liu,R.,Fu,A.,Deng,Z.,Li,Y.andLiu,T.,2020.Promising methods for detection of novel
 coronavirus SARS-CoV-2.*View*.1(1).<u>https://doi.org/10.1002/viw2.4</u>.

Prem,K.,Liu,Y.,Russell,T.W.,Kucharski,A.J.,Eggo,R.M.,Davies,N.,Flasche,S.,Clifford,S.,Pearso
 n,C.A.,Munday,J.D.andAbbott,S.,2020.Theeffectofcontrolstrategiestoreducesocialmixing
 on outcomes of the COVID-19 epidemic in Wuhan,China:a modelling study. *The Lancet*

253 Public Health. <u>https://doi.org/10.1016/S2468-2667(20)30073-6</u>.

254 Wei, W., Jiang, J., Liang, H., Gao, L., Liang, B., Huang, J., Zang, N., Liao, Y., Yu, J., Lai, J. and Qin, F., 2016. 255 Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis 256 257 incidence in Heng County, China. PloS 11(6). one, 258 https://doi.org/10.1371/journal.pone.0156768.

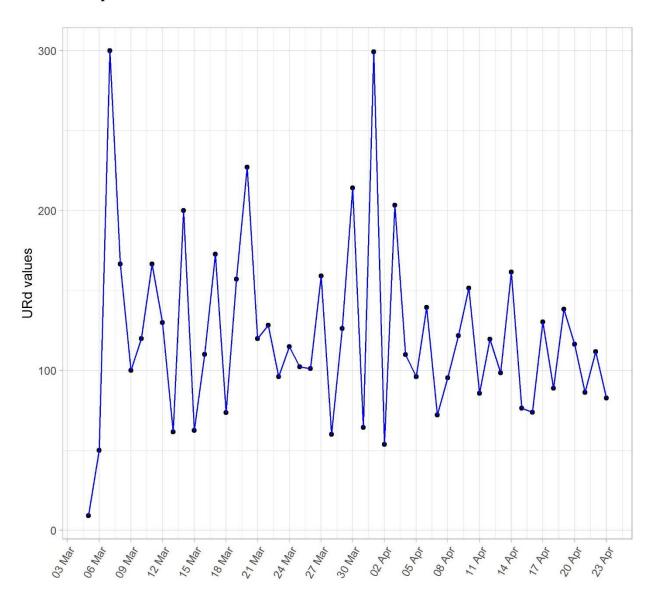
Wu, J. T., Leung, K. and Leung, G. M., 2020. Nowcasting and forecasting the potential domestic
and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a
modelling study. *The Lancet.* 395(10225), 689-697. <u>https://doi.org/10.1016/S0140-</u>
6736(20)30260-9.

5D cycle				7D cycle					
	Start	End	Peak	Total cases		Start	End	Peak	Total cases
Phase 1	5 March	16 March	4 March	90	Phase 1	5 March	29 March	22 March	996
Phase 2	17 March	28 March	21 March	800					
Phase 3	29 March	9 April	2 April	4947	Phase 2	30 March	13 April	3 April	8328
Phase 4	10 April	-	13 April	15835*	Phase 3	14 April	-	-	12348*

Table 1. List of Phases as delineated by  $AR_{cn}$  of 5D and 7D cycle for India

