Real-Time COVID-19 Forecasting for Four States of India Using a Regression Transmission Model

Lincoln Priyadarshi Choudhury1, B. Ranjeeth Kumar2

1Karshapana Consultants Private Limited, Moti Nagar, New Delhi, India
2Karshapana Consultants Private Limited, Saidabad, Hyderabad Telangana, India
Email: lincolnchoudhury@gmail.com, ranjeethpaul01@gmail.com

Abstract

Introduction: More than a million people are reported to have been infected with COVID in India, since the beginning of the pandemic. However, the epidemic is not the same across the country. Though there are state-level variations rapidly changing disease dynamics and the response has created uncertainty towards appropriate use of models to project for the future.

Method: This paper aims at using a validated semi-mechanistic stochastic model to generate short term forecasts. This analysis used data available at the respective state government bulletins for four states. The analysis used a simplified transmission model using Markov Chain Monte Carlo simulation with Metropolis-Hastings updating. Results: Two weeks were used to compare the results with the actual data. The forecasted results are well within the 25th and 75th percentile of the actual cases reported by the respective states. The results indicate a reliable method for a real-time short term forecasting of COVID-19 cases. The 1st week projected interquartile range and actual; reported cases for the state of Kerala, Tamil Nadu, Andhra Pradesh and Odisha were (1064 - 2532) 2234, (17,503 - 50,125) 27,214, (5225 - 11,003) 9563, (2559 - 4461) 3925, respectively. Similarly, the 2nd week projected interquartile range and actual; reported cases were (1055 - 7803) 4221, (18,298 - 73,952) 31,488, (4705 - 23,224) 13,357, (2701 - 9037) 4175 respectively.

Conclusion: This real-time forecast can be used as an early warning tool for projecting the changes in the epidemic in the near future triggering proactive management steps.

Keywords
COVID-19, Forecasting, Regression Transmission, Epidemic Model
1. Introduction

As on 18th July 2020, there are more than million cases of COVID-19 reported from India [1]. The first case was reported on 31st of January 2020 [2]. As the pandemic of COVID continued to sweep across the world the country took a series of measures to address it. This included improving the testing capacity and having a testing strategy to identify the cases [3]. However, the uncertainty over the duration and the burden of the pandemic is visible with reports both peer-reviewed and not peer-reviewed [4]. These reports indicate the epidemic to be a range from few hundred thousand to few hundred millions with peak varying between April to July 2020 [5] [6]. Since the first reported case, India has taken several non-pharmaceutical interventions to address the pandemic [7].

India follows a federal structure, where health is a state subject and the centre plays its supporting role at the time of need. While the earlier models provide the bigger picture, the planning for response needed short term projections that can keep a close eye on the upcoming wave of cases, which can help them in local decision making. The states in India are different in population size, density and connectivity to the other parts of the world directly [8]. The burden of COVID-19 both in terms of cases and death differ from state to state [1]. Also the strength of health system is not uniform across the country, thus the local response is expected to be different [9].

The short term forecasting is being found useful for similar infectious diseases earlier [10] [11] [12]. These models use the reported cases as inputs and use various methods like discrete time stochastic model, a conditional intensity of accumulation of cases using non parametric probability method or generation dependent growth factors to develop simple but robust models for forecasting disease respectively. Short term projections using growth models and modified SIR model (Susceptible-Infected-recovered) have also been used for early epidemic [13] [14]. Keeping the above facts in view, this paper is focused on creating real-time and short-term projections for COVID-19 in the near future that can be helpful for the states in India and applicable not only for early but also for later part of the epidemic.

