

# Estimation of COVID-19 recovery and decease periods in Canada using machine learning algorithms

Subhendu Paul<sup>1,\*</sup> and Emmanuel Lorin<sup>1,2</sup>

<sup>1</sup>School of Mathematics and Statistics, Carleton University, Ottawa, K1S 5B6, Canada

<sup>2</sup>Centre de Recherches Mathématiques, Université de Montréal, Montréal, H3T 1J4, Canada

\*subhendu.paul@carleton.ca

## ABSTRACT

We derive a novel model escorted by large scale compartments, based on a set of coupled delay differential equations with extensive delays, in order to estimate the incubation, recovery and decease periods of COVID-19, and more generally any infectious disease. This is possible thanks to machine learning algorithms applied to publicly available database of confirmed corona cases, recovered cases and death toll. In this purpose, we separate i) the total cases into 14 groups corresponding to 14 incubation periods, ii) the recovered cases into 406 groups corresponding to a combination of incubation and recovery periods, and iii) the death toll into 406 groups corresponding to a combination of incubation and decease periods. In this paper, we focus on recovery and decease periods and their correlation with the incubation period. The estimated mean recovery period we obtain is 22.14 days (95% Confidence Interval(CI): 22.00 to 22.27), and the 90th percentile is 28.91 days (95% CI: 28.71 to 29.13), which is in agreement with statistical supported studies. The bimodal gamma distribution reveals that there are two groups of recovered individuals with a short recovery period, mean 21.02 days (95% CI: 20.92 to 21.12), and a long recovery period, mean 38.88 days (95% CI 38.61 to 39.15). Our study shows that the characteristic of the decease period and the recovery period are alike. From the bivariate analysis, we observe a high probability domain for recovered individuals with respect to incubation and recovery periods. A similar domain is obtained for deaths analyzing bivariate distribution of incubation and decease periods.

## Main

The outbreak of coronavirus disease 2019 (COVID-19), reported early in Wuhan (China)<sup>1</sup> and spread around the world, is creating dramatic and daily changes with profound impacts worldwide. As a consequence the outbreak was declared a pandemic by the World Health Organization (WHO) in March 2020<sup>2</sup>, and by the end of 2020, COVID-19 has infected about 79.2 millions of people in the world, with an approximate cumulative global mortality of 3.2%<sup>2</sup>. To limit the impact of this deadly virus, a rapid and widespread vaccination of the population is now in place. However, it is established that vaccine are not 100% effective to stop the transmission or infection of COVID-19. In addition, huge numbers of global SARS-CoV-2 infections have led to the emergence of variants, notably Alpha (B.1.1.7 UK), Beta (B.1.351 S. Africa), Gamma (P.1 Brazil), Epsilon (B.1.429 California), Iota (B.1.526 New York), Delta and Kappa (B.1.617.2 and B.1.617.1 India) which make the situation more challenging. In this circumstance to get a complete feature of COVID-19, it is essential to fully understand the key (incubation, recovery and decease) periods.

We already successfully estimated the incubation period of COVID-19 in Canada<sup>3</sup>. In the present context, we focus on the recovery and decease periods and their correlation with the incubation period. In the current framework, we define the recovery period as the time from the contraction of the coronavirus to recovery, i.e., the incubation period plus the onset time from the symptom to recovery; the latter is the same as the viral shedding of SARS-CoV-2. We describe the decease period in the same way as the recovery period. Understanding the recovery period of disease is very useful information in the struggle against the disease. If the incidence of a disease is remarkably high and the recovery period of the disease is also high then the prevalence of the disease in the country is likely to increase which in turn puts extra health, economic and social burden on this country. Understanding the recovery period of the disease will help governments to plan proper strategies to counter the disease and to organize the requirements such as hospitals, doctors, medical staffs, medical equipment's, etc. It will also help to implement different social and economic policies which will be essential to fight the disease.

There are several statistical studies<sup>4-11</sup>, based on various samples of patients such as severe, non-severe, ICU, non-ICU, large size, small size, meta-analysis, estimated the recovery time of the current pandemic. In addition to those statistical approaches, there are numerous analytical and computational studies based on mathematical models, involving Ordinary

Differential Equations (ODE)<sup>12–20</sup> as well as Delay Differential Equations (DDE)<sup>21–26</sup>, to calculate the basic reproduction number  $R_0$  and understand the underlying dynamics of the epidemic. Researchers usually consider single-delay models, occasionally two delays.

To the best of our knowledge, we demonstrate for the first time a substantial compartment based model, with a total 830 partitions, in order to estimate the key (incubation, recovery, decease periods) periods of COVID-19 as well as the bivariate distribution of incubation and recovery periods, and the bivariate distribution of incubation and decease periods. This will be achieved using publicly available database<sup>27</sup> of the total number of corona-positive cases, recovery and death toll. Using the novel model, demonstrated here, we divide the publicly available database into thousands of groups, and these separated classes are the key source for estimating all the key periods. This approach is free from any special type of samples in order to produce the distributions of those periods; it only involves large scale computations for estimating about thousand model parameters. After a single calculation of this method, we can generate the current distributions as well as previous distributions of those periods. In the statistical based approaches, it is usually difficult to consider large incubation, recovery and decease periods if the sample size is small. However, in our approach, we can go well beyond 14 days, the maximum incubation period that we have set in this paper, and beyond the interval 2 weeks to 6 weeks, the range of recovery as well as decease periods that we have considered in the current computations. As of May 23, 2021, the World Health Organization (WHO) had confirmed a total of 1,359,180 cases of COVID-19 in Canada, including 25,231 deaths<sup>2</sup>. As of May 23, 2021 there are five provinces in Canada with death toll more than 1000 (Fig. 1(a)), and the recovery and death rates are respectively 96.1% and 0.7% (Fig. 1(c)). During the first wave of COVID-19 in Canada, January 22, 2020 to July 16, 2020, the recovery and death rates were respectively 66.5% and 8.1% (Fig. 1(b)). Here, we assume that the recovery and decease periods of COVID-19 remain unchanged i.e., these periods during the first wave and the present time are almost identical, and under this assumption we merely consider the database of first wave for the calculation.

There are various studies on recovery period, and no result is reported (to the best of our knowledge) on bivariate distributions as mentioned above. The key periods may depend on age<sup>28</sup> (median-age / country), hard immunity, public health system, corona testing capacities, daily corona cases, etc. For a better estimation of the key periods for a particular region, we need to study local patients. Data collection is a bottleneck in studying those key periods for COVID-19 or other infectious diseases using clinical survey, and we need a sample of large size for bivariate analysis. However, key periods can easily be estimated using the approach we propose here, the publicly available database along with machine learning algorithms.

## Results

The proposed model assists us to generate new refined recovery and death toll database,  $R_{ik}$  and  $D_{ik}$ , by dividing the total recovered individuals and the total number of deaths as of July 16, 2020 into myriad of groups. The new database is the key source for studying all kinds of distributions, reported in the article.

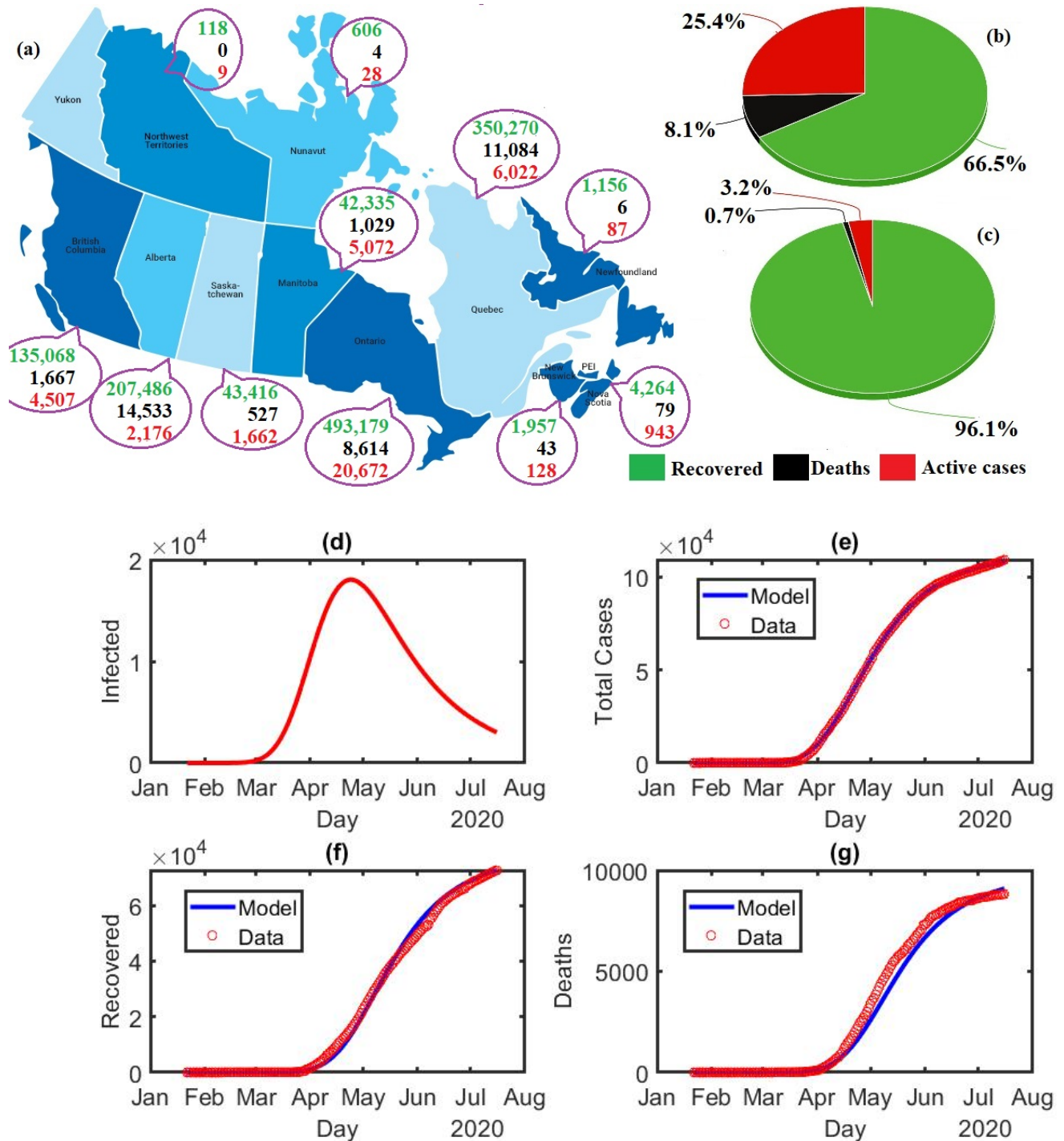
### Validation of the proposed model

After estimating the model parameters with sufficiently small values of error functions, we obtain a good agreement (Figs. 1(e), 1(f) and 1(g)) between the calculated values of the model variables such as total corona-positive cases, number of recovered individuals, etc. and the available data<sup>27</sup>. The population of the infected group gradually increased until end of April 2020, and thereafter slowed down (Fig. 1(d)).

