medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A new estimation method for COVID-19 time-varying reproduction number using active cases

Agus Hasan[1,*], Hadi Susanto[2,3], Venansius Tjahjono[4], Rudy Kusdiantara[5], Endah Putri[4], Panji Hadisoemarto[6], and Nuning Nuraini[5]

[1]Mærsk McKinney Møller Institute, University of Southern Denmark, Odense, Denmark
[2]Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates
[3]Department of Mathematical Sciences, University of Essex, Colchester, United Kingdom
[4]Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[5]Department of Mathematics, Institut Teknologi Bandung, Bandung, Indonesia
[6]School of Medicine, Universitas Padjadjaran, Sumedang, Indonesia
[*]agha@mmmi.sdu.dk

ABSTRACT

We propose a new method to estimate the time-varying effective (or instantaneous) reproduction number of the novel coronavirus disease (COVID-19). The method is based on a discrete-time stochastic augmented compartmental model that describes the virus transmission. A two-stage estimation method, which combines the Extended Kalman Filter (EKF) to estimate the reported state variables (active and removed cases) and a low pass filter based on a rational transfer function to remove short term fluctuations of the reported cases, is used with case uncertainties that are assumed to follow a Gaussian distribution. Our method does not require information regarding serial intervals, which makes the estimation procedure simpler without reducing the quality of the estimate. We show that the proposed method is comparable to common approaches, e.g., age-structured and new cases based sequential Bayesian models. We also apply it to COVID-19 cases in the Scandinavian countries: Denmark, Sweden, and Norway, where we see a delay of about four days in predicting the epidemic peak.

Introduction

The coronavirus disease 2019 (COVID-19), i.e., a disease outbreak of atypical pneumonia that originated from Wuhan, China[1], has caused globally at least 90 million confirmed cases, including an estimated 2 million deaths in approximately 216 countries by the end of 2020. The World Health Organization declared the COVID-19 crisis a pandemic on 11 March 2020.

In modelling the disease's transmission as well as to inform and evaluate control policies, it is particularly important to estimate its reproduction number. Early estimates for COVID-19 basic reproduction number $\mathscr{R}_0$, that denotes the transmission potential of infectious disease when introduced to a completely susceptible population, ranged from 1.4 to 6.49[2]. The effective (or instantaneous) reproduction number $\mathscr{R}_t$, on the other hand, reflects the extent of transmission in the presence of population immunity or intervention. Thus, the estimation of $\mathscr{R}_t$ is important for evaluating public measure success. However, estimation of $\mathscr{R}_t$ is sensitive to the model structure and parameter assumptions[3]. As a case in point, due to incorporation of more individual case information and travel data, the estimate for $\mathscr{R}_0$ in Wuhan was revised upward from 2.2-2.7 to 5.7[4]. On the other hand, data inavailability or poor quality often hinders the use of certain estimation methods, such as serial interval data that are usually needed to estimate $\mathscr{R}_t$ (e.g., Fraser[5], Wallinga and Teunis[6], Cauchemez et al.[7], White and Pagano[8]).

In the course of calculating the exact value of $\mathscr{R}_t$, especially when the data has not yet reached its peak, precise assumptions and data estimates are needed. Nishiura et al.[9] discussed a likelihood-based approach to estimate $\mathscr{R}_t$ from early epidemic growth data. Using the compartmental Susceptible-Infectious-Recovered (SIR) model, Bettencourt and Ribeiro[10] use the incidence data to estimate $\mathscr{R}_0$ and $\mathscr{R}_t$. In this paper, based on the Susceptible-Infectious-Recovered-Dead (SIRD) model as a reference, we develop a novel approach to estimate $\mathscr{R}_t$ of COVID-19. It uses information on the number of infected or active (I), recovered (R), and death (D) cases, which are readily available for all affected countries, so that they can be accessed rather easily.

