An infectious diseases hazard map for India based on mobility and transportation networks

Onkar Sadekar, Mansi Budamagunta, G. J. Sreejith, Sachin Jain and M. S. Santhanam*

Indian Institute of Science Education and Research, Pune 411 008, India

We propose a risk measure and construct an infectious diseases hazard map for India. Given an outbreak location, a hazard index is assigned to each city using an effective distance that depends on inter-city mobilities instead of geographical distance. We demonstrate its utility using an SIR model augmented with air, rail and road data among the top 446 cities. Simulations show that the effective distance from outbreak location reliably predicts the time of arrival of infection in other cities. The hazard index predictions compare well with the observed spread of SARS-CoV-2. This hazard map can be used in other outbreaks as well.

Keywords: COVID-19, effective distance, hazard map, infectious diseases, transportation networks.

As of July 2021, more than 19 crore people – about one in every 40 humans – have been infected by SARS-CoV-2, and about 39 lakh people have died[1]. COVID-19 has escalated from a cluster of cases in China in the late 2019 into an unprecedented global public health crisis. In India, starting with a few cases in February 2020, the infection had spread to about 65 lakh people in a span of about eight months when the first wave peaked. With the resurgence of the second wave in India since March 2021, the number of infections and deaths has witnessed a steep increase[2].

The Spanish flu of 1918 was one of the biggest pandemics to hit India, arriving in Bombay with the British-Indian army returning from the First World War in Europe[3]. The then Annual Report of the Sanitary Commissioner to the Government of India observed that 'There is ample evidence during the first epidemic of the introduction of infection into a locality from another infected locality. The railway played a prominent part, as was inevitable. During the panic caused by the epidemic, the trains were filled with emigrants from infected centres, many of them being ill. The Post office also was an important agency in disseminating infection, also very largely through the Railway Postal Service. Lucknow, Lahore, Simla and other cities are said to have been infected in this manner'[4]. Further, the Report states that 'there is ample evidence to prove that infection in India

during the second epidemic was carried from province to province and place to place within each province by travellers by rail, riverboats, carts and on foot'[4]. This mode of spread is also confirmed by other studies based on detailed data recorded then in Bombay and other provinces of British India[3]. Nearly one century after the Spanish flu, long-distance travel is even more common. This has resulted in rapid spread of infections to remote corners of the world[5–7]. It is expected that irrespective of the innate capacity of a virus to infect, the spread from one geographical area to another is primarily caused by the mobility of the people[8–11].

The influence of transportation on the pattern of infection spread is evident in SARS-CoV-2 and earlier infectious diseases[12–15]. One might identify two concurrent but distinct processes: (i) evolution of infection within a small, well-mixed geographical region (city/town), and (ii) intercity transmission between the regions. The latter will depend crucially on the transport networks and the mobility patterns of people within the country[16–19]. A rather impractical limit is when transportation systems are entirely stopped leading to suppression of infection spread. Most modelling efforts focused on prediction of caseloads in India, rather than geographical spreading patterns[20–25].

In this study, we propose an infectious diseases hazard map for India based on a reliable predictor of the arrival time of infections from a known outbreak location[12]. Though the first official COVID-19 cases were detected in Kerala, significant outbreaks (several hundred cases) were reported in April 2020 from Mumbai and Delhi[2]. Being large transport hubs, the infection quickly spread into the rest of the country from these two cities. While Mumbai or Delhi could be the outbreak location now, in a general scenario, it can be anywhere. It is natural to define hazard indices for every city/town based on different potential outbreak locations.

Let us pose the following question: Consider a network of M cities/towns $(X_1, X_2, X_3, \ldots, X_M)$, and the outbreak location as X_1: then can a hazard value be assigned to other cities/towns reflecting not their geographical proximity, but an 'effective proximity' incorporating mobility patterns? We discuss one solution and validate it using models incorporating extensive transportation network data.

*For correspondence. (e-mail: santh@iiserpune.ac.in)

Note that the proposed hazard index (on which the hazard map is based) depends on the outbreak location and mobility patterns. The latter is a time-dependent factor. However, as the number of cases does not appreciably change in less than a day and the data are made public only on a granularity of a day, we construct the hazard map assuming mobility averaged over a few days to be representative for all times. For a hazard map at a subcontinental spatial scale such as India, each city/town is assumed to be well-mixed. In this study, the mobility data are applied to obtain a hazard map for 446 cities/towns with a population more than one lakh[26].