### Univariate distributions

The groups of recovered individuals  $R_{ik}$ ,  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , corresponding to the incubation period (in days)  $\tau_i$ ,  $1 \leq \tau_i \leq 14$ , and recovery period (in days)  $\zeta_k$ ,  $14 \leq \zeta_k \leq 42$  can be represented in a matrix form (Fig. 2(a)). We use the data set  $\sum_{i=1}^{14} R_{ik}$  for  $\zeta_k = 14, 15, \dots, 42$  to obtain the frequency distribution for recovery period and the corresponding fitted gamma distributions, unimodal (Fig. 2(b))  $\Gamma(\zeta, K_r, \theta_r)$  and bimodal (Fig. 2(c))  $0.9365\Gamma(\zeta, K_{r1}, \theta_{r1}) + 0.0635\Gamma(\zeta, K_{r2}, \theta_{r2})$ . Here, the variable  $\zeta$  indicates the recovery period and the parameters  $K_r = 18.62067$ ,  $\theta_r = 1.18892$ ,  $K_{r1} = 34.55447$ ,  $\theta_{r1} = 0.60847$ ,  $K_{r2} = 226.40545$  and  $\theta_{r2} = 0.17171$  with statistical  $p$  value less than 0.01. The mean recovery period we obtain using an unimodal gamma distribution is 22.14 days (95% CI 22.00 to 22.27); the median of the recovery period is 21.74 days (95% CI 21.61 to 21.87); the 90th percentile is 28.91 days (95% CI 28.71 to 29.13); the 95th percentile is 31.20 days (95% CI 30.95 to 31.45). For a better estimation, we use a bimodal distribution, a linear combination of  $\Gamma(\zeta, K_{r1}, \theta_{r1})$  and  $\Gamma(\zeta, K_{r2}, \theta_{r2})$ . The mean of  $\Gamma(\zeta, K_{r1}, \theta_{r1})$  and  $\Gamma(\zeta, K_{r2}, \theta_{r2})$  are 21.02 days (95% CI 20.92 to 21.12) and 38.88 days (95% CI 38.61 to 39.15), respectively. The percentile curves of unimodal and bimodal gamma distributions show (Fig. 2(d)) that the median of unimodal and bimodal are the same, although there are slight differences other than the median.

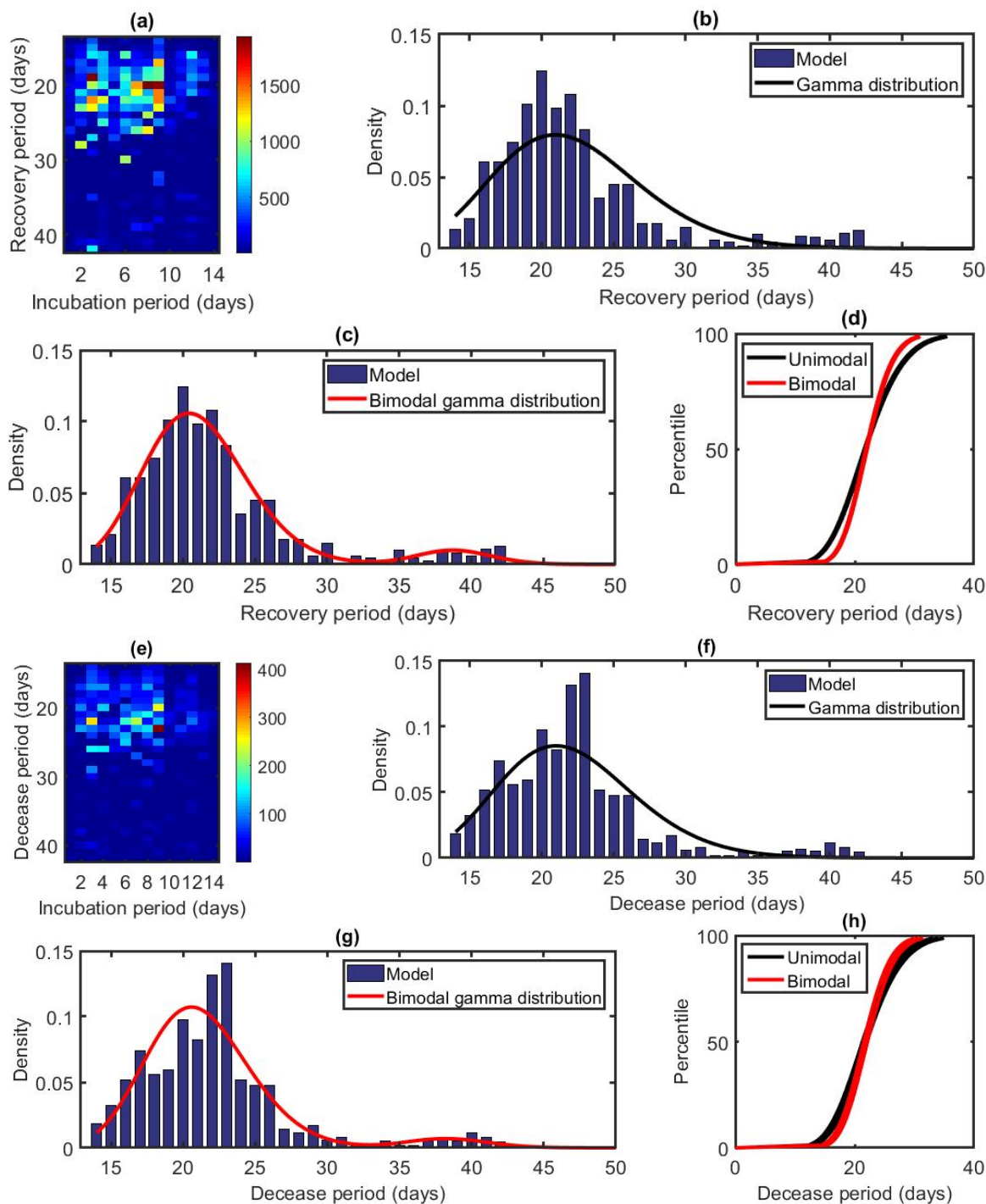
The groups death toll groups  $D_{ik}$ ,  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , and corresponding incubation period (in days)  $\tau_i$ ,  $1 \leq \tau_i \leq 14$ , and decease period (in days)  $\eta_k$ ,  $14 \leq \eta_k \leq 42$  can be represented as a matrix (Fig. 2(e)). We use the data set  $\sum_{i=1}^{14} D_{ik}$  for  $\eta_k = 14, 15, \dots, 42$  to obtain the frequency distribution for decease period and corresponding fitted gamma



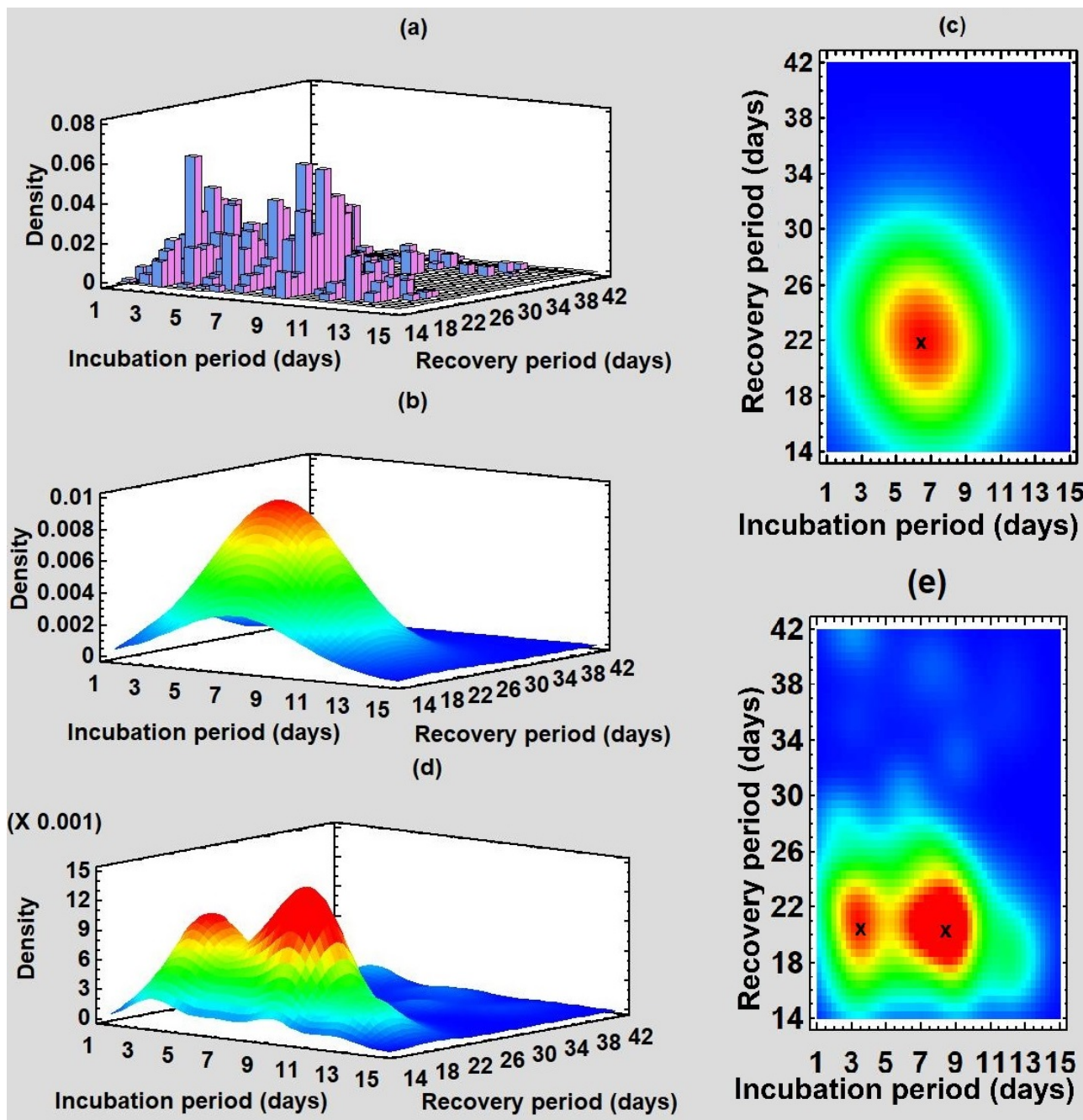
**Figure 1. COVID-19 pandemic in Canada:** (a) As of May 23, 2021 Canadian out break at-a-glance; the green, black and red digits are represented the number of recovered individuals, death toll and active cases, respectively. (b) Percentage of recovered (green), deaths (black) and active cases (red) in Canada during the first wave, January 22, 2020 to July 16, 2020. (c) Percentage of recovered (green), deaths (black) and active cases (red) in Canada as of May 23, 2021. **Model calculation for Canada during the first wave, January 22, 2020 to July 16, 2020:** (d) Estimation of the number of infected individuals. (e) Estimation of the total number of coronavirus cases compared to the available data<sup>27</sup>. (f) Estimation of the total number of recovered compared to the available data<sup>27</sup>. (g) Estimation of the total number of deaths compared to the available data<sup>27</sup>.