The reproduction number is estimated from reported cases under uncertainties using a two-stage estimation method based on the Extended Kalman Filter (EKF) and a low-pass filter. The method not only considers the nominal number of reported cases, but also its daily pattern. To show our method's practical ability, we apply it to COVID-19 cases in the Scandinavian countries, i.e., Denmark, Sweden, and Norway, and compare the results with two commonly used Bayesian methods due to Bettencourt and Ribeiro[10] and Cori et al.[5,11]. We show that the results are indeed comparable.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A Discrete-Time Stochastic Augmented Compartmental Model

Our estimation method is based on the compartmental SIRD model that can be written as the following first-order nonlinear differential equations:

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

$$\dot{I}(t) = \frac{\beta I(t)S(t)}{N} - (\gamma+\kappa)I(t), \tag{2}$$

$$\dot{R}(t) = \gamma I(t), \tag{3}$$

$$\dot{D}(t) = \kappa I(t), \tag{4}$$

where S denotes the number of susceptible individuals, N is the total population, β is the average number of contacts per person per time, γ and κ are the recovery and death rate, respectively. Remark that the value of β is time-varying due to intervention, i.e., β=$\beta(t)$. To use the model, we require information on the average infectious time T_i and the case fatality rate CFR, so that

$$\gamma = \frac{1-\text{CFR}}{T_i}, \kappa = \frac{\text{CFR}}{T_i}. \tag{5}$$

For COVID-19, we take $T_i = 9$ as the infectious period on average lasts for 9 days (i.e., 7-11 days with 95% CI)[12]. The time-varying effective reproduction number is then given by:

$$\mathscr{R}_t(t) = \frac{S(t)}{N}\left(\frac{\beta(t)}{\gamma+\kappa}\right) \approx \frac{\beta(t)}{\gamma+\kappa}. \tag{6}$$

The approximation is under the assumption that government intervention is taken at an early stage so that the susceptible is relatively the same over time as the total population. This is the case especially for emerging diseases. We modify the SIRD model by augmenting the following two equations into the system:

$$\dot{E}(t) = (\gamma+\kappa)I(t) - E(t), \quad \dot{\mathscr{R}}_t(t) = 0. \tag{7}$$

The former equation takes into account the daily number of reported new cases E, while the latter one says that the effective reproduction number $\mathscr{R}_t$ is assumed to be a piece-wise constant function with jump every one day time interval.

Discretizing the model using the forward Euler method, we obtain the following discrete-time augmented SIRD model:

$$S(k+1) = \left(1 - \frac{(\gamma+\kappa)\Delta t}{N}\mathscr{R}_t(k)I(k)\right)S(k), \tag{8}$$

$$I(k+1) = (1-(\gamma+\kappa)\Delta t)I(k) + \frac{(\gamma+\kappa)\Delta t}{N}\mathscr{R}_t(k)I(k)S(k), \tag{9}$$

$$R(k+1) = R(k) + \gamma\Delta t I(k), \tag{10}$$

$$D(k+1) = D(k) + \kappa\Delta t I(k), \tag{11}$$

$$E(k+1) = (\gamma+\kappa)\Delta t I(k) + (1-\Delta t)E(k), \tag{12}$$

$$\mathscr{R}_t(k+1) = \mathscr{R}_t(k). \tag{13}$$

Our method computes a new estimate of $\mathscr{R}_t$ based on new reported cases. Since their frequency is low (could be once a day), the reported data can be interpolated using, e.g., a modified Akima cubic Hermite interpolation, such that it fits with the time step Δt. In our simulation, the time step Δt is chosen as 0.01, i.e., 100 time discretization within one day interval. The confidence interval of our estimated $\mathscr{R}_t$ is determined by computing the reproduction number for different values of the infectious period T_i within a certain interval.

To simplify the presentation, we define the augmented state vector

$$\boldsymbol{x}(k+1) = \begin{pmatrix} S(k+1) \\ I(k+1) \\ R(k+1) \\ D(k+1) \\ E(k+1) \\ \mathscr{R}_t(k+1) \end{pmatrix}, \tag{14}$$

and as such, the discrete-time augmented SIRD model (8)-(13) can be written as follows

$$\boldsymbol{x}(k+1) = \boldsymbol{f}(\boldsymbol{x}(k)) + \boldsymbol{w}(k), \tag{15}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

where $\boldsymbol{f}$ is the nonlinear term written in the right hand side of (8)-(13) and $\boldsymbol{w}$ is introduced as an uncertainty to model the inaccuracies due to simplification in the modelling. The uncertainty is assumed to be a zero mean Gaussian white noise with known covariance $\boldsymbol{Q}_F$. In practice, $\boldsymbol{Q}_F$ can be considered as a tuning parameter for the EKF. Thus, the transmission model becomes a discrete-time stochastic augmented SIRD model.