\* as of 23 April

Unadjusted Rd Plot



**Figure 1**. Unadjusted  $AR_{cn}$  from 5 March to 23 April.

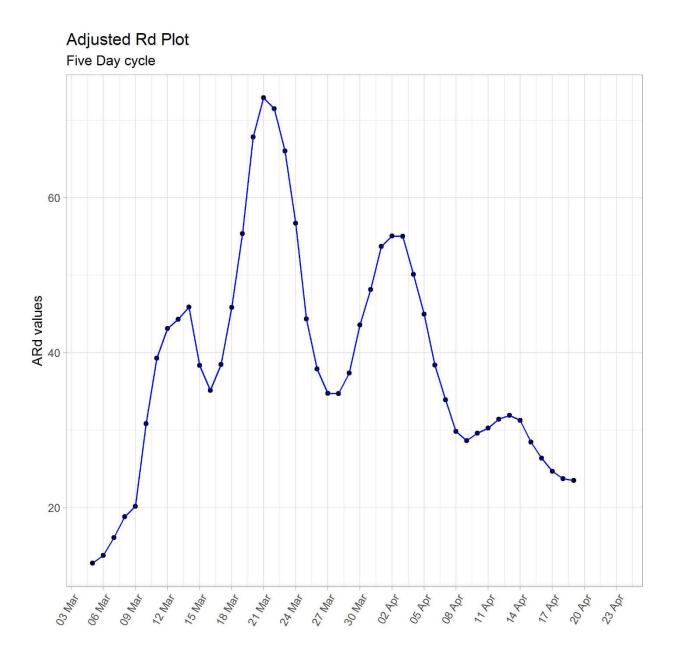
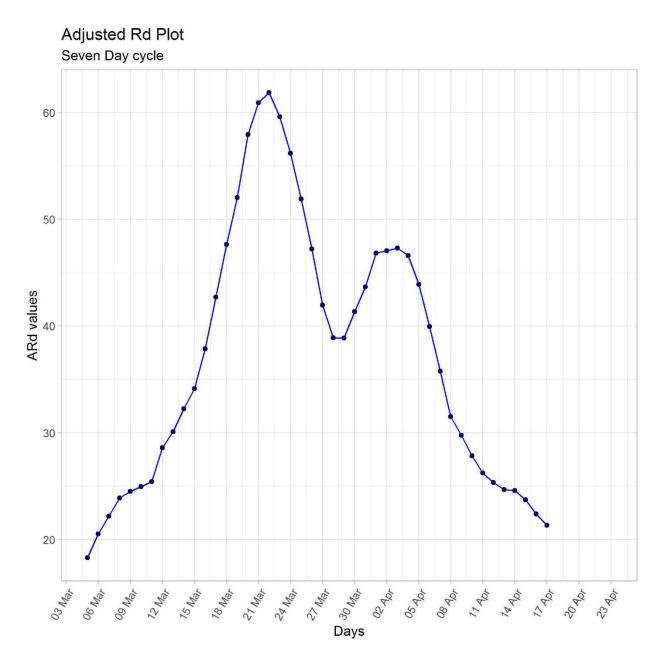


Figure 2. *AR<sub>cn</sub>* plot for five day (5D) cycle from 5 March to 19 March.