2. Methodology

2.1. Data

The data required for the model was available from several publicly available domains. However, due to reported discrepancies, the authors decided to choose the ones that have the desired information and reported by government through daily bulletins [15]. We collected daily updated data on the number of confirmed cases from all the respective state government daily bulletins and dashboard, which were reporting daily from the first case identified in their state. The data was available in the respective state government websites [16] [17] [18] [19]. These states included Kerala, Tamil Nadu, Andhra Pradesh, and Odisha. The time period of data collection was from 30th January 2020 to 18th
2.2. Model

We re-calibrated the semi-mechanistic discrete-time stochastic compartmental disease model, the details of the model can be found elsewhere [20]. The model consisted of two integer states or compartments of “Susceptible” - “Exposed” - “Infectious”, thus can be considered as mechanistic. The mean latent period was assumed to be 2.5 days. The duration of infectiousness for the model was obtained from the literature search showed pre-symptomatic period was 5 - 6 days with 97% of people infected persons showing infections before 12 days [21] and in India the disease detection from day of sample collection was around 4 days [22]. As soon as people were detected they were removed from the non-infected population through quarantine measures, thus removing them from further transmitting COVID-19. Thus, we assumed the infectious period to be 8 days with range 4 to 12 days.

All the new infections entered the “Exposed” compartment as a Poisson distribution with mean as a product of time-varying reproductive rate $r(t)$ and proportion infected at the time. The model used the available information on the latent period of the disease and the infectious period to move people between the compartments using independent geometric transitions with first-order Euler method [23]. Random-walk methods are used elsewhere for modelling the reproductive rate in outbreak situations [24]. The model used a multiplicative normal random walk with a log-linear drift to generate the $r(t)$ parameter. Assuming the uniform prior distribution of parameters the model was fit to the number of reported cases [10] [25]. The reported cases were adjusted at each state space using particle Markov Chain Monte Carlo simulation with Metropolis-Hastings updating [26].

The interquartile range, 25th-75th and in the 5-95th percentile ranges were calculated. The model used a minimum of 400 particles (400 - 800) and 30,000 iterations until the overall acceptance value was within the acceptable range of 20 - 30 per cent (Table 1) [27] [28]. The starting timeline and the initial cases for the states varied as the initial case detection was a different period for different states. Credible intervals for the reported cases, forecasted cases and time-varying reproduction number were generated from the posterior distribution samples. Two weeks forecast starting from 5th July 2020 were generated and were validated during the actual reported data during the forecasted period. R-version 4.0 was used for the analysis [29]. C++ using the Rcpp package for computational efficiency, and ggplot2 package was used to produce charts [30].

3. Results

The results show an increasing number of cases for all the states with varying level except in Tamil Nadu, where is expected to remain stable [Figures 1-4]. For every figure, the input incidence data is denoted by black dots the lower left. The time-varying reproductive number $r(t)$ during the same period is denoted.
with black lines and shaded regions in the upper left. The forecasted results for
the \( r(t) \) and the number of cases is illustrated with blue lines and shaded regions.
In all shaded regions, the central line indicates the median, the darker shaded
region indicates the interquartile range and the lighter shaded region indicates
the 5-95th percentile range.

The common result for all the state indicates towards a clear time-varying \( r(t) \)
with more than 1 and showing a stable trend for the forecasted period. The number
of cases is showing an increasing trend in the rest of the three states (Table 2).
All the forecasted figures are within the 25th to 75th percentile/interquartile range.
However, the first week of projection matches more closely with the actual values
in comparison to the following week.

### Table 1. Parameters and Priors.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Prior values</th>
<th>Notes and Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r(0) )</td>
<td>2 - 8: uniform</td>
<td>[26] [31] [32]</td>
</tr>
<tr>
<td>Latent period</td>
<td>[1 - 5] uniform</td>
<td>[8] [33] [34]</td>
</tr>
<tr>
<td>Infectious period</td>
<td>[4 - 12] uniform</td>
<td>[35] [36]</td>
</tr>
<tr>
<td>The initial number of new infections</td>
<td>[1 - 10]</td>
<td>Reported cases, from the respective state government bulletin</td>
</tr>
<tr>
<td>Population</td>
<td>The total state population in million</td>
<td>Kerala-35 Tamil Nadu-67, Andhra Pradesh-49.7, Odisha-46.</td>
</tr>
<tr>
<td>Particles</td>
<td>[100 - 800]</td>
<td>Target acceptable level is 20% - 30% range.</td>
</tr>
</tbody>
</table>

### Table 2. Comparison of Day-wise projections, upper and lower interquartile range with actual reported COVID-19 cases.