Reported cases, such as the number of active cases and the cumulative numbers of recovered and death, can be incorporated into the model using the following output vector

$$\boldsymbol{y}(k+1) = \boldsymbol{C}\boldsymbol{x}(k) + \boldsymbol{v}(k). \tag{16}$$

Here, $\boldsymbol{v}$ denotes uncertainties due to false testing results. We also assume the uncertainty to be a zero mean Gaussian white noise with known covariance $\boldsymbol{R}_F$. As well as $\boldsymbol{Q}_F$, $\boldsymbol{R}_F$ can also be considered as a tuning parameter. Following the available data that include I, R, D, and E, the data/measurement matrix $\boldsymbol{C}$ is taken to be

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{17}$$

A Two-Stage Filtering Method

A two-stage filtering method is used to estimate the daily reproduction number $\mathscr{R}_t$. The method consists of the EKF and a low-pass filter. In the first stage of estimation, the EKF is used to estimate the state variables and the value of $\mathscr{R}_t$ under uncertainties in the number of reported cases. Afterwards, the low pass filter is used to remove short term fluctuations of the reported cases that can be caused by delays in the reporting. For example, suddenly in Denmark there were 893 recovered patients reported on 1 April 2020, in contrast to the previous days from 16 February 2020 onwards when there was no recovery reported at all. Such an accumulated delay can cause a falsely decreasing value of $\mathscr{R}_t$.

The EKF is an extension of Kalman filter for nonlinear systems. The Kalman filter itself is based on a recursive Bayesian estimation and is an optimal linear filter. The idea of EKF is to linearize the non-linearity around its estimate. Due to that linearization, the optimality and stability of the EKF cannot be guaranteed. However, if the non-linearity is not severe, the EKF can give a reasonably good estimate.

Let us denote $\hat{\boldsymbol{x}}(k)$ as an estimated vector state from the EKF. Applying first-order Taylor series expansion to $\boldsymbol{f}$ at $\hat{\boldsymbol{x}}(k)$, we obtain

$$\boldsymbol{f}(\boldsymbol{x}(k)) = \boldsymbol{f}(\hat{\boldsymbol{x}}(k)) + \boldsymbol{J}_{\boldsymbol{f}}(\hat{\boldsymbol{x}}(k))(\boldsymbol{x}(k) - \hat{\boldsymbol{x}}(k)), \tag{18}$$

where $\boldsymbol{J}_{\boldsymbol{f}}(\hat{\boldsymbol{x}}(k))$ is the Jacobian matrix of $\boldsymbol{f}$, given by:

$$\boldsymbol{J}_{\boldsymbol{f}}(\hat{\boldsymbol{x}}(k)) = \begin{pmatrix} J_{11}(\hat{\boldsymbol{x}}(k)) & J_{12}(\hat{\boldsymbol{x}}(k)) & 0 & 0 & 0 & J_{16}(\hat{\boldsymbol{x}}(k)) \\ J_{21}(\hat{\boldsymbol{x}}(k)) & J_{22}(\hat{\boldsymbol{x}}(k)) & 0 & 0 & 0 & J_{26}(\hat{\boldsymbol{x}}(k)) \\ 0 & \gamma\Delta t & 1 & 0 & 0 & 0 \\ 0 & \kappa\Delta t & 0 & 1 & 0 & 0 \\ 0 & (\gamma+\kappa)\Delta t & 0 & 0 & 1-\Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{19}$$

where

$$J_{11}(\hat{\boldsymbol{x}}(k)) = 1 - \frac{(\gamma+\kappa)\Delta t}{N}\hat{\mathscr{R}}_t(k)\hat{I}(k), \tag{20}$$

$$J_{12}(\hat{\boldsymbol{x}}(k)) = -\frac{(\gamma+\kappa)\Delta t}{N}\hat{\mathscr{R}}_t(k)\hat{S}(k), \tag{21}$$