Augmented SIR model framework

Our framework is based on the susceptible–infected–recovered (SIR) compartmental model augmented with connectivity information between towns and cities[12,27]. For a well-mixed population, the SIR model is given as[28,29]

$$\frac{\partial S(t)}{\partial t} = -\alpha \frac{S(t)I(t)}{N},$$

$$\frac{\partial I(t)}{\partial t} = +\alpha \frac{S(t)I(t)}{N} - \beta I(t),$$

$$\frac{\partial R(t)}{\partial t} = +\beta I(t). \tag{1}$$

Here, $S(t)$, $I(t)$ and $R(t)$ denote the susceptible, infected, and recovered population respectively, at time t. α and β denote the infection and recovery rate respectively. The total population $N = S(t) + I(t) + R(t)$ remains constant over time. However, the population in a large region like India is not well-mixed. But the population within each city/town can be assumed to be well-mixed and eqs (1) are applicable within each of M cities/towns. A small part of this population can travel between cities/towns according to

$$\frac{\partial N_n(t)}{\partial t} = \sum_{m=1}^{M} [F_m^n - F_n^m], \ n, m = 1, 2, ..., M, \tag{2}$$

where $N_n(t)$ denotes the population of the nth city/town at time t and F_n^m denotes the number of people travelling from n to m in unit amount of time. F_n^m together with the convention $F_n^n = 0$ defines the traffic matrix. The population of a city will be constant if its total influx and outflux are equal. In this study the traffic matrix is inferred from limited, available real-life data, and generally they do not strictly satisfy this balance. However, the timescales over which our simulations have been performed are small enough that such imbalances in fluxes do not change the city populations appreciably; accordingly, we assume them to be constant in the rest of the work.

Little reliable data are available regarding infection acquired during transit. The probability of getting infected during transit is assumed to be zero. In our model, a susceptible traveller leaving city n would remain so upon reaching city m (similarly for infected or recovered travellers). With these assumptions, the SIR model incorporating inter-city mobilities can be written as[30]

$$\frac{\partial S_n(t)}{\partial t} = -\alpha \frac{S_n(t)I_n(t)}{N_n}$$

$$+ \sum_m \left[\frac{F_m^n}{N_m} S_m(t) - \frac{F_n^m}{N_n} S_n(t) \right],$$

$$\frac{\partial I_n(t)}{\partial t} = +\alpha \frac{S_n(t)I_n(t)}{N_n} - \beta I_n(t)$$

$$+ \sum_m \left[\frac{F_m^n}{N_m} I_m(t) - \frac{F_n^m}{N_n} I_n(t) \right],$$

$$\frac{\partial R_n(t)}{\partial t} = +\beta I_n(t) + \sum_m \left[\frac{F_m^n}{N_m} R_m(t) - \frac{F_n^m}{N_n} R_n(t) \right],$$

$$n, m = 1, 2, ..., M. \tag{3}$$

Upon adding the above set of equations, we find that the total city population $S_n(t) + I_n(t) + R_n(t) = N_n$ is a constant up to small deviations on account of the imbalance between influxes and outfluxes. These equations extend the SIR model to a network in which the population is well-mixed only within each city. These equations provide one of the well-studied among several models of large-scale infection spread[31,32].

Note that eq. (3) is not India-specific and has been applied on a global scale[12,30]. In rest of this study, eq. (3) will be the central framework supplemented with India-specific traffic matrix. Estimating the entries F_n^m of the traffic matrix (**F**) is particularly difficult due to the insufficient availability of real data, the details of which are elaborated in the Supplementary Material[34].

Transportation network and data

Hereafter, the discussion will be India-specific. We include air, rail and road data in the traffic matrix; inland waterways and other modes are ignored. A directed network of cities/towns with a population above 1 lakh (according to the 2011 census)[26] and having at least one of air, rail or road connectivity is created. The network has $M = 446$ nodes (cities/towns), and 46,448 weighted edges. Each pair of cities can have up to two oppositely directed edges between them, with weights representing the total traffic

(all modes) in that direction. Further details of the edge data are given in the <u>Supplementary Material</u>[34].

The air, rail and road transportation data are combined to obtain the averaged daily traffic matrix **F**, whose element F_n^m represents the net direct traffic (number of people) from city n to m on a 'typical' day. We ignore any effect of the differences in the travel times associated with different modes of traffic (for instance, air travel being faster than road travel). Figure 1 shows the 500 busiest intercity routes based on the sum of forward and backward traffic.

The matrix **F** constructed from real data is not symmetric: $F_n^m \neq F_m^n$, i.e. the forward and backward traffic between n and m is unequal. The line thickness in Figure 1 indicates its weight – thicker lines represent more traffic. We also summarize some of the key statistics of transportation networks in Table 1.

Infectious diseases hazard index

The central idea in constructing the hazard map is the notion of 'effective distance' introduced by Brockmann and Helbing[12]. If F^n and F_n, are the net rates of people travelling in and out of city n, then the one-step conditional probability that a person leaving city n travels to m is given by

$$P_n^m = \frac{F_n^m}{F_n}. \tag{4}$$

We define pair distance d_n^m from city n to m as

$$d_n^m = 1 - \log P_n^m. \tag{5}$$

If the cities are not directly connected, i.e. nobody travels from n to m directly, then $P_n^m = 0$ and $d_n^m \to \infty$. In contrast, large traffic between the cities (relative to the population of the origin n) makes d_n^m small. Note that d_n^m is not necessarily symmetric between the cities.