<table>
<thead>
<tr>
<th>State</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
<th>11th</th>
<th>12th</th>
<th>13th</th>
<th>14th</th>
<th>15th</th>
<th>16th</th>
<th>17th</th>
<th>18th</th>
<th>1st Week</th>
<th>2nd Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerala</td>
<td>182</td>
<td>201</td>
<td>216</td>
<td>234</td>
<td>251</td>
<td>271</td>
<td>290</td>
<td>313</td>
<td>335</td>
<td>361</td>
<td>388</td>
<td>417</td>
<td>447</td>
<td>480</td>
<td>1645</td>
<td>2741</td>
</tr>
<tr>
<td>Lower</td>
<td>145</td>
<td>147</td>
<td>150</td>
<td>151</td>
<td>152</td>
<td>153</td>
<td>152</td>
<td>151</td>
<td>151</td>
<td>151</td>
<td>149</td>
<td>147</td>
<td>1046</td>
<td>1055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>227</td>
<td>263</td>
<td>302</td>
<td>348</td>
<td>405</td>
<td>437</td>
<td>550</td>
<td>642</td>
<td>759</td>
<td>892</td>
<td>1051</td>
<td>1241</td>
<td>1467</td>
<td>1751</td>
<td>2532</td>
<td>7803</td>
</tr>
<tr>
<td>Reported</td>
<td>225</td>
<td>193</td>
<td>272</td>
<td>301</td>
<td>339</td>
<td>416</td>
<td>488</td>
<td>435</td>
<td>449</td>
<td>608</td>
<td>623</td>
<td>722</td>
<td>791</td>
<td>593</td>
<td>2234</td>
<td>4221</td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>4055</td>
<td>4086</td>
<td>4109</td>
<td>4104</td>
<td>4113</td>
<td>4095</td>
<td>4078</td>
<td>4027</td>
<td>3901</td>
<td>3813</td>
<td>3708</td>
<td>3588</td>
<td>3468</td>
<td>3209</td>
<td>28,640</td>
<td>25,714</td>
</tr>
<tr>
<td>Lower</td>
<td>3098</td>
<td>2878</td>
<td>2683</td>
<td>2482</td>
<td>2296</td>
<td>2114</td>
<td>1952</td>
<td>1795</td>
<td>1645</td>
<td>1510</td>
<td>1389</td>
<td>1259</td>
<td>1160</td>
<td>9540</td>
<td>17,503</td>
<td>18,298</td>
</tr>
<tr>
<td>Upper</td>
<td>5484</td>
<td>5979</td>
<td>6533</td>
<td>7107</td>
<td>7688</td>
<td>8298</td>
<td>9036</td>
<td>9770</td>
<td>11,272</td>
<td>11,948</td>
<td>12,675</td>
<td>13,241</td>
<td>13,640</td>
<td>1406</td>
<td>50,125</td>
<td>73,952</td>
</tr>
<tr>
<td>Reported</td>
<td>4150</td>
<td>3827</td>
<td>3616</td>
<td>3745</td>
<td>4231</td>
<td>3680</td>
<td>3965</td>