**Table 1.** Comparison of several studies (including the present work) for infectious period along with sample size, mean/median and ranges. Here SD and IQR stand for standard deviation and interquartile range, respectively.

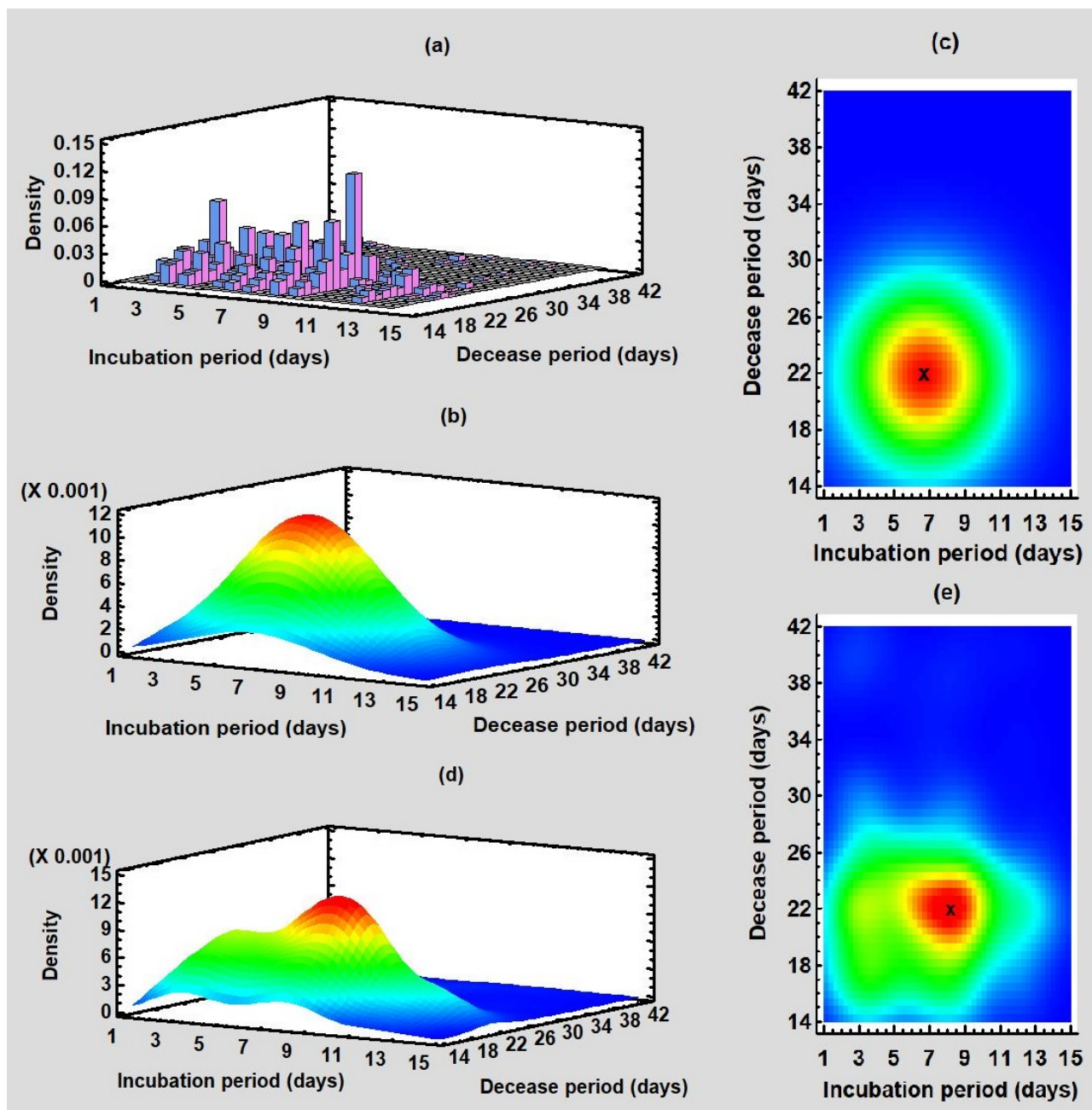
Author	Location	Data size	Parameter (days)	Reported	Variation (days)	Comment	
Present work	Canada	72,680	15.40	Mean	95% CI 14.87 to 15.92	Onset time from symptom to recovery (OTSR)	
			14.28			95% CI 13.79 to 14.77	Short OTSR
			32.14			95% CI 31.48 to 32.80	Long OTSR
Voinsky <i>et. al.</i> <sup>4</sup>	Israeli	5,769		Mean	13.24 to 14.81	Days from first positive to first negative COVID test	
Barman <i>et. al.</i> <sup>5</sup>	India	221	21	Mean	95% CI 12.82 to 29.32	Days of hospitalization Age < 60 yrs.	
			25			95% CI 17.22 to 32.78	Days of hospitalization Age > 60 yrs.
Cai <i>et. al.</i> <sup>6</sup>	China	298	14	Median	9 to 19 (IQR)	Days of treatment within hospital setting	
Fang <i>et. al.</i> <sup>7</sup>	China	24	15.7	Mean	6.7(SD)	Days of hospitalization, non-ICU	
Wu <i>et. al.</i> <sup>8</sup>	China	74	16.1	Mean	6.7(SD)	Days of hospitalization, severe and non-severe	
Bi <i>et. al.</i> <sup>9</sup>	China	391	21		95% CI 20 to 22	Median time to recovery	
Alinaghi <i>et. al.</i> <sup>10</sup>	Iran	478	13.5		IQR: 9	Median time to recovery	



**Figure 2. Distribution of the recovery and decease periods:** Results based on the total recovered cases of the first 177 days during the pandemic in Canada starting from January 22, 2020 i.e., cumulative data as of July 16, 2020. **(a)** Splitting values of recovered individuals as a function of incubation and recovery periods. **(b)** Probability density function of the gamma distribution  $\Gamma(\zeta, K, \theta)$  with  $K = 18.62067$  and  $\theta = 1.18892$ . The blue bars indicate the densities obtained from the model calculation. **(c)** Probability density function of the bimodal gamma distribution  $0.9365\Gamma(\zeta, K_1, \theta_1) + 0.0635\Gamma(\zeta, K_2, \theta_2)$  with  $K_1 = 34.55447$ ,  $\theta_1 = 0.60847$ ,  $K_2 = 226.40545$  and  $\theta_2 = 0.17171$ . The blue bars indicate the densities obtained from the model calculation. **(d)** Percentile curves for unimodal and bimodal gamma distributions. **(e)** Splitting values of the deaths as a function of incubation and decease periods. **(f)** Probability density function of the gamma distribution  $\Gamma(\eta, K, \theta)$  with  $K = 21.33660$  and  $\theta = 1.03174$ . The blue bars indicate the densities obtained from the model calculation. **(g)** Probability density function of the bimodal gamma distribution  $0.9508\Gamma(\eta, K_1, \theta_1) + 0.0492\Gamma(\eta, K_2, \theta_2)$  with  $K_1 = 35.00855$ ,  $\theta_1 = 0.60511$ ,  $K_2 = 186.11379$  and  $\theta_2 = 0.20636$ . The blue bars indicate the densities obtained from the model calculation. **(h)** Percentile curves for unimodal and bimodal gamma distributions.



**Figure 3. Bivariate distribution of the incubation and recovery periods:** (a) Histogram of the estimated data  $R_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  using the model. (b) Fitted bivariate normal distribution. (c) Two-dimensional display of (b); the red region is the highly probable domain for recovery, and  $x$  (6.43, 21.91) denotes the center of the region. (d) Fitted nonparametric density estimate with wide 33%; two peaks show that there are two distinguishable high probable regions. (e) Two-dimensional display of (d); two  $x$  represent the centers of two high probable regions (3.49, 20.52) and (8.38, 20.35).