The fastest path for an infection may pass through other cities. This motivates the notion of an effective shortest distance between any pair of cities as follows: For any path Γ_n^m (a sequence of cities starting and ending at n and m) through the network between n and m, $\lambda(\Gamma_n^m)$ represents the sum of the pair distances between successive cities. The effective distance D_{eff}^{nm} between a pair of cities is defined as the shortest among all paths Γ_n^m.

$$D_{\text{eff}}^{nm} = \min_{\text{all paths } \Gamma_n^m} \lambda(\Gamma_n^m). \tag{6}$$

Infection spread between the cities is likely to depend on the traffic between them, rather than geographical distances. The effective distance depends on the mobi-

lities rather than the geographical distance. By definition, it takes into account the multiple paths that may connect a pair of cities, just as the infection may reach a city through another one rather than directly from the outbreak. It is therefore natural to expect a high correlation between aspects of infection spread and D_{eff}.

To make this precise, we define the 'time of arrival' T_A^{nm} of the infection in a city m (from a given outbreak location n) as the first time when the number of (active) infected cases crosses a predefined threshold I^c. In studies of infection spread through global air-traffic patterns, time of arrival at location m from an outbreak location n was found to be proportional to the effective distance D_{eff}^{nm} between them[12]. Naively T_A^{nm} between cities, which is obtained by solving eq. (3) is expected to have a complex dependence on the traffic. It is surprising that T_A^{nm} can instead be reliably predicted from a simple functional D_{eff}^{nm}.

Figure 1. Averaged composite transportation network estimated based on data from pre-COVID years 2017–2019. The lines represent the busiest 500 connections between cities and their thickness is proportional to the total volume of traffic in the forward and backward directions between each pair of cities.

Table 1. Properties of transportation network and mobility data used in this study. Not surprisingly, air travel constitutes a small fraction of the overall mobility. Majority of the long-distance travel is accounted for by trains, and road is the dominant mode for short-distance travel

Property	Air	Rail	Road	Combined
Number of nodes	85	435	446	446
Number of edges	1182	41,594	9128	46,448
Average degree	13	95	20	104
Passengers per day	7.5×10^5	8.8×10^6	2.5×10^6	1.2×10^7
Fraction of total	0.06	0.73	0.21	1.0

Extensive simulations performed in this study and summarized in Figure 2 show that for a wide range of realistic α and β and Indian traffic patterns, T_A has a high linear correlation with D_{eff}. Predicting the arrival of infection at a given location is not only of academic interest but is also of immense practical value. In the rest of the study, we present our analysis of this idea using Indian traffic data.

Given an outbreak location, the risk of infection in another location can be quantified in many ways, time of arrival being a natural one. Under reasonable and realistic assumptions, including uniform infection parameters, D_{eff} provides a reliable and robust predictor of T_A, as evident from our simulations. D_{eff} can be mapped using available transportation data, unlike T_A that can be obtained either from extensive simulations or a posteriori knowledge of infection spread.
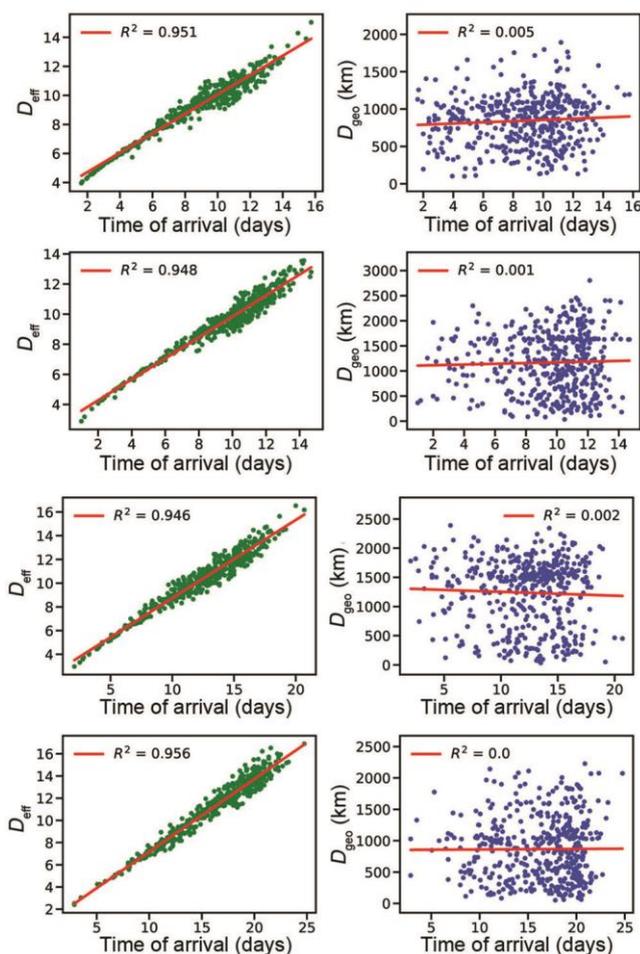

Figure 2. Plots showing strong linear correlation between effective distance D_{eff} and time of arrival T_A. Left column shows $D_{\text{eff}}^{i_0 m}$ plotted against T_A of the infection at city m from outbreak at city i_0. T_A is obtained by solving eq. (3) with infection parameters $\alpha = 1.5$, $\beta = 1.0$ and $I^c = 10$. Outbreak locations considered in the rows from top to bottom are Delhi, Mumbai, Patna and Tirupati. Right column shows the geographical distance $D_{\text{geo}}^{i_0 m}$ from i_0 plotted against T_A. R^2 is a measure of goodness of the linear fits (red lines), with $R^2 = 1$ being a perfect linear fit.