$$J_{16}(\hat{\boldsymbol{x}}(k)) = -\frac{(\gamma+\kappa)\Delta t}{N}\hat{I}(k)\hat{S}(k), \tag{22}$$

$$J_{21}(\hat{\boldsymbol{x}}(k)) = \frac{(\gamma+\kappa)\Delta t}{N}\hat{\mathscr{R}}_t(k)\hat{I}(k), \tag{23}$$

$$J_{22}(\hat{\boldsymbol{x}}(k)) = 1 - (\gamma+\kappa)\Delta t + \frac{(\gamma+\kappa)\Delta t}{N}\hat{\mathscr{R}}_t(k)\hat{S}(k), \tag{24}$$

$$J_{26}(\hat{\boldsymbol{x}}(k)) = \frac{(\gamma+\kappa)\Delta t}{N}\hat{I}(k)\hat{S}(k). \tag{25}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The EKF consists of two steps: predict and update. The discrete-time stochastic augmented SIRD model is used to predict the next state and covariance and update them after obtaining new data/measurement. The EKF can be considered as one of the simplest dynamic Bayesian networks. While the EKF calculates estimates of the true values of states recursively over time using incoming measurements and a mathematical process model, recursive Bayesian estimation calculates estimates of an unknown probability density function recursively over time using incoming measurements and a mathematical process model[13]. Let $\hat{\boldsymbol{x}}(n|m)$ denotes the estimate of $\boldsymbol{x}$ at time n given observations up to and including at time $m \leq n$. The Kalman filter algorithm is given as follows[14]

Predict

$$\hat{\boldsymbol{x}}(k+1|k) = \boldsymbol{f}(\hat{\boldsymbol{x}}(k|k)) \tag{26}$$

$$\boldsymbol{P}(k+1|k) = \boldsymbol{J_f}(\hat{\boldsymbol{x}}(k|k))\boldsymbol{P}(k|k)\boldsymbol{J_f}(\hat{\boldsymbol{x}}(k|k))^{\mathsf{T}} + \boldsymbol{Q}_F(k) \tag{27}$$

Update

$$\tilde{\boldsymbol{y}}(k+1) = \boldsymbol{y}(k+1) - \boldsymbol{C}\hat{\boldsymbol{x}}(k+1|k) \tag{28}$$

$$\boldsymbol{K}(k+1) = \boldsymbol{P}(k+1|k)\boldsymbol{C}^{\mathsf{T}}\left(\boldsymbol{C}\boldsymbol{P}(k+1|k)\boldsymbol{C}^{\mathsf{T}} + \boldsymbol{R}_F(k)\right)^{-1} \tag{29}$$

$$\hat{\boldsymbol{x}}(k+1|k+1) = \hat{\boldsymbol{x}}(k+1|k) + \boldsymbol{K}(k+1)\tilde{\boldsymbol{y}}(k+1) \tag{30}$$

$$\boldsymbol{P}(k+1|k+1) = \left(\boldsymbol{I} - \boldsymbol{K}(k+1)\boldsymbol{C}\right)\boldsymbol{P}(k+1|k) \tag{31}$$

Here $\boldsymbol{P}(k|k)$ denotes a posteriori estimate covariance matrix. In the second stage, a low pass filter based on a rational transfer function is used to remove short term fluctuation at time step k, and is given by

$$\hat{\boldsymbol{y}}(k) = \frac{1}{y_n}\left(\hat{\boldsymbol{x}}(k) + \hat{\boldsymbol{x}}(k) + \cdots + \hat{\boldsymbol{x}}(k - y_n + 1)\right), \tag{32}$$

where y_n is a window length along the data. In our case, we choose $y_n = \frac{3}{\Delta t}$.

To evaluate the quality of the estimate, we calculate a Relative Root Mean Square Error (RRMSE) between the estimated and reported cases. The RRMSE is defined as

$$\text{RRMSE} = \frac{1}{N_d}\sum_{i=1}^{N_d}\frac{\|X_i - \hat{X}_i\|^2}{\|X_i\|^2}, \tag{33}$$

where N_d is the number of observed days and $X \in \{I, D, R, E\}$.