**Figure 4. Bivariate distribution of the incubation and decease periods:** (a) Histogram of the estimated data  $D_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  using the model. (b) Fitted bivariate normal distribution. (c) Two-dimensional display of (b); the red region is the highly probable domain for decease, and  $x$  (6.56, 21.64) denotes the center of the region. (d) Fitted nonparametric density estimate with wide 40%; two peaks show that there are two distinguishable high probable regions. (e) Two-dimensional display of (d); the  $x$  represent the centers of the high probable red region (8.17, 21.86).

distributions, unimodal (Fig.2(f))  $\Gamma(\eta, K_d, \theta_d)$  and bimodal (Fig.2(g))  $0.9508\Gamma(\eta, K_{d1}, \theta_{d1}) + 0.0492\Gamma(\eta, K_{d2}, \theta_{d2})$ . Here, the variable  $\eta$  indicates the decrease period and the parameters  $K_d = 21.33660$ ,  $\theta_d = 1.03174$ ,  $K_{d1} = 35.00855$ ,  $\theta_{d1} = 0.60511$ ,  $K_{d2} = 186.11379$  and  $\theta_{d2} = 0.20636$  with statistical  $p$  value less than 0.01 for  $\Gamma(\eta, K_d, \theta_d)$ ,  $\Gamma(\eta, K_{d1}, \theta_{d1})$  and equal to 0.18 for  $\Gamma(\eta, K_{d2}, \theta_{d2})$ . The mean decrease period we obtain using an unimodal gamma distribution is 22.01 days (95% CI 21.64 to 22.39); the median of the decrease period is 21.67 days (95% CI 21.31 to 22.04); the 90th percentile is 28.30 days (95% CI 27.72 to 28.89); the 95th percentile is 30.39 days (95% CI 29.71 to 31.10). For better estimation, we use a bimodal distribution, a linear combination of  $\Gamma(\eta, K_{d1}, \theta_{d1})$  and  $\Gamma(\eta, K_{d2}, \theta_{d2})$ . The mean of  $\Gamma(\eta, K_{d1}, \theta_{d1})$  and  $\Gamma(\eta, K_{d2}, \theta_{d2})$  are 21.18 days (95% CI 20.90 to 21.47) and 38.41 days (95% CI 37.41 to 39.40), respectively. The percentile curves show (Fig.2(h)) that the percentiles of unimodal and bimodal distributions are almost the same.

## Bivariate distributions

To analyze the bivariate distribution, we use the software Statgraphics<sup>29</sup>, based on the statistical package **R**. Using the elements  $R_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , we obtain a bivariate histogram (Fig.3(a)) for the incubation and recovery periods. There are two peaks at the points (3, 19), i.e., for  $\tau_i = 3$  and  $\zeta_k = 19$ , and (8, 20), i.e., for  $\tau_i = 8$  and  $\zeta_k = 20$ , corresponding to the high densities of recovered individuals. We estimate the histogram using a bivariate normal distribution  $\mathcal{N}(m_\tau^{(r)}, m_\zeta^{(r)}, \sigma_\tau^{(r)}, \sigma_\zeta^{(r)}, \rho_{\tau\zeta})$  (Fig.3(b)) where the variables  $\tau$  and  $\zeta$  represent the incubation and recovery periods, respectively. The mean  $m_\tau^{(r)}$  and standard deviation  $\sigma_\tau^{(r)}$  of the incubation period are 6.43 (95% CI 6.27 to 6.59) and 3.06 (95% CI 2.96 to 3.18), respectively; the mean  $m_\zeta$  and standard deviation  $\sigma_\zeta$  of the recovery period are 21.91 (95% CI 21.63 to 22.18) and 5.33 (95% CI 5.14 to 5.53), respectively; the correlation between incubation and recovery periods  $\rho_{\tau\zeta}$  is -0.11. The two dimensional representation of the bivariate normal distribution (Fig.3(c)) shows that the highly probable recovery region (red in the figure) is a nested domain of  $\tau = 6.43$  and  $\zeta = 21.91$ . To precisely analyze the highly probable region, we estimate the histogram (Fig.3(a)) using a nonparametric density function with a width of 33%, low and high percentage give a more local and global estimation, respectively, and we obtain a distribution with two peaks (Fig.3(d)). Two distinguishable peaks indicate that there are two separate highly probable regions surrounding the points  $\tau = 3.49$ ,  $\zeta = 20.52$  and  $\tau = 8.38$ ,  $\zeta = 20.35$  (Fig.3(e)). The bivariate mixture distribution analysis shows that we can estimate the histogram of  $R_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  using a combination of two bivariate normal distributions,  $0.94\mathcal{N}(m_\tau^{(r1)}, m_\zeta^{(r1)}, \sigma_\tau^{(r1)}, \sigma_\zeta^{(r1)}, \rho_{\tau\zeta}^{(r1)}) + 0.06\mathcal{N}(m_\tau^{(r2)}, m_\zeta^{(r2)}, \sigma_\tau^{(r2)}, \sigma_\zeta^{(r2)}, \rho_{\tau\zeta}^{(r2)})$  where the superscript 1 (resp. 2) represents the parameters for the first (resp. second) component, respectively. The parameters of the first component are  $m_\tau^{(r1)} = 6.43$ ,  $m_\zeta^{(r1)} = 20.83$ ,  $\sigma_\tau^{(r1)} = 3.05$ ,  $\sigma_\zeta^{(r1)} = 3.36$ ,  $\rho_{\tau\zeta}^{(r1)} = -0.16$ , and those of the second component are  $m_\tau^{(r2)} = 6.42$ ,  $m_\zeta^{(r2)} = 37.83$ ,  $\sigma_\tau^{(r2)} = 3.25$ ,  $\sigma_\zeta^{(r2)} = 3.45$ ,  $\rho_{\tau\zeta}^{(r2)} = -0.47$ .

Using the elements  $D_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , we obtain a bivariate histogram (Fig.4(a)) for the incubation and decrease periods. There are two peaks at the points (3, 22), i.e., for  $\tau_i = 3$  and  $\eta_k = 22$ , and (9, 23), i.e., for  $\tau_i = 9$  and  $\eta_k = 23$ , corresponding to the high densities of deaths. We estimate the histogram using a bivariate normal distribution  $\mathcal{N}(m_\tau^{(d)}, m_\eta, \sigma_\tau^{(d)}, \sigma_\eta, \rho_{\tau\eta})$  (Fig.4(b)) where the variables  $\tau$  and  $\eta$  represent the incubation and decrease periods, respectively. The mean  $m_\tau^{(d)}$  and standard deviation  $\sigma_\tau^{(d)}$  of the incubation period are 6.56 (95% CI 6.36 to 6.76) and 3.00 (95% CI 2.86 to 3.15), respectively; the mean  $m_\eta$  and standard deviation  $\sigma_\eta$  of the decrease period are 21.64 (95% CI 21.33 to 21.94) and 4.43 (95% CI 4.23 to 4.65), respectively; the correlation between incubation and decrease periods  $\rho_{\tau\eta}$  is -0.008. The two dimensional representation of the bivariate normal distribution (Fig.4(c)) shows that the highly probable decrease region (red in the figure) is a nested domain of  $\tau = 6.56$  and  $\eta = 21.64$ . To precisely analyze the highly probable regions, we estimate the histogram (Fig.4(a)) using a nonparametric density function with a width of 40%, low and high percentage give a more local and global estimation, respectively, and obtain a distribution with two peaks (Fig.4(d)), one in the high probability region (red in figure) and another one in the second high probability region (yellow in figure). The highly probable region is surrounding the point  $\tau = 8.17$ ,  $\eta = 21.86$  (Fig.4(e)). The bivariate mixture distribution analysis shows that we can estimate the histogram of  $D_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  using a combination of two bivariate normal distributions,  $0.97\mathcal{N}(m_\tau^{(d1)}, m_\eta^{(1)}, \sigma_\tau^{(d1)}, \sigma_\eta^{(1)}, \rho_{\tau\eta}^{(1)}) + 0.03\mathcal{N}(m_\tau^{(d2)}, m_\eta^{(2)}, \sigma_\tau^{(d2)}, \sigma_\eta^{(2)}, \rho_{\tau\eta}^{(2)})$  where the superscript 1 (resp. 2) represents the parameters for first (resp. second) component. The parameters of the first component are  $m_\tau^{(d1)} = 6.58$ ,  $m_\eta^{(1)} = 21.19$ ,  $\sigma_\tau^{(d1)} = 2.98$ ,  $\sigma_\eta^{(1)} = 3.49$ ,  $\rho_{\tau\eta}^{(1)} = 0.03$ , and those of the second component are  $m_\tau^{(d2)} = 5.81$ ,  $m_\eta^{(2)} = 38.61$ ,  $\sigma_\tau^{(d2)} = 3.33$ ,  $\sigma_\eta^{(2)} = 2.15$ ,  $\rho_{\tau\eta}^{(2)} = -0.29$ .

## Onset time from symptom to recovery

Using the fact that  $\tau + \theta = \zeta$  and the property of expectation  $E(T + \Theta) = E(T) + E(\Theta)$ , we calculate the mean Onset Time from Symptom to Recovery (OTSR)  $E(\Theta)$  (Table1), where  $\theta$  is the variable corresponding to  $\theta_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ ;  $T$  and  $\Theta$  are the random variables corresponding the incubation period and OTSR, respectively. There is a good agreement between the calculated values, mean of OTSR, short OTSR and long OTSR, with the reported works (Table1)



of earlier studies. However, these calculated values do not show excellent concordance with some other studies, because we consider all recovery cases, mild to moderate, severe, hospitalized (ICU, non-ICU), non hospitalized, in Canada. For example, Voinsky *et. al.*<sup>4</sup> reported a study with a sample of 5769 patients, not including severe COVID-19 cases. In fact, they mentioned that severe cases were reported to be discharged from the hospital on average 8 days longer than mild to moderate patients requiring hospitalization.

## Discussion

In the present context, we estimate the recovery as well as decease periods using a novel compartment based model and publicly available database. Here, we consider a maximum length of the incubation period of 14 days, and the ranges of the recovery and decease periods are from 2 to 6 weeks. However, in our method, we can go well beyond all those ranges; the longer ranges simply require a long computational time. Notice that our method could apply the proposed model to estimate key periods for any infectious disease, as along as similar data are available.