Results

In order to validate the utility of the effective distance, we numerically evolve the coupled differential eq. (3) using fourth-order Runge–Kutta method. The initial infected population I_{i_0} ($t = 0$) in the outbreak city i_0 is taken to be a fraction (0.0001) of the local population. We perform such simulations for different choices of the outbreak locations and infection parameters. T_A for each city is evaluated in each case by finding the time when the infected population in that city crosses a threshold, I^c is taken to be 10. Qualitative results are independent of choices of I_{i_0} and I^c.

In Figure 2 we show the results assuming infection parameters $\alpha = 1.5$, $\beta = 1.0$ giving $R_0 = 1.5$, a typical value that was witnessed for SARS-CoV-2 (ref. 33). In Figure 2 (left panel), the effective distance $D_{\text{eff}}^{i_0 m}$, where i_0 is the outbreak location, is plotted against the time of arrival at city m. This is shown for four different outbreak locations of varying size, namely Delhi, Mumbai, Patna and Tirupati. We find a good linear relation between $D_{\text{eff}}^{i_0 m}$ and $T_A^{i_0 m}$, as indicated by high $R^2 \simeq 0.94$. These are in striking contrast to the right panel in Figure 2 which shows T_A against the geographical distance from the outbreak.

Similar observations were made by Brockmann and Helbing[12], who considered key global air traffic patterns alone.

Remarkably, within India, considering multiple modes of transport, with air travel being the least popular mode accounting for less than 10% of relevant mobility, the linearity holds good. Smaller D_{eff} to the outbreak then suggests a higher risk to a city, manifested as earlier arrival of the infection. The demonstration of the ability of D_{eff} to predict the time of arrival is a key result of this study.

Table 2 shows T_A for the same outbreak locations as in Figure 2. In each case, the list of the top cities in terms of risk (i.e. smallest D_{eff}) is also shown. For outbreaks from poorly connected cities, the surrounding regions face the first brunt of infection; followed by bigger cities. On the other hand, outbreaks from big metros which are well connected, quickly reach far corners. For instance, infection from Tirupati reached Bengaluru in ~5 days, whereas the outbreak from Mumbai or Delhi spread to Bengaluru in ~2.5 days. The hazard map in Figure 3 shows this visually for the same four outbreak locations. The size of the circles represents the hazard (larger the circle, greater the risk). The hazard (i.e D_{eff}) is easily estimated for all the cities; only the top 10 cities are shown to avoid clutter.

It is interesting to consider the hazard map assuming that only one mode of transport is operating. The transportation mode-specific hazard map is shown for two outbreak locations, Bengaluru (Figure 4) and Guwahati (Figure 5). As expected, air traffic takes the infection to distant big cities, whereas road traffic restricts the infection in geographical proximity. When all the data are

Table 2. Time of arrival (T_A; in days), for each of the four outbreak locations in Figure 2, showing cities with 10 largest values of T_A. The parameters used are $\alpha = 1.5$, $\beta = 1.0$ and $I^c = 10$