**Figure 3**.  $AR_{cn}$  plot for seven day (7D) cycle from 5 March to 17 March.

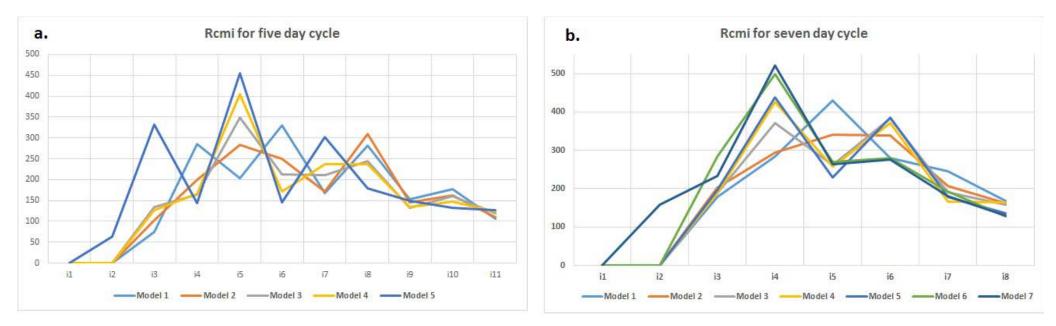
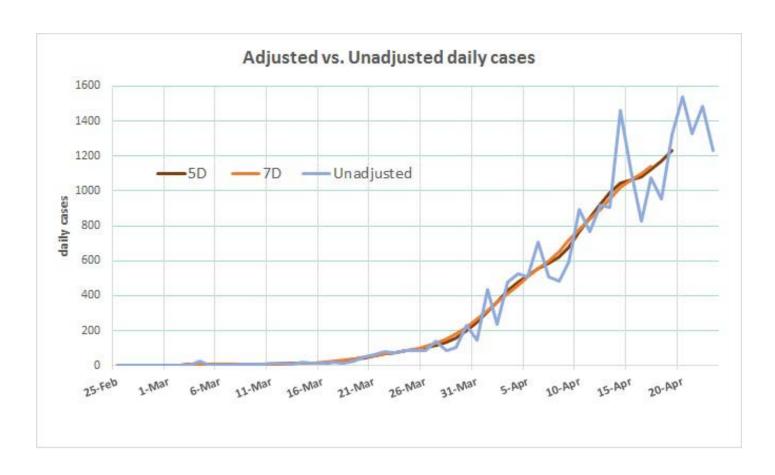
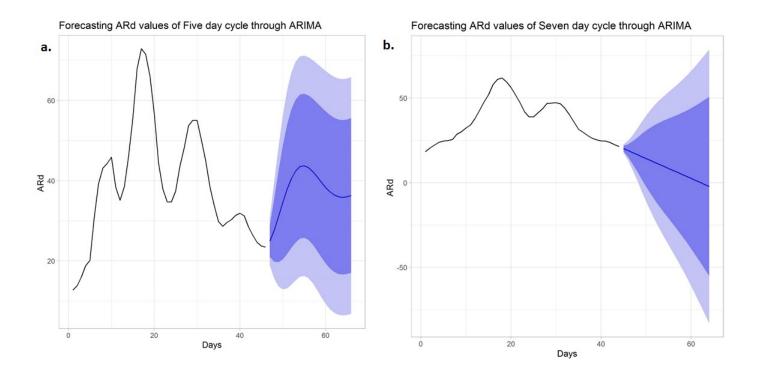


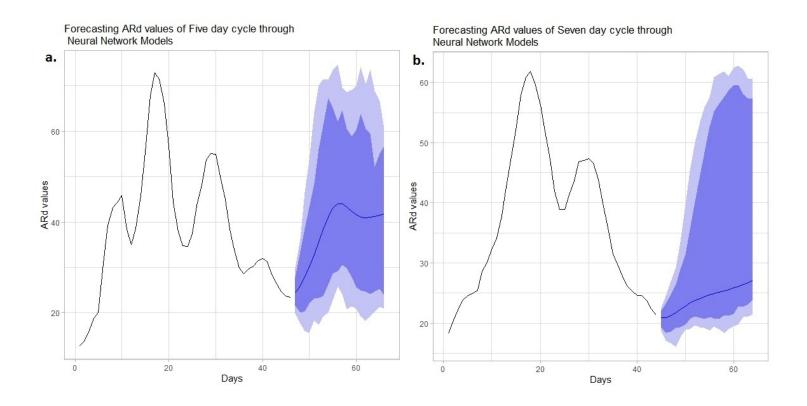
Figure 4. a)  $R_{cmi}$  plots for five day cycle (5D) across 11 intervals, b)  $R_{cmi}$  plot for seven days cycle (7D) across 8 intervals



**Figure 5**. Plot of Unadjusted daily cases  $(Udc_n)$  (up to 23 April) and Adjusted daily cases  $(Adc_n)$  estimated from 5D cycle (up to 19 April) and 7D cycle (17 April).



**Figure 6**. Forecasting  $AR_{cn}$  for five day cycle (5D) 20 March to 9 May and seven day cycle (7D) from 18 March to 7 May through ARIMA.



**Figure 7**. Forecasting  $AR_{cn}$  for five day cycle (5D) 20 March to 9 May and seven day cycle (7D) from 18 March to 7 May through Neural Network Models.