The multi-group database  $R_{ik}$ ,  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , generated from the model, is the key source to compute all types of distribution of the recovery period, univariate, bimodal and bivariate. The bimodal gamma distribution of the recovery period,  $0.9365\Gamma(\zeta, K_{r1}, \theta_{r1}) + 0.0635\Gamma(\zeta, K_{r2}, \theta_{r2})$ , demonstrates that the recovery period of 93.65% recovered individuals obeys the distribution  $\Gamma(\zeta, K_{r1}, \theta_{r1})$ , and that of 6.35% recovered individuals obeys the distribution  $\Gamma(\zeta, K_{r2}, \theta_{r2})$ . Thus, there are two groups of recovered individuals with short recovery period, 21.02 days (on average), and long recovery period, 38.88 days (on average). The characteristics of those two groups may depend on age, underlying health condition, immunity, etc. The database of numerous groups  $D_{ik}$ ,  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$ , generated from the model, is the key source to compute all types of distribution of the decease period, univariate, bimodal and bivariate. The bimodal gamma distribution of the decease period,  $0.9508\Gamma(\zeta, K_{d1}, \theta_{d1}) + 0.0492\Gamma(\zeta, K_{d2}, \theta_{d2})$ , demonstrates that the decease period of 95.08% deaths obeys the distribution  $\Gamma(\zeta, K_{d1}, \theta_{d1})$ , and that of 4.92% deaths obeys the distribution  $\Gamma(\zeta, K_{d2}, \theta_{d2})$ . Thus, there are two groups of deaths with short decease period, 21.18 days (on average), and long decease period, 38.41 days (on average). The characteristics of those two groups may depend on age, underlying health condition, immunity, etc. The calculated results employing the proposed model show that the recovery and decease periods are the same. It seems that the survival period of the coronavirus is the same as that of human, in the form of immunity.

The bivariate normal distribution of incubation and recovery periods indicates a recovery window of  $4.82 \leq \tau \leq 8.49$  and  $19.27 \leq \zeta \leq 25.72$  as the highly probable domain for recovery. The bivariate normal distribution of incubation and decease periods indicates a decease window of  $4.55 \leq \tau \leq 8.45$  and  $19.35 \leq \eta \leq 24.85$  as the highly probable domain of deaths. The study shows that the recovery and decease windows almost coincide within these key periods. To determine precisely the recovery as well as the decease windows, we use nonparametric distributions. Under the nonparametric analysis we identify two recovery windows,  $2.27 \leq \tau \leq 4.38$ ,  $18.41 \leq \zeta \leq 22.79$  and  $6.42 \leq \tau \leq 9.63$ ,  $17.81 \leq \zeta \leq 23.93$ , and one decease window,  $6.34 \leq \tau \leq 9.41$ ,  $20.17 \leq \eta \leq 23.69$ . Nonparametric analysis provides some discrepancy between the recovery and decease windows.

The bivariate mixed distribution,  $0.94\mathcal{N}(m_{\tau}^{(r1)}, m_{\zeta}^{(1)}, \sigma_{\tau}^{(r1)}, \sigma_{\zeta}^{(1)}, \rho_{\tau\zeta}^{(1)}) + 0.06\mathcal{N}(m_{\tau}^{(r2)}, m_{\zeta}^{(2)}, \sigma_{\tau}^{(r2)}, \sigma_{\zeta}^{(2)}, \rho_{\tau\zeta}^{(2)})$ , of the incubation and recovery periods demonstrates that 94% recovered individuals obey  $\mathcal{N}(m_{\tau}^{(r1)}, m_{\zeta}^{(1)}, \sigma_{\tau}^{(r1)}, \sigma_{\zeta}^{(1)}, \rho_{\tau\zeta}^{(1)})$ , the bivariate normal distribution, with recovery window  $m_{\tau}^{(r1)} - \sigma_{\tau}^{(r1)} \leq \tau \leq m_{\tau}^{(r1)} + \sigma_{\tau}^{(r1)}$ ,  $m_{\zeta}^{(1)} - \sigma_{\zeta}^{(1)} \leq \zeta \leq m_{\zeta}^{(1)} + \sigma_{\zeta}^{(1)}$  and 6% recovered individuals obey the bivariate normal distribution  $\mathcal{N}(m_{\tau}^{(r2)}, m_{\zeta}^{(2)}, \sigma_{\tau}^{(r2)}, \sigma_{\zeta}^{(2)}, \rho_{\tau\zeta}^{(2)})$  with recovery window  $m_{\tau}^{(r2)} - \sigma_{\tau}^{(r2)} \leq \tau \leq m_{\tau}^{(r2)} + \sigma_{\tau}^{(r2)}$ ,  $m_{\zeta}^{(2)} - \sigma_{\zeta}^{(2)} \leq \zeta \leq m_{\zeta}^{(2)} + \sigma_{\zeta}^{(2)}$ . The bivariate mixed distribution,  $0.97\mathcal{N}(m_{\tau}^{(d1)}, m_{\eta}^{(1)}, \sigma_{\tau}^{(d1)}, \sigma_{\eta}^{(1)}, \rho_{\tau\eta}^{(1)}) + 0.03\mathcal{N}(m_{\tau}^{(d2)}, m_{\eta}^{(2)}, \sigma_{\tau}^{(d2)}, \sigma_{\eta}^{(2)}, \rho_{\tau\eta}^{(2)})$ , of the incubation and decease periods demonstrates that 97% deaths obey the bivariate normal distribution  $\mathcal{N}(m_{\tau}^{(d1)}, m_{\eta}^{(1)}, \sigma_{\tau}^{(d1)}, \sigma_{\eta}^{(1)}, \rho_{\tau\eta}^{(1)})$  with decease window  $m_{\tau}^{(d1)} - \sigma_{\tau}^{(d1)} \leq \tau \leq m_{\tau}^{(d1)} + \sigma_{\tau}^{(d1)}$ ,  $m_{\eta}^{(1)} - \sigma_{\eta}^{(1)} \leq \eta \leq m_{\eta}^{(1)} + \sigma_{\eta}^{(1)}$  and 3% deaths obey the bivariate normal distribution  $\mathcal{N}(m_{\tau}^{(d2)}, m_{\eta}^{(2)}, \sigma_{\tau}^{(d2)}, \sigma_{\eta}^{(2)}, \rho_{\tau\eta}^{(2)})$  with decease window  $m_{\tau}^{(d2)} - \sigma_{\tau}^{(d2)} \leq \tau \leq m_{\tau}^{(d2)} + \sigma_{\tau}^{(d2)}$ ,  $m_{\eta}^{(2)} - \sigma_{\eta}^{(2)} \leq \eta \leq m_{\eta}^{(2)} + \sigma_{\eta}^{(2)}$ .

In summary, we have developed a novel compartment based model to divide the publicly available database of total confirmed cases, recovered cases, and number of deaths into numerous subgroups to obtain univariate and bivariate distributions. The model itself and the procedure to solve it, are the core of this work, and it can be applied to any infectious disease in any region provided that similar data are available. In conclusion, we obtain the distributions of the key periods from the population, considering all types of cases (non-hospitalized, non-ICU, ICU) of recovered individuals and deaths, which is naturally better than any sample-dependent result. In this approach, we do not need any clinical survey; the publicly available data on confirmed cases, recovery and death toll, are sufficient to analyze univariate and bivariate distributions. The current model can be extended to study age-based key periods, and for this purpose we need an age dependent database. The monotonic iteration scheme, introduced for better estimation, can be applied to machine learning as well as numerical analysis problems.

## Methods

In this section, we introduce a compartment based infectious disease model including a large number of partitions, Lockdown, Susceptible, Removed, Infected, fourteen compartments of Confirmed cases, hundreds compartments of Recovered and Deaths. The model is constructed as a set of coupled delay differential equations involving few thousands of variables and parameters, and will be used, not as a prediction tool, but i) for constructing the myriad groups of recovered individuals and death tools and ii) estimating accurately the recovery and decease periods. This model will however have to be parameterized and validated using existing data, in order to justify its accuracy and its application in the proposed methodology.

### The Model

Modeling the spread of epidemic is an essential tool for projecting its outcome. By estimating important epidemiological parameters using the available database and machine learning techniques, we can make predictions of different intervention scenarios. Compartment based model, where the population of a region is distributed into several population groups, such as susceptible, infected, total cases, etc., is a simple but useful tool to demonstrate the panorama of an epidemic.

The proposed model is an extension of our previous work<sup>3</sup>, including a very large number of compartments of recovered and deaths individuals; the schematic diagram of the model is presented in Fig. 5(a). The following are the underlying principles of the present model.

- The total population is constant (neglecting the migrations, births and unrelated deaths) and initially every individual is assumed susceptible to contract the disease.
- The disease is spread through the direct (face-to-face meeting) or indirect (through air current, common used or delivery items like door handles, grocery products) contact of susceptible individuals with the infected individuals.
- The quarantined area or the compartment for corona cases contains only members of the infected population who are tested corona-positive.
- The virus kills a part of the people it infects; the survivors represent the recovered group.
- There is a non-pharmaceutical policy (stay at home), commonly known as *lockdown*, to stop the spread of the disease.
- The group of asymptomatic patients is a part of infected individuals, and the never-tested recovered asymptomatic patients can be removed from the infected group. If an asymptomatic patient dies, it is counted after investigation.

Based on the above principles, we consider the following compartments:

- Lockdown (insusceptible) ( $L$ ): the group of persons who are keeping themselves safe.
- Susceptible ( $S$ ): the group of individuals who can be infected.
- Infected ( $I$ ): the group of people who are spreading the contiguous disease.
- Removed ( $V$ ): the group of recovered asymptomatic patients without testing.
- Confirmed cases ( $C$ ): the group of individuals who tested corona-positive.
- Recovered ( $R$ ): the group of recovered individuals who tested corona-positive.
- Deaths ( $D$ ): the group of deaths individuals who tested corona-positive.