<td>4244</td>
<td>4328</td>
<td>4526</td>
<td>4496</td>
<td>4549</td>
<td>4538</td>
<td>4807</td>
<td>27,214</td>
<td>31,488</td>
</tr>
<tr>
<td>Andhra Pradesh</td>
<td>929</td>
<td>980</td>
<td>1027</td>
<td>1072</td>
<td>1120</td>
<td>1173</td>
<td>1227</td>
<td>1276</td>
<td>1332</td>
<td>1391</td>
<td>1442</td>
<td>1500</td>
<td>1544</td>
<td>1595</td>
<td>7528</td>
<td>10,080</td>
</tr>
<tr>
<td>Lower</td>
<td>764</td>
<td>759</td>
<td>786</td>
<td>748</td>
<td>741</td>
<td>731</td>
<td>726</td>
<td>715</td>
<td>702</td>
<td>694</td>
<td>678</td>
<td>669</td>
<td>651</td>
<td>641</td>
<td>5225</td>
<td>4750</td>
</tr>
<tr>
<td>Upper</td>
<td>1115</td>
<td>1242</td>
<td>1383</td>
<td>1535</td>
<td>1707</td>
<td>1901</td>
<td>2120</td>
<td>2359</td>
<td>2622</td>
<td>2926</td>
<td>3257</td>
<td>3623</td>
<td>3999</td>
<td>4438</td>
<td>11,003</td>
<td>23,224</td>
</tr>
<tr>
<td>Reported</td>
<td>998</td>
<td>1322</td>
<td>1178</td>
<td>1062</td>
<td>1555</td>
<td>1608</td>
<td>1813</td>
<td>1933</td>
<td>1935</td>
<td>1916</td>
<td>2432</td>
<td>2539</td>
<td>2602</td>
<td>3963</td>
<td>9536</td>
<td>13,357</td>
</tr>
<tr>
<td>Odisha</td>
<td>405</td>
<td>429</td>
<td>454</td>
<td>478</td>
<td>505</td>
<td>530</td>
<td>560</td>
<td>587</td>
<td>621</td>
<td>652</td>
<td>685</td>
<td>718</td>
<td>755</td>
<td>791</td>
<td>3361</td>
<td>4809</td>
</tr>
<tr>
<td>Lower</td>
<td>350</td>
<td>356</td>
<td>362</td>
<td>367</td>
<td>372</td>
<td>375</td>
<td>377</td>
<td>381</td>
<td>383</td>
<td>385</td>
<td>386</td>
<td>389</td>
<td>389</td>
<td>388</td>
<td>2559</td>
<td>2701</td>
</tr>
<tr>
<td>Upper</td>
<td>467</td>
<td>514</td>
<td>566</td>
<td>625</td>
<td>687</td>
<td>761</td>
<td>841</td>
<td>932</td>
<td>1028</td>
<td>1142</td>
<td>1265</td>
<td>1403</td>
<td>1551</td>
<td>1716</td>
<td>4461</td>
<td>9037</td>
</tr>
<tr>
<td>Reported</td>
<td>469</td>
<td>456</td>
<td>571</td>
<td>527</td>
<td>577</td>
<td>755</td>
<td>570</td>
<td>595</td>
<td>616</td>
<td>543</td>
<td>618</td>
<td>494</td>
<td>718</td>
<td>591</td>
<td>3925</td>
<td>4175</td>
</tr>
</tbody>
</table>
4. Discussion