| Delhi			Mumbai			Patna			Tirupati		
City	D_{eff}	T_A	City	D_{eff}	T_A	City	D_{eff}	T_A	City	D_{eff}	T_A
Kanpur	3.96	1.62	Thane	2.89	1.00	Gaya	2.98	2.06	Chittoor	2.41	2.88
Mumbai	4.06	1.69	Pune	3.17	1.19	Dinapur Nizamat	3.32	2.50	Chennai	2.53	2.88
Gurgaon	4.25	1.94	Delhi	3.7	1.62	Arrah	3.58	2.75	Hyderabad	3.04	3.50
Lucknow	4.33	2.00	Surat	4.07	2.00	Delhi	3.75	2.81	Vellore	4.23	5.31
Faridabad	4.34	2.00	Ahmedabad	4.08	2.00	Bhagalpur	3.99	3.25	Bengaluru	4.25	5.06
Jhansi	4.54	2.19	Pimpri Chinchwad	4.25	2.19	Kolkata	4.09	3.25	Tiruvannamalai	4.50	5.81
Rohtak	4.58	2.31	Nashik	4.33	2.25	Darbhanga	4.44	3.88	Kadapa	4.84	6.50
Ludhiana	4.70	2.38	Vasai	4.43	2.38	Jehanabad	4.47	3.94	Vijayawada	4.89	6.62
Moradabad	4.70	2.44	Vasco Da Gama	4.47	3.00	Begusarai	4.57	4.00	Anantapur	5.00	6.75
Bengaluru	4.71	2.38	Bengaluru	4.49	2.38	Biharsharif	4.60	3.94	Madanapalle	5.01	6.75
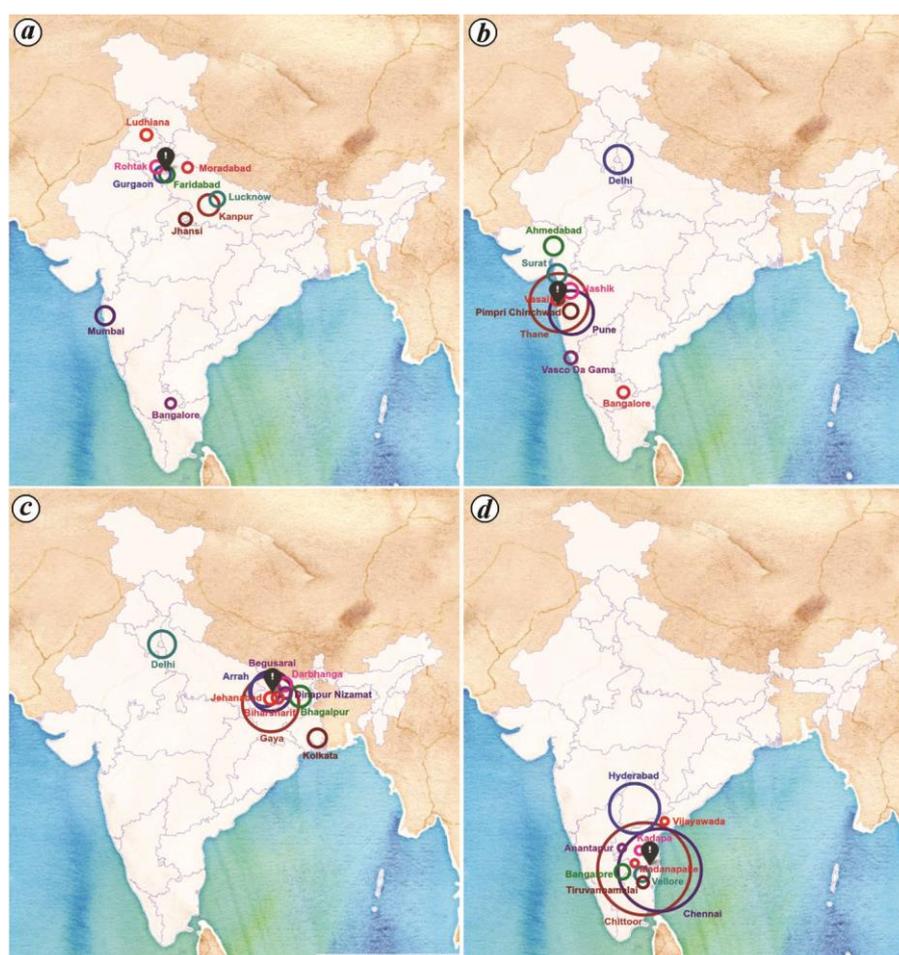
Figure 3. A visual depiction of the information in Table 2 in the form of an infectious diseases hazard map, with outbreak locations at (**a**) Delhi, (**b**) Mumbai, (**c**) Patna and (**d**) Tirupati (shown as black-coloured location icon). The radius of the circle is proportional to the hazard index of the city/town. Larger the circle, greater is the hazard and their colour does not carry any information. Only the cities/towns with the top-ten hazard values are shown.

combined, the map is largely influenced by rail and road traffic patterns due to their higher contribution to the total traffic (Figures 4 *d* and 5 *d*).

Earlier works which used mobility to study the spread have exclusively used airline mobility, which is justified in the global context[12]. India has not just one of the largest railway networks, but it is also used by a significant fraction of people. Hence, an analysis of the type presented here is most desirable in the Indian context.

In Table 3, the results of our framework are compared with real data of T_A for the first wave of the SARS-CoV-2 pandemic.

Table 3. Comparison of the time of arrival of infection (T_A; days) for real data and that from the simulation framework proposed in this study. The top 12 cities at most risk obtained through D_{eff} and real-life data are shown. The ranks based on T_A for the top-12 cities according to D_{eff} are also given. Correspondingly, a list is prepared for the top-12 cities based on T_A. The rows in bold denote cities common to both lists. Note that 9 out of 12 cities (~75%) are common to both the lists, showing that the proposed framework has predictive capability. Mumbai is taken to be the outbreak location and real-life data is at the granularity of districts

| Rank based on D_{eff}	City	D_{eff}	Rank based on T_A	Rank based on T_A	City	T_A (days)	Rank based on D_{eff}
1	**Thane**	**2.88**	**4**	1	**Delhi**	**11**	**3**
2	**Pune**	**3.18**	**5**	2	**Ahmedabad**	**13**	**5**
3	**Delhi**	**3.70**	**1**	3	**Chennai**	**16**	**12**
4	Surat	4.06	222	4	**Thane**	**25**	**1**
5	**Ahmedabad**	**4.08**	**2**	5	**Pune**	**46**	**2**
6	**Nashik**	**4.29**	**10**	6	**Hyderabad**	**57**	**10**
7	Vasai	4.42	No Data	7	**Bangalore**	**65**	**9**
8	Vasco Da Gama	4.47	No Data	8	Guwahati	70	55
9	**Bangalore**	**4.49**	**7**	9	**Kolkata**	**79**	**11**
10	**Hyderabad**	**4.62**	**6**	10	**Nashik**	**85**	**6**
11	**Kolkata**	**4.91**	**9**	11	Guntur	88	141
12	**Chennai**	**4.93**	**3**	12	Kurnool	89	131
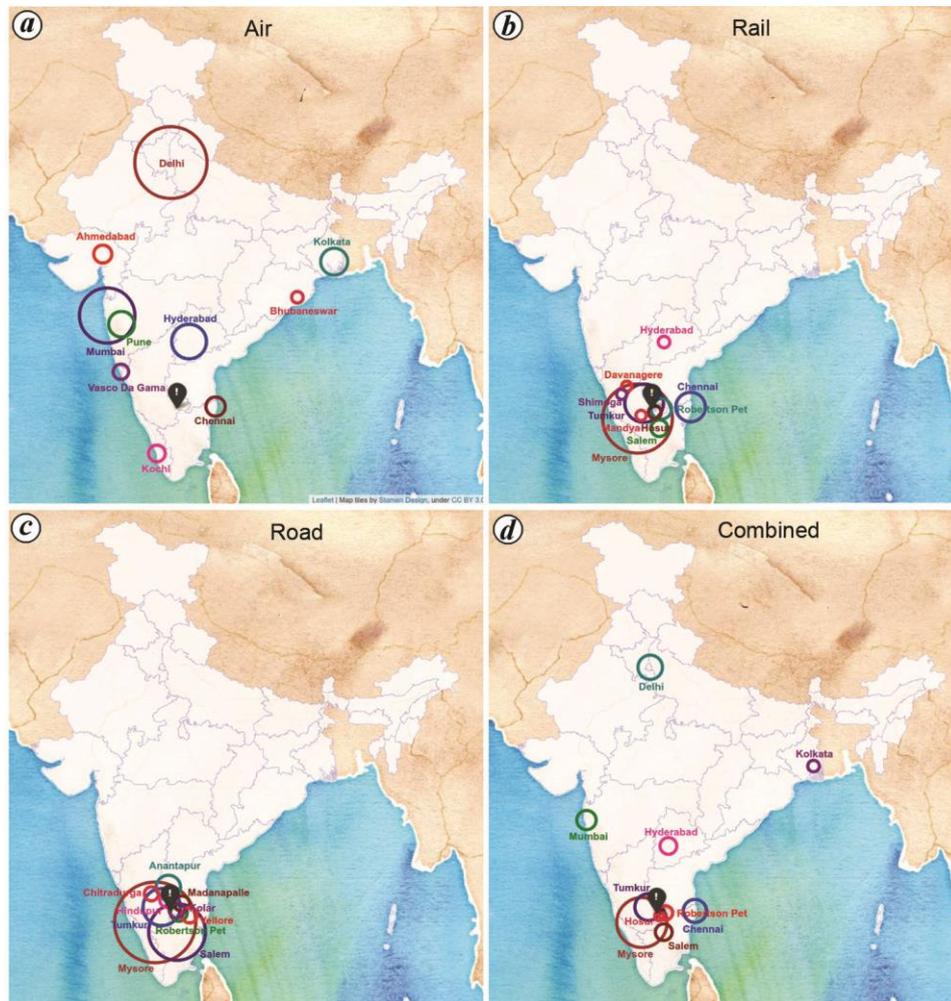
Figure 4. Transportation mode-specific hazard maps with Bengaluru as the outbreak location. Corresponding to (**a**) air, (**b**) rail, (**c**) road and (**d**) combined modes of transport. The radius of the circle is proportional to the hazard index of the city/town. Larger the circle, greater is the hazard and their colour does not carry any information. Only the cities/towns with the top 10 hazard values are shown.
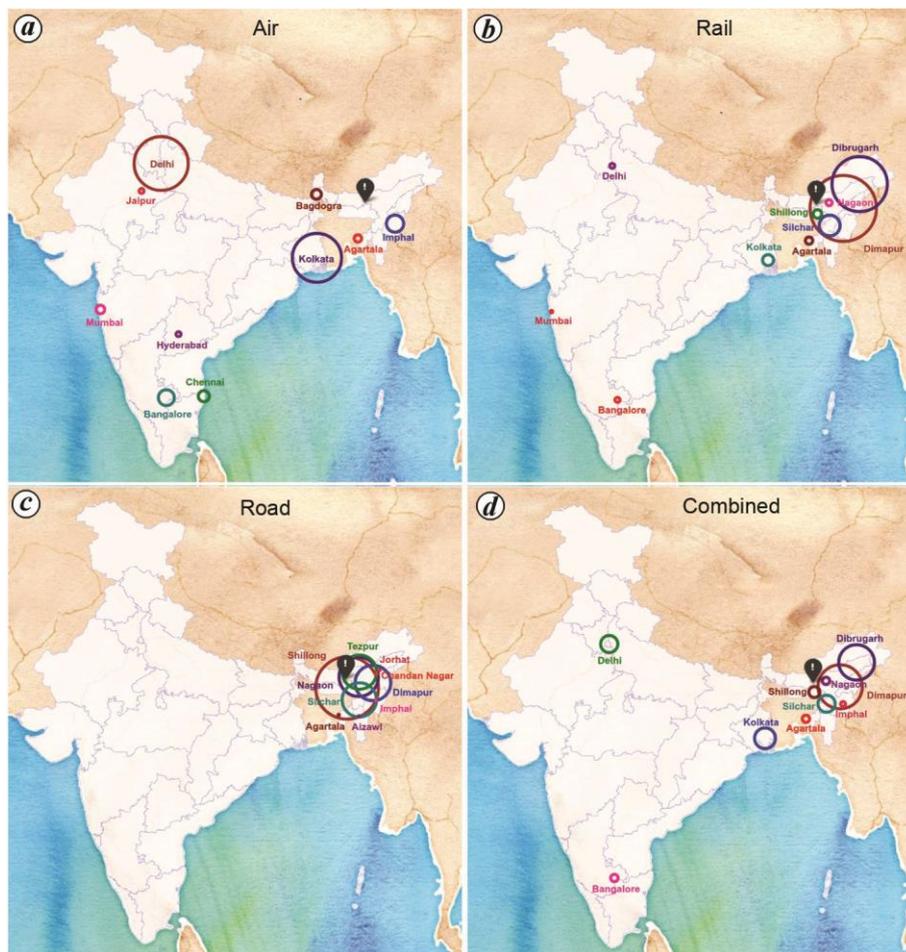
Figure 5. Transportation mode-specific hazard maps with Guwahati as the outbreak location. Corresponding to (*a*) air, (*b*) rail, (*c*) road and (*d*) combined modes of transport. The radius of the circle is proportional to the hazard index of the city/town. Larger the circle, greater is the hazard and their colour does not carry any information. Only the cities/towns with the top 10 hazard values are shown.

District-wise data are available from 26 April 2020 (ref. 2). Mumbai crossed the threshold first and has been taken as the outbreak location. There were ~4000 active cases in Mumbai on 26 April 2020. The time of arrival (T_A) for a city is when its three-day average caseload crosses a threshold taken to be $I^c = 4000$. Table 3 presents a comparison of the real data with predictions from D_{eff}. In Table 3 (left), the top 12 cities at most risk based on the D_{eff} framework are shown. This is compared with their ranks in terms of real-life time of arrival of infections (T_A). We find that 9 out 12 cities also appear in the top-12 based on estimates from T_A. To present a different means of comparison, Table 3 (right) shows the top-12 cities based on time of arrival. Again, we find that 9 out of 12 cities appear in the top-12 based on D_{eff}. It is remarkable, given the uncertainties in the traffic data and the approximations made to fill in missing data, that ~75% of cities obtained from simulations match with those in the list obtained from real data. This provides the proof-of-concept that it is possible to develop a systematic predictive framework to objectively estimate the risk in Indian cities[34].

Conclusion

If an infectious disease breaks out in one city, how long does it take to reach other cities and towns? This length of time can be a simple measure of the risk in other cities – the longer it takes, the lesser the risk. One may estimate this from careful simulations involving detailed traffic patterns. However, this time is easily predicted by a quantity called the effective distance, which can be calculated if we know the prevailing traffic patterns. Larger the effective distance of a city from an outbreak location, lower is its risk of early infection.

Based on this idea, we have constructed an infectious disease hazard map for India using data from the intercity transportation network in the country. Further details about the map can be found at https://www.iiserpune.ac.in/~hazardmap/.

Real data from air, road and rail transportation networks between the most populous 446 Indian cities were used in this calculation. We relied on publicly available data sources and used simple assumptions and algorithms to fill in the missing attributes of typical Indian traffic patterns.

We used extensive simulations to validate the usefulness of the idea. We find, in agreement with similar past studies, that the effective distance of a city, from the origin is proportional to the time of appearance of first infections in that city and is thus a reliable measure of its risk. Comparison with the early patterns of spread of COVID-19 in India showed surprisingly good agreement between the predictions from effective distances and real data. This adds further credence to the idea of effective distance.

The results of this study prompt several interesting, questions, both of academic and practical value. While effective distance predicts relative order in which cities are affected, the rate of spread through this sequence is determined by details of the infection parameters (α, β) as well as average mobility. A conceptual framework that explains the empirical observations regarding these (from simulations) is missing. Moreover, a good explanation for the linear relationship between the effective distance and time of arrival may be needed in order to know the limits of its applicability. Lastly, a generalization of the notion of effective distance to a scenario of multiple outbreak locations will make this an invaluable tool in designing efficient mitigation measures – for instance in determining which traffic routes to close down with higher priority.

1. https://covid19.who.int/, https://coronavirus.jhu.edu/map.html (accessed in July 2021).
2. https://www.covid19india.org/ (accessed in April 2021).
3. Mills, I. D., The 1918–1919 influenza pandemic – the Indian experience. *Indian Econ. Soc. Hist. Rev.*, **23**, 1, 1986; https://doi.org/10.1177%2F001946468602300102
4. The Annual Report of the Sanitary Commissioner with the Government of India (1918), Calcutta, 1920.
5. Gautreau, A., Barrat, A. and Barthelemy, M., Global disease spread: statistics and estimation of arrival times. *J. Theor. Biol.*, 2008, **251**, 509–522.
6. Barthlemy, M., Spatial networks. *Phys. Rep.*, 2011, **499**, 1101.
7. Feng, L., Zhao, Q. and Zhou, C., Epidemic in networked population with recurrent mobility pattern. *Chaos Solitons Fract.*, 2020, **139**, 110016.
8. Belik, V., Geisel, T. and Brockmann, D., Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev.* 2011, **X1**, 011001.
9. Coltart, C. E. and Behrens, R. H., The new health threats of exotic and global travel. *Br. J. Gen. Pract.*, 2012, **62**, 512–513.
10. Helbing, D., Globally networked risks and how to respond. *Nature*, 2013, **497**, 51–59.
11. Barbosa, H. *et al.*, Human mobility: models and applications. *Phys. Rep.*, 2018, **734**, 174.
12. Brockmann, D. and Helbing, D., The hidden geometry of complex, network–driven contagion phenomena. *Science*, 2013, **342**, 1337–1342.
13. Arenas, A. *et al.*, Modeling the spatiotemporal epidemic spreading of COVID-19 and the impact of mobility and social distancing interventions. *Phys. Rev. X*, 2020, **10**, 041055.
14. Althouse, B. M., Wenger, E. A., Miller, J. C., Scarpino, S. V., Allard, A., Hbert-Dufresne, L. and Hu, H., Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2, 2020; arXiv:2005.13689.
15. Garcia-Gasulla, D. *et al.*, Global Data Science Project for COVID-19 Summary Report, 2020, arXiv:2006.05573.
16. Mishra, R., Gupta, H. P. and Dutta, T., Analysis, modeling and representation of COVID-19 spread: a case study on India. *IEEE Trans. Comput. Soc. Syst.*, 2021, doi:10.1109/TCSS.2021.3077701.
17. Sarkar, K., Khajanchi, S. and Nieto, J. J., Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fract.*, 2020, **139**, 110049.
18. Pujari, B. S. and Shekatkar, S., Multi-city modeling of epidemics using spatial networks: application to 2019-nCov (COVID-19) coronavirus in India. medRxiv, 2020; https://doi.org/10.1101/2020.03.13.20035386.
19. Gupta, S. *et al.*, An India-specific compartmental model for COVID-19: projections and intervention strategies by incorporating geographical, infrastructural and response heterogeneity; 2020, arXiv:2007.14392.
20. Gopal, R., Chandrasekar, V. K. and Lakshmanan, M., Dynamical modelling and analysis of COVID-19 in India. *Curr. Sci.*, 2021, **120**(8), 1342–1349.
21. Bedi, P., Dhiman, S., Gole, P. and Jindal, V., Prediction of COVID-19 trend in India and its four worst-affected states using modified SEIRD and LSTM models. *SN Comput. Sci.*, 2021, **2**, 224.
22. Khajanchi, S. and Sarkar, K., Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India. *Chaos: Interdiscip. J. Nonlinear Sci.*, 2020, **30**, 071101.
23. Das, A., Dhar, A., Goyal, S., Kundu, A. and Pandey, S., COVID-19: analytic results for a modified SEIR model and comparison of different intervention strategies. *Chaos Solitons Fract.*, 2021, 110595.
24. Jha, V., Forecasting the transmission of COVID-19 in India using a data driven SEIRD model, 2020, arXiv:2006.04464.
25. Khajanchi, S., Sarkar, K., Mondal, J. and Perc, M., Dynamics of the COVID-19 pandemic in India, 2020, arXiv:2005.06286.
26. https://censusindia.gov.in/2011-common/censusdata2011.html (accessed in July 2020).
27. Gong, Y., Song, Y. and Jiang, G., Epidemic spreading in metapopulation networks with heterogeneous infection rates. *Physica A*, 2014, **416**, 208–218.
28. Ronald, R. and Hilda, P., An application of the theory of probabilities to the study of a priori pathometry – Part III. *Proc. R. Soc. London A*, 1917, **93**, 225–240.
29. Kermack, W. O. and McKendrick, A. G., A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London Ser. A*, 1927, **115**, 700–721.
30. Colizza, V., Pastor-Satorras, R. and Vespignani, A., Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Phys.*, 2007, **3**, 276–282.
31. Taylor, D., Klimm, F., Harrington, H. A., Kramr, M., Mischaikow, K., Porter, M. A. and Mucha, P. J., Topological data analysis of contagion maps for examining spreading processes on networks. *Nature Commun.*, 2015, **6**, 111.
32. Toli, D., Kleineberg, K.-K. and Antulov-Fantulin, N., Simulating SIR processes on networks using weighted shortest paths. *Sci. Rep.*, 2018, **8**, 110.
33. Marimuthu, S., Joy, M., Malavika, B., Nadaraj, A., Asirvatham, E. and Jeyaseelan, L., Modelling of reproduction number for COVID-19 in India and high incidence states. *Clin. Epidemiol. Global Health*, 2021, **9**, 57–61.
34. https://www.iiserpune.ac.in/~hazardmap/ (The Supplementary Material can found on this site).
35. Leaflet – map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.