In the present context, we assume that there is no overlap between these two compartments, infected ( $I$ ) and confirmed cases ( $C$ ). In other words, tested corona-positive individuals are assumed to be unable to substantially spread the disease due to isolation and are immune to re-infection after recovery<sup>30</sup>. The aim of the present work is to estimate the distribution of the recovery and decease periods of COVID-19. In this goal, we split the compartment  $C$  into  $J$  subcomponents  $C_1, \dots, C_J$ , the compartment  $R$  into  $J \times M$  subcomponents  $R_{ik}$  for  $i = 1, \dots, J$  and  $k = 1, \dots, M$  and the compartment  $D$  into  $J \times M$  subcomponents  $D_{ik}$  for  $i = 1, \dots, J$  and  $k = 1, \dots, M$  where

$$C(t) = \sum_{i=1}^J C_i(t) \text{ or } C^{(m)} = \sum_{i=1}^J C_i^{(m)}, \quad (1)$$

$$R(t) = \sum_{i=1}^J \sum_{k=1}^M R_{ik}(t) \text{ or } R^{(m)} = \sum_{i=1}^J \sum_{k=1}^M R_{ik}^{(m)}, \quad (2)$$

$$D(t) = \sum_{i=1}^J \sum_{k=1}^M D_{ik}(t) \text{ or } D^{(m)} = \sum_{i=1}^J \sum_{k=1}^M D_{ik}^{(m)}. \quad (3)$$

In (1), (2) and (3)  $m$  represents the time index, and  $C_i^{(m)}$ ,  $R_{ik}^{(m)}$  and  $D_{ik}^{(m)}$  represents the total corona-positive cases corresponding the incubation period  $\tau_i$ , recovered individuals corresponding the incubation period  $\tau_i$  and onset time  $\theta_{ik}$  i.e., recovery period  $\zeta_k = \tau_i + \theta_{ik}$  and death toll corresponding the incubation period  $\tau_i$  and onset time  $\mu_{ik}$  i.e., decease period  $\eta_k = \tau_i + \mu_{ik}$ , respectively, presented in Fig.5(a).

The time-dependent model is the following set of coupled delay differential equations, for  $i = 1, \dots, J$ :

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -\beta \frac{SI}{N} - \alpha S + \nu L \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \beta \sum_{i=1}^J \delta_i \frac{S(t-\tau_i)I(t-\tau_i)}{N} \\ \frac{dV}{dt} = \gamma I \\ \frac{dC_i}{dt} = \delta_i \beta \frac{S(t-\tau_i)I(t-\tau_i)}{N} - \beta \sum_{k=1}^M \lambda_{ik} \delta_i \frac{S(t-\tau_i-\theta_{ik})I(t-\tau_i-\theta_{ik})}{N} \\ \quad - \beta \sum_{k=1}^M \kappa_{ik} \delta_i \frac{S(t-\tau_i-\mu_{ik})I(t-\tau_i-\mu_{ik})}{N} \\ \frac{dR_{ik}}{dt} = \lambda_{ik} \delta_i \beta \frac{S(t-\tau_i-\theta_{ik})I(t-\tau_i-\theta_{ik})}{N} \\ \frac{dD_{ik}}{dt} = \kappa_{ik} \delta_i \beta \frac{S(t-\tau_i-\mu_{ik})I(t-\tau_i-\mu_{ik})}{N} \\ \frac{dL}{dt} = \alpha S - \nu L \end{array} \right. \quad (4)$$

where the real positive modeling parameters  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta_i$ ,  $\lambda_{ik}$ ,  $\kappa_{ik}$  and  $\nu$  are the rate of lockdown, the rate of infection, the rate of recovery from the asymptomatic group, the rate of tested corona-positive corresponding the incubation period  $\tau_i$ , the rate of recovery corresponding the recovery period  $\zeta_k$ , the rate of decease corresponding the decease period  $\eta_k$  and the rate of transit from lockdown compartment to susceptible compartment, respectively. The variables  $S(t-\tau_i)$  and  $I(t-\tau_i)$  denote the cumulative data of  $(t-\tau_i)$  days, i.e., total number of suspected and infected individuals of  $(t-\tau_i)$  days. The factors  $\delta_i \beta S(t-\tau_i)I(t-\tau_i)/N$ ,  $\lambda_{ik} \delta_i \beta S(t-\tau_i-\theta_{ik})I(t-\tau_i-\theta_{ik})/N$ ,  $\kappa_{ik} \delta_i \beta S(t-\tau_i-\mu_{ik})I(t-\tau_i-\mu_{ik})/N$  convey the rate of individuals who were infected  $\tau_i$  days ago, the rate of individuals who were infected  $\tau_i + \theta_{ik}$  days ago and recovered, the rate of individuals who were infected  $\tau_i + \mu_{ik}$  days ago and died, respectively. It follows from (4), that for any  $t$

$$L(t) + S(t) + I(t) + V(t) + C(t) + R(t) + D(t) = N, \quad (5)$$

where  $N$  (constant) is the total population size. We can define a group of new variable  $T_i$  for  $i = 1, \dots, J$  such that

$$T_i = C_i + \sum_{k=1}^M R_{ik} + \sum_{k=1}^M D_{ik}, \quad (6)$$

and

$$T(t) = \sum_{i=1}^J T_i(t) \text{ or } T^{(m)} = \sum_{i=1}^J T_i^{(m)}, \quad (7)$$

where,  $T$ , total confirmed cases, is the group of individuals who tested corona positive (active cases + recovered + deaths).

From Eq.(4) we can generate three different sets of coupled delay differential equations for  $i = 1, \dots, J$  and  $k = 1, \dots, M$

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} - \alpha S + \nu L \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \beta \sum_{i=1}^J \delta_i \frac{S(t-\tau_i)I(t-\tau_i)}{N} \\ \frac{dT_i}{dt} = \delta_i \beta \frac{S(t-\tau_i)I(t-\tau_i)}{N} \\ \frac{dL}{dt} = \alpha S - \nu L \end{cases} \quad (8)$$

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} - \alpha S + \nu L \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \beta \sum_{i=1}^J \delta_i \frac{S(t-\tau_i)I(t-\tau_i)}{N} \\ \frac{dR_{ik}}{dt} = \lambda_{ik} \delta_i \beta \frac{S(t-\tau_i-\theta_{ik})I(t-\tau_i-\theta_{ik})}{N} \\ \frac{dL}{dt} = \alpha S - \nu L \end{cases} \quad (9)$$

and

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} - \alpha S + \nu L \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \beta \sum_{i=1}^J \delta_i \frac{S(t-\tau_i)I(t-\tau_i)}{N} \\ \frac{dD_{ik}}{dt} = \kappa_{ik} \delta_i \beta \frac{S(t-\tau_i-\mu_{ik})I(t-\tau_i-\mu_{ik})}{N} \\ \frac{dL}{dt} = \alpha S - \nu L \end{cases} \quad (10)$$

where Eq.(8), (9) and (10) can be used to calculate incubation period<sup>3</sup>, recovery period and decease period, respectively. In the present context, we focus on recovery as well as decease periods. We solve Eq.(8), (9) and (10) using `matlab` inner-embedded function `dde23` with particular sets of model parameters. To solve the initial value problem, in the interval  $[t_0, t_1]$ , we consider  $L(t_0)$ ,  $S(t_0)$ ,  $I(t_0)$ ,  $T(t_0)$ ,  $R(t_0)$  and  $D(t_0)$  as follows:

$$\begin{cases} L(t_0) = 0, \\ S(t_0) = N - L(t_0) - I(t_0) - V(t_0) - T(t_0), \\ I(t_0) = q, \\ V(t_0) = 0, \\ T(t_0) = \tilde{T}(t_0), \\ R(t_0) = \tilde{R}(t_0), \\ D(t_0) = \tilde{D}(t_0), \end{cases} \quad (11)$$

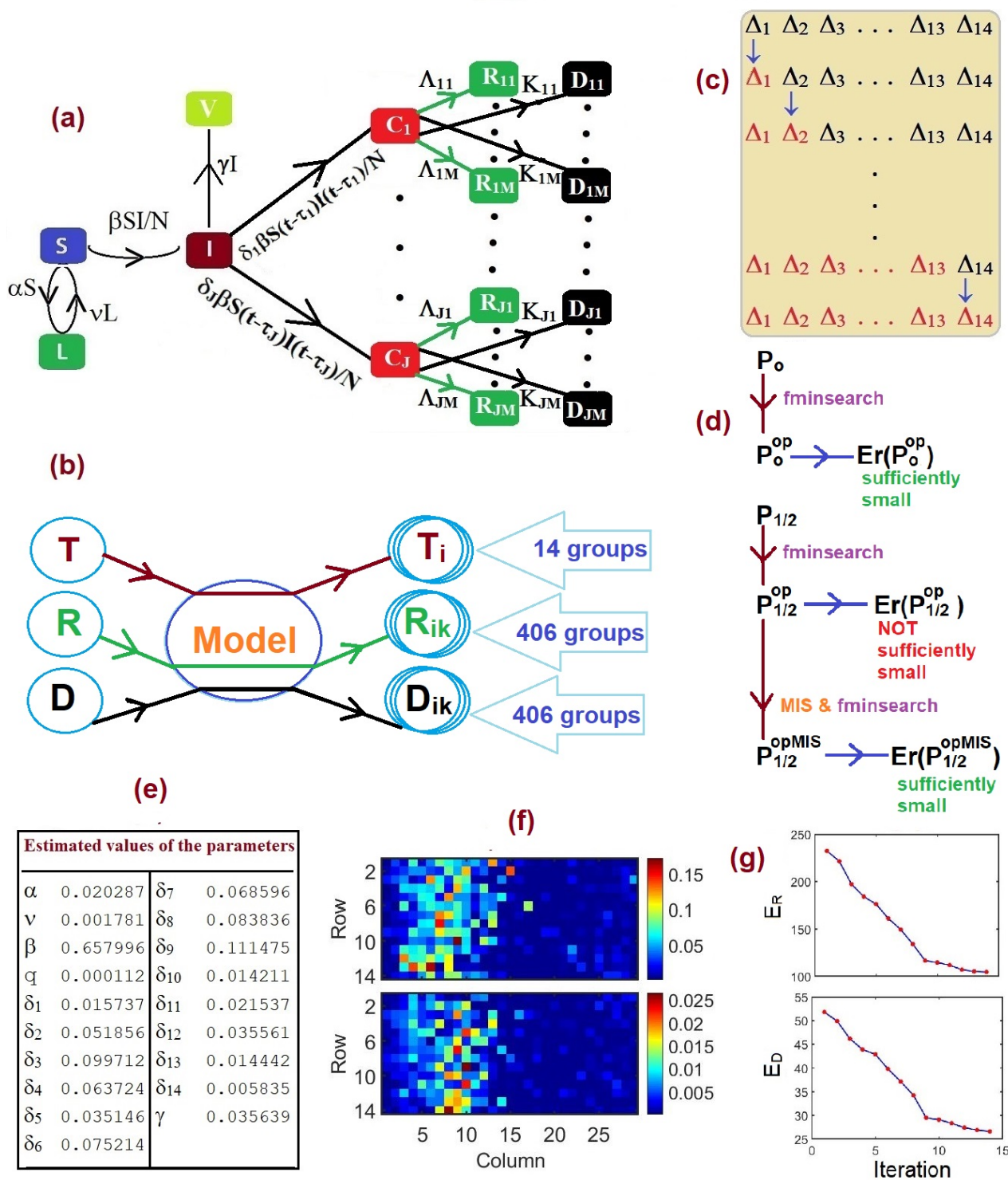
where  $\tilde{T}(t_0)$ ,  $\tilde{R}(t_0)$  and  $\tilde{D}(t_0)$  are the available data at time  $t_0$ , and  $q$  is the initial value adjusting parameters. Initially, there are no lockdown individual and no removed individuals from asymptomatic group so that we can consider  $L(t_0) = 0$  and  $V(t_0) = 0$ . It follows from (7) and (11)