The study covered nearly 200 million population at risk. Though all the states are in the coastal region of India are different in language, socioeconomic dynamics and health system capacity [9] [37]. Though the first case of COVID was reported from Kerala in late January, most of the other states in the study reported the first cases in March. This analysis uses reported cases from the public health system, shows the heterogeneity of the epidemic movement through the time-varying effective reproductive rate and near-future forecast of COVID-19 burden, which closely matches with the actual number of cases.

**Figure 1.** Two weeks forecast: Time varying reproduction number and new cases in Kerala.

**Figure 2.** Two weeks forecast: Time varying reproduction number and new cases in Tamil Nadu.
4.1. Epidemic Model and Validity

The results from four different states show that the model is robust and can be deployed for other states too. The variations in the actual number and the forecasted figures need to be analysed with the perspective of finding reasonable answers [38]. Most of the models are based on the assumption that “given current situation remains the same for future”, which is difficult to achieve, particularly concerning changing policy in testing and response strategy in the states [39] [40]. Changing case definitions, testing strategy and response to the pandemic is known to have influenced our understanding of the trajectory of the
epidemic not only in India but also around the world [41]. The social distancing and lock down measures had shown to reduce the contagion, which was reported to be reversed once the started to open up [42] [43]. As this model demonstrates that the near future cases can be predicted with close certainty, repeated application of the model in constant time intervals can provide vital information on the projected number of cases in short runs and thus used for deployment of mitigation strategies. The criticism that outputs of mathematical models are not always useful may be considered in the right spirit, with a reasonable understanding of the models and the need for thinking beyond the one-time application of models [44] [45].

4.2. Effective Reproduction Number

The effective reproductive number indicates the risk of the epidemic in a given point of time. Though \( r(t) \) lower than unit indicates towards the loss of force in the epidemic, it remains sensitive to the fact that most of the population is not infected thus providing little protection, long way from herd immunity and thus the potential risk of transmission. It is also important to mention that \( r(t) \) lower than unit does not exclude the potential of localised outbreaks, which can influence the epidemic trajectory [33] [46]. However, the trend in \( r(t) \) does provide advance information on the highly unpredictable nature of this epidemic.

4.3. Need for Better Data

Any model output is as good as the data is as the assumptions are. The details about the day of symptom onset today of sample collection and the day of diagnosis and day of reporting was not available in the respective public domains of the states, thus were not considered in the model. Availability of this information may improve the model performance further. Also, as the capacity for testing was increased gradually the delay between the sample collection and testing had declined [3]. Having local data for each state with specific serial interval or generation time could have helped in the improvement of this analysis. The other published studies from India have also relied on external information source [7] [47]. It is expected that those who will be replicating the study can focus on further improving the model by addressing the gaps. The states’ response was mixed with some states going for only institutional quarantine while others going for home and institutional quarantine. An additional limitation of this model is reliant on the published information, which was dependence on the published case reports, thus sensitive to the local level variation in the testing strategy.

4.4. Consensus on Models

There are many types of forecasting and predictive modelling has been published on the COVID epidemic in India [48] [49]. All these models are relevant in the scientific quest and add value to the knowledge base of this novel virus.
However, different models will provide different results creating confusion, and making it difficult for the policymakers to take decisions. However, there are other epidemic response initiatives in India which focuses on a single tool for the epidemic projection leading to harmony in the response [50]. The other model of managing a large number of real-time scientific publication is by aggregating them all and adding mutual accountability [51]. Most importantly the real-time forecasts can happen only if it is repeated again and again regular intervals [14].

5. Conclusion

The real-time short-term forecasting used in the four states provides a good approximation of the near future epidemic trajectory. The tool is available in the public domain and needs to be used on a given interval repeatedly to ensure tracking of the epidemic at the local level. This can be used for understanding the future burden.

Acknowledgements

The constant guidance and sharing of original codes by Jason Asher are highly appreciated. The paper also acknowledges the critical inputs of Dr. Shailaja Tetali and Dr. Jammy Rajesh.

Ethics Statement

The analysis is done using publicly available secondary data and no patient information was taken for analysis, thus ethical clearance was not required.

Conflicts of Interest

The authors did not receive any funding from any agency or individual for the work and the authors declare no conflict of interest.

References


[40] Covid Testing Numbers in Karnataka Down Sharply and So Are Fresh Cases. The Economic Times. 

https://doi.org/10.1016/S2468-2667(20)30089-X

https://doi.org/10.1016/S0140-6736(20)31689-5


https://doi.org/10.1056/NEJMp2016822

https://doi.org/10.3201/eid2501.171901

https://doi.org/10.4103/ijph.IJPH_502_20

https://doi.org/10.1017/dmp.2020.115

https://doi.org/10.1016/j.scitotenv.2020.138762


[51] COVID-19 Forecasts: Cumulative Deaths CDC. 