Case study: Scandinavian countries

In this section, we apply our method to study viral transmission of COVID-19 in Denmark, Sweden, and Norway. All datasets and MATLAB code are available on GitHub (https://github.com/agusisma/covid19). As the country of residence of the corresponding author, Denmark has a high estimated percentage of reported symptomatic COVID-19 cases, i.e., 61% (42%-87% of 95% CI). Norway is at 73% with 43%-98% of 95% CI. On the contrary, Sweden only has 15% (11%-19% of 95% CI). The countries also have a different approach in their public measures in responding to COVID-19, i.e., Sweden did not implement a strict lockdown, unlike its Nordic neighbouring countries.

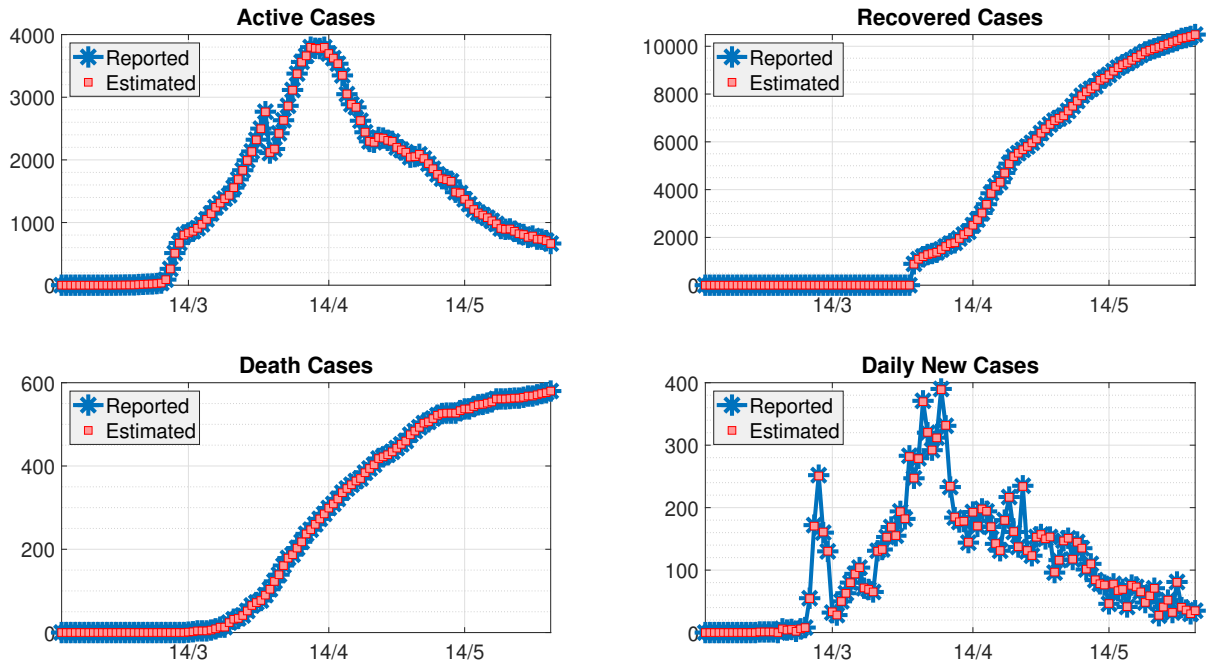
We plot the observed incidence of COVID-19 in Denmark, Sweden, and Norway in Fig. 1. We also plot in the same figure estimated numbers computed using our method, where good agreement is obtained. For all estimation, the process and observation covariance matrices are considered as tuning parameters and are chosen as $\boldsymbol{Q}_F = \text{diag}(10\ 10\ 10\ 10\ 5\ 0.2)$ and $\boldsymbol{R}_F = \text{diag}(100\ 10\ 10\ 5\ 1)$, respectively. These parameters are chosen such that the RRMSE between the estimated and reported data are sufficiently small. In our case study, the RRMSE are shown in Table 1.

	RRMSE				
Country	I	R	D	E	Total
Denmark	7.8943e-05	0.1396	1.1640e-04	6.0490e-05	0.1399
Sweden	4.9208e-05	0.0155	1.3375e-04	0.0102	0.0259
Norway	0.0011	0.0682	7.9477e-05	0.1631	0.2326

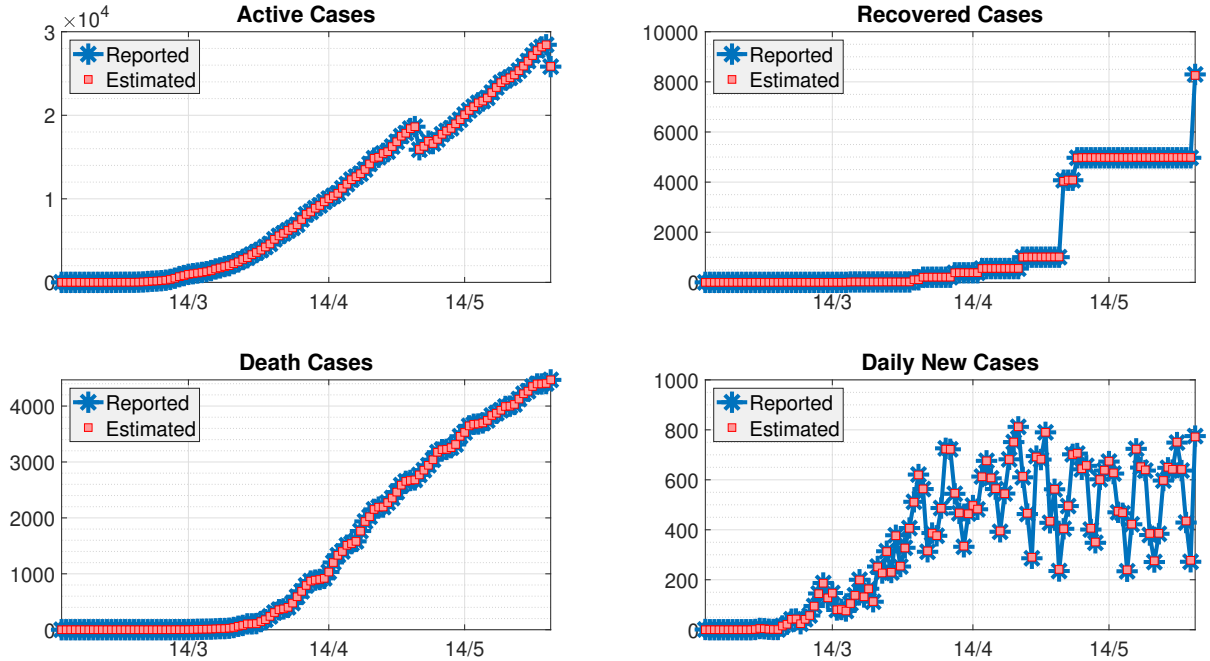
Table 1. RRMSE of the two-stage filtering method for Scandinavian countries.

In applying our method, we also compare it with two commonly used methods to estimate transmission parameters, namely the sequential Bayesian method of Bettencourt and Ribeiro[10] that provides an approximation of the basic reproduction number,

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



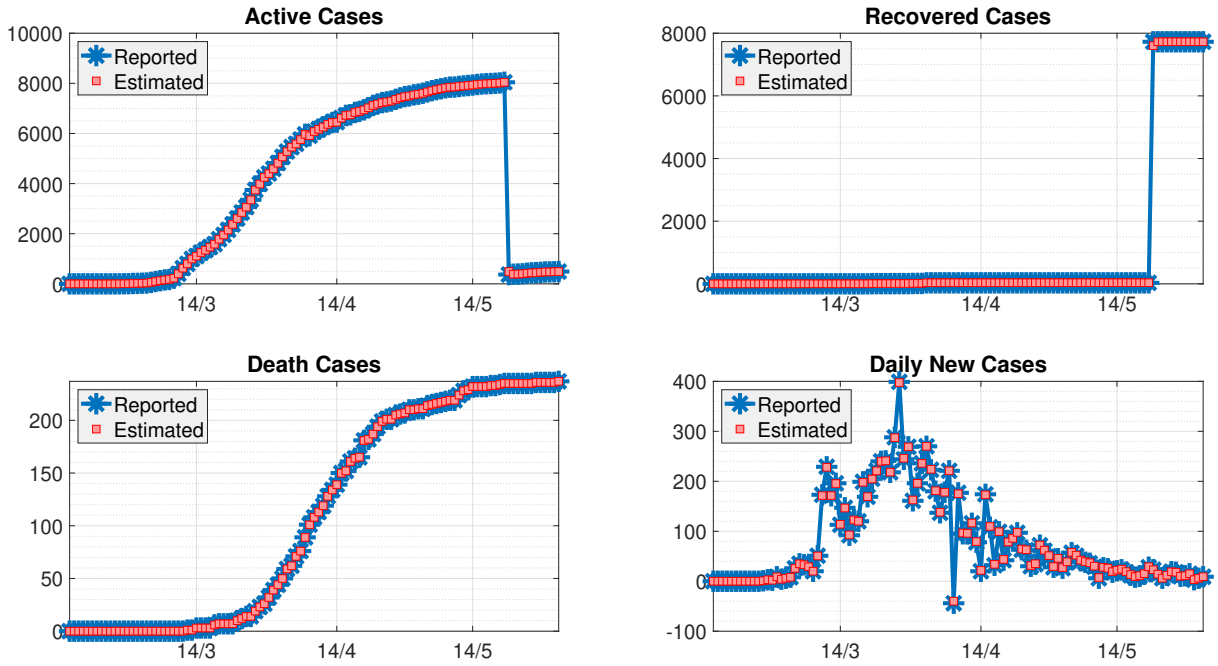
(a)



(b)

Figure 1. (Continued)

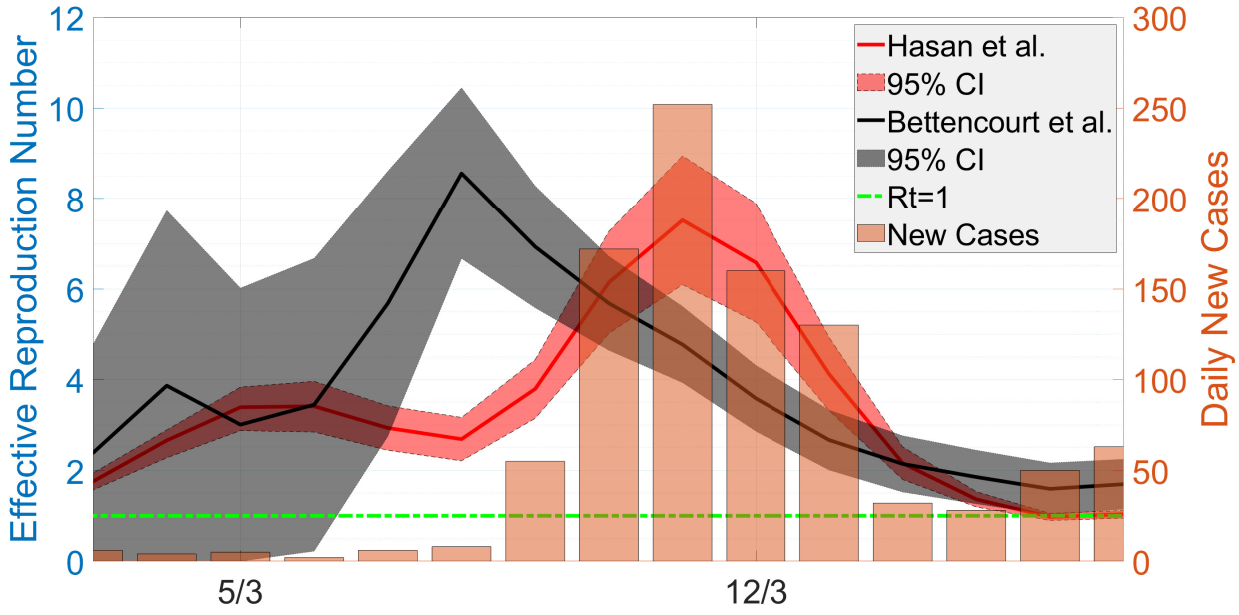
medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