$$\sum_{i=1}^J T_i(t_0) = T(t_0) = \tilde{T}(t_0). \quad (12)$$

In the present context  $\tilde{T}(t_0) = 0$ , since there were no corona-positive cases reported on January 22, 2020. As a consequence, we also take  $T_i(t_0) = 0$  for  $i = 1, 2, \dots, J$ , and the similar assumptions are valid for  $R_{ik}(t_0)$  and  $D_{ik}(t_0)$  i.e.,  $R_{ik}(t_0) = 0$  and  $D_{ik}(t_0) = 0$  for  $i = 1, 2, \dots, J$  and  $k = 1, 2, \dots, M$ .

### Parameter estimation

We focus on the exponential growth phase of the COVID-19 epidemic in Canada; one can use this approach to estimate the incubation, recovery period and decease periods for any region affected by this infectious disease. The time resolved (daily



**Figure 5. Model, methodology and estimated values of the parameters:** (a) Schematic diagram of the present compartmental based model, total 830 compartments. Here  $\Lambda_{ik} = \lambda_{ik} \delta_i \beta S(t - \tau_i - \theta_{ik}) I(t - \tau_i - \theta_{ik}) / N$  and  $K_{ik} = \kappa_{ik} \delta_i \beta S(t - \tau_i - \mu_{ik}) I(t - \tau_i - \mu_{ik}) / N$  for  $i = 1, 2, \dots, J$  and  $k = 1, 2, \dots, M$ . We consider  $J = 14$  and  $M = 29$ . (b) Bubble diagram of the foundation of the present work, splitting publicly available database, total cases (T), recovered individuals (R) and death toll (D), into myriad groups. (c) Sketch of the Monotonic Iteration Scheme (MIS); for 'recovery' calculation  $\Delta_i = \{\lambda_{ik} | k = 1, 2, \dots, 29\}$  and for 'decease' calculation  $\Delta_i = \{\kappa_{ik} | k = 1, 2, \dots, 29\}$  and  $i = 1, 2, \dots, 14$ . (d) Sketch of the optimization scheme for the primary,  $P_0$ , and secondary,  $P_1$  and  $P_2$ , parameters.  $P_{1/2}$  indicates either  $P_1$  or  $P_2$ . (e) Estimated values of the primary parameters. (f) Estimated values of the secondary parameters, upper panel:  $\lambda_{ik}$  and lower panel:  $\kappa_{ik}$ . (g) Iteration verses error function in MIS, upper panel: estimating  $\lambda_{ik}$  and lower panel: estimating  $\kappa_{ik}$ .

updated) database<sup>27</sup> provides the number of total corona-positive cases, the number of recovered individuals and the death toll. We define two groups of model parameters: primary parameters, the parameters involved in Eq.(8) i.e.,  $q, \alpha, \beta, \gamma, \delta_i$  for  $i = 1, 2, \dots, J$  and  $v$ , and secondary parameters, the parameters involved in Eq.(9) and (10) other than the primary parameters i.e.,  $\lambda_{ik}$  and  $\kappa_{ik}$  for  $i = 1, 2, \dots, J$  and  $k = 1, 2, \dots, M$ . We use the estimated values of the primary parameters to optimize the secondary parameters. The optimal values of the primary parameters  $\mathbf{P}_0 = (q, \alpha, \beta, \delta_1(t), \dots, \delta_J(t), v)^T$ ,  $q$  is the initial value of  $I(t)$ , is obtained by minimizing the error function  $Er(\mathbf{P}_0)$ , defined as

$$Er(\mathbf{P}_0) = \frac{1}{K} \sqrt{\sum_{m=1}^K (T^{(m)}(\mathbf{P}_0) - \tilde{T}^{(m)})^2}, \quad (13)$$

where  $\tilde{T}^{(m)}$  is the available data of total corona-positive cases on the particular  $m$ th day, and  $T^{(m)}$  is the calculated results obtained from the system (8). The integer  $K$ , used in (13), is the size of the data set. Due to the complexity of the error function, the minimization using the `matlab` function `fminsearch` requires a very large number of iterations. We use the similar error functions  $Er(\mathbf{P}_1)$  and  $Er(\mathbf{P}_2)$  to optimize the secondary parameters  $\mathbf{P}_1 = (\lambda_{11}, \dots, \lambda_{JM})^T$  and  $\mathbf{P}_2 = (\kappa_{11}, \dots, \kappa_{JM})^T$ , defined as

$$Er(\mathbf{P}_1) = \frac{1}{K} \sqrt{\sum_{m=1}^K (R^{(m)}(\mathbf{P}_0^{\text{op}}, \mathbf{P}_1) - \tilde{R}^{(m)})^2}, \quad (14)$$

and

$$Er(\mathbf{P}_2) = \frac{1}{K} \sqrt{\sum_{m=1}^K (D^{(m)}(\mathbf{P}_0^{\text{op}}, \mathbf{P}_2) - \tilde{D}^{(m)})^2}, \quad (15)$$

where  $\mathbf{P}_0^{\text{op}}$  is the estimated values of  $\mathbf{P}_0$ ;  $\tilde{R}^{(m)}$  and  $\tilde{D}^{(m)}$  are the available data of total number of recovered individuals and total number of death toll;  $R^{(m)}$  and  $D^{(m)}$  are the calculated results obtained from the Eq.(9) and (10), respectively.

### Numerical experiment

In this section, we present a detailed description of the computational procedure for the proposed model. On 23 January 2020, a 56-year old man admitted to Toronto hospital emergency department in Toronto with a new onset of fever and nonproductive cough, and returning from Wuhan, China, the day prior<sup>31,32</sup>. It is believed this is the first confirmed case of 2019-nCoV in Canada, and according to the government report, the novel coronavirus arrived on the Canadian coast on January 25, 2020, first reported case. The above information suggests that the start date of the current pandemic in Canada is possibly to be January 22, 2020. Additionally, some research studies reported that the estimation of the incubation period is from 2 days to 14 days, and recovery as well as decrease period of COVID-19 is from 2 weeks to 6 weeks<sup>2,33</sup>. As a consequence, in the present study we consider  $J = 14$  i.e., 14 delays for the incubation period, and  $M = 29$  i.e., 29 delays for the recovery as well as decrease periods. Here we consider a calculation of 177 days, from January 22, 2020 to July 16, 2020, duration of the first wave in Canada. The purpose of the model is to separate the publicly available database  $T$ ,  $R$  and  $D$  into myriad groups  $T_i$ ,  $R_{ik}$  and  $D_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  (Fig. 5(b)).

Then the local minimum computed by the optimization algorithm depends on the initial values of the parameters: for  $q, \alpha, \beta, v$  we consider any positive random number less than unity, where as a choice of  $\delta = (\delta_1, \dots, \delta_{14})^T$  is tricky. For this purpose, we consider a vector of 14 positive random numbers  $\delta$  such that  $\delta_1 < \dots < \delta_4 < \delta_5 > \delta_6 > \dots > \delta_{14}$  and  $\sum_{i=1}^{14} \delta_i = 0.9$ . We observe, from numerous numerical experiments, the renormalization factor 0.9 works perfectly for the computation. The estimated values of the primary parameters  $\mathbf{P}_0^{\text{op}}$  are presented in Fig.5(e), and the value of the error function  $Er(\mathbf{P}_0) = 41.64$ . The estimated values of the primary parameters are related to Eq.(8), the set of coupled delay differential equations, and Eq.(13), the error function. Using the estimated values of the primary parameters, we optimize the secondary parameters  $\lambda_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  related to Eq.(9) and (14). The choice of the initial values of  $\lambda_{ik}$  is such that for any fixed  $i, i = 1, 2, \dots, 14$ , the first fourteen  $\lambda_{ik}$ s i.e.,  $\{\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{i14}\}$  are in ascending order, and the rest i.e.,  $\{\lambda_{i15}, \lambda_{i16}, \dots, \lambda_{i29}\}$  are in descending order; and  $\sum_{k=1}^{29} \lambda_{ik} = 0.72$ . After optimization, we obtain the value of the error function  $Er(\mathbf{P}_1^{\text{op}}) = 236.47$ . Using the estimated values of the primary parameters, we optimize the secondary parameters  $\kappa_{ik}$  for  $i = 1, 2, \dots, 14$  and  $k = 1, 2, \dots, 29$  related to Eq.(10) and (15). The choice of the initial values of  $\kappa_{ik}$  is such that for any fixed  $i, i = 1, 2, \dots, 14$ , the first fourteen  $\kappa_{ik}$ s i.e.,  $\{\kappa_{i1}, \kappa_{i2}, \dots, \kappa_{i14}\}$  are in ascending order, and the rest i.e.,  $\{\kappa_{i15}, \kappa_{i16}, \dots, \kappa_{i29}\}$  are in descending order; and  $\sum_{k=1}^{29} \kappa_{ik} = 0.1$ . After optimization, we obtain the value of the error function  $Er(\mathbf{P}_2^{\text{op}}) = 52.82$ . The values of the error functions  $Er(\mathbf{P}_1^{\text{op}})$  and  $Er(\mathbf{P}_2^{\text{op}})$  are not sufficiently small. To overcome that difficulties, here, we introduce a Monotonic Iteration Scheme (MIS).

### Monotonic Iteration Scheme

To optimize the parameters  $\xi_{ik}$  for  $i = 1, 2, \dots, J$  and  $k = 1, 2, \dots, M$ , we use a MIS with  $J = 14$  and  $M = 29$ . However, the method can be applied for any finite integer values of  $J$  and  $M$ . The schematic diagram of MIS is presented in Fig.5(c), and consists of the following steps.

- **Step 1:** We decompose the parametric domain  $\Delta = \{\xi_{ik} | i = 1, 2, \dots, 14; k = 1, 2, \dots, 29\}$  into 14 subdomains  $\Delta_i = \{\xi_{ik} | k = 1, 2, \dots, 29\}$  so  $\Delta = \{\Delta_1, \Delta_2, \Delta_3, \dots, \Delta_{14}\}$ .
- **Step 2:** We optimize the subdomain  $\Delta_1$  and consider the other parameters  $\Delta_2, \Delta_3, \dots, \Delta_{14}$ , as constants. After first iteration, we get estimated parameters  $\Delta_1^{\text{op}}$ ; the entire parametric domain is  $\Delta^{(1)} = \{\Delta_1^{\text{op}}, \Delta_2, \Delta_3, \dots, \Delta_{14}\}$ , and the error function  $Er(\Delta^{(1)})$ .
- **Step 3:** In the second iteration, we optimize the subdomain  $\Delta_2$  and keeping the other subdomains of  $\Delta^{(1)}$  unchanged. After second iteration, we get estimated parameters  $\Delta_2^{\text{op}}$ ; the entire parametric domain is  $\Delta^{(2)} = \{\Delta_1^{\text{op}}, \Delta_2^{\text{op}}, \Delta_3, \dots, \Delta_{14}\}$ , and the error function  $Er(\Delta^{(2)})$ .
- **Step 4:** Repeated the same procedure discussed in **Step 3**.

The optimization of the subdomain  $\Delta_2$ , demonstrated in **Step 3**, is related to minimizing the error function such that  $Er(\Delta^{(1)}) \geq Er(\Delta^{(2)})$ ; the equality sign holds for  $\Delta_2^{\text{op}} = \Delta_2$ . The error function of the  $n + 1$ th iteration,  $Er(\Delta^{(n+1)})$  cannot be greater than that of  $n$ th iteration,  $Er(\Delta^{(n)})$ , because of this characteristic of the error function we define the approach as MIS. The flow chart of the optimization scheme is presented in Fig. 5(d). The upper and lower panels of Fig.5(f) show the estimated values of the secondary parameters  $\lambda_{ik}$  and  $\kappa_{ik}$ , respectively, obtained from the MIS. The upper and lower panels of Fig.5(g) show the values of the error functions  $E_R$ , using MIS to optimize  $\lambda_{ik}$ , and  $E_D$ , using MIS to optimize  $\kappa_{ik}$ , for fourteen iteration steps and  $Er(\mathbf{P}_1^{\text{opMIS}}) = 104.07$ ,  $Er(\mathbf{P}_2^{\text{opMIS}}) = 26.66$  where  $\mathbf{P}_1^{\text{opMIS}}$  and  $\mathbf{P}_2^{\text{opMIS}}$  are the estimated values of  $\mathbf{P}_1^{\text{op}}$  and  $\mathbf{P}_2^{\text{op}}$ , respectively, using MIS. Fig. 5(g) shows that the MIS works efficiently to get better estimations.

### References

1. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in wuhan, China. *The lancet* **395**, 497–506 (2020).
2. Organization, W. H. Rolling updates on coronavirus disease (COVID-19).[cited 2020 april 14] available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>. (2020).
3. Paul, S. & Lorin, E. Distribution of incubation periods of COVID-19 in the Canadian context. *Sci. Reports* **11**, 1–9 (2021).
4. Voinsky, I., Baristaite, G. & Gurwitz, D. Effects of age and sex on recovery from COVID-19: Analysis of 5769 israeli patients. *J. Infect.* **81**, e102–e103 (2020).
5. Barman, M. P., Rahman, T., Bora, K. & Borgohain, C. COVID-19 pandemic and its recovery time of patients in India: A pilot study. *Diabetes & Metab. Syndr. Clin. Res. & Rev.* **14**, 1205–1211 (2020).
6. Cai, Q. *et al.* COVID-19 in a designated infectious diseases hospital outside Hubei Province, China. *Allergy* **75**, 1742–1752 (2020).
7. Fang, Z. *et al.* Comparisons of viral shedding time of SARS-CoV-2 of different samples in ICU and non-ICU patients. *J. Infect.* **81**, 147–178 (2020).
8. Wu, Y. *et al.* Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The lancet Gastroenterol. & hepatology* **5**, 434–435 (2020).
9. Bi, Q. *et al.* Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infect. Dis.* **20**, 911–919 (2020).
10. SeyedAlinaghi, S. *et al.* Predictors of the prolonged recovery period in COVID-19 patients: a cross-sectional study. *Eur. J. Med. Res.* **26**, 1–10 (2021).
11. Byrne, A. W. *et al.* Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ open* **10**, e039856 (2020).
12. Paul, S. & Lorin, E. Lockdown: a non-pharmaceutical policy to prevent the spread of COVID-19. Mathematical modeling and computation. *Res. Gate preprint* DOI: [10.13140/RG.2.2.13857.22882](https://doi.org/10.13140/RG.2.2.13857.22882) (2020).

13. Bouchnita, A. & Jebrane, A. A hybrid multi-scale model of COVID-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. *Chaos, Solitons & Fractals* 109941 (2020).
14. Ng, K. Y. & Gui, M. M. COVID-19: Development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and resusceptibility. *Phys. D: Nonlinear Phenom.* 132599 (2020).
15. Eshragh, A., Alizamir, S., Howley, P. & Stojanovski, E. Modeling the dynamics of the COVID-19 population in Australia: A probabilistic analysis. *arXiv preprint arXiv:2005.12455* (2020).
16. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases* (2020).
17. Chen, T.-M. *et al.* A mathematical model for simulating the phase-based transmissibility of a novel coronavirus. *Infect. diseases poverty* **9**, 1–8 (2020).
18. Tang, B. *et al.* Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *J. clinical medicine* **9**, 462 (2020).
19. Sun, G.-Q. *et al.* Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources. *Nonlinear Dyn.* **101**, 1981–1993 (2020).
20. Bastos, S. B. & Cajueiro, D. O. Modeling and forecasting the early evolution of the COVID-19 pandemic in Brazil. *Sci. Reports* **10**, 1–10 (2020).
21. Qaddura, Y. & Mavinga, N. Analysis of a vector-borne diseases model with a two-lag delay differential equation. *The North Carol. J. Math. Stat.* **4**, 12–28 (2018).
22. Vyasarayani, C. P. & Chatterjee, A. New approximations, and policy implications, from a delayed dynamic model of a fast pandemic. *arXiv preprint arXiv:2004.03878* (2020).
23. Yin, Z., Yu, Y. & Lu, Z. Stability analysis of an age-structured seirs model with time delay. *Mathematics* **8**, 455 (2020).
24. Menendez, J. Elementary time-delay dynamics of COVID-19 disease. *medRxiv* (2020).
25. Chen, Y., Cheng, J., Jiang, Y. & Liu, K. A time delay dynamic system with external source for the local outbreak of 2019-ncov. *Appl. Analysis* 1–12 (2020).
26. Dell’Anna, L. Solvable delay model for epidemic spreading: the case of COVID-19 in Italy. *Sci. reports* **10**, 1–10 (2020).
27. <https://resources-covid19canada.hub.arcgis.com/datasets/case-accumulation/data>.
28. Tan, W. Y. T., Wong, L. Y., Leo, Y. S. & Toh, M. P. H. S. Does incubation period of COVID-19 vary with age? a study of epidemiologically linked cases in singapore. *Epidemiol. & Infect.* **148** (2020).
29. Statgraphics, “centurion XIX Version 19.2.01. March 10, 2021,” statgraphics technologies, inc. <https://www.statgraphics.com>.
30. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Medicine* **26**, 1205–1211 (2020).
31. Marchand-Sénécal, X. *et al.* Diagnosis and management of first case of COVID-19 in Canada: Lessons applied from SARS. *Clin. Infect. Dis.* (2020).
32. Silverstein, W. K., Stroud, L., Cleghorn, G. E. & Leis, J. A. First imported case of 2019 novel coronavirus in Canada, presenting as mild pneumonia. *The Lancet* **395**, 734 (2020).
33. <https://www.worldometers.info/coronavirus/coronavirus-incubation-period/>.

## Data availability

The data used to estimate the model parameters are publicly available and are available with the code.

## Code availability

All code is available in the GitHub repository for the project at <https://github.com/SPAUL2021/COVID19RECOVERY/tree/SPAUL2021-patch-1>

## Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).



## **Author contributions statement**

S.P. has derived the model, has developed the matlab code, has analyzed the calculated results, and has prepared all figures and tables. S.P. and E.L. have drafted the original article. Both authors have contributed to the editing of the article. Both authors have read and approved the final article.

## **Competing interests**

The authors declare that they have no competing interests.

## **Additional information**

Correspondence and requests for materials should be addressed to S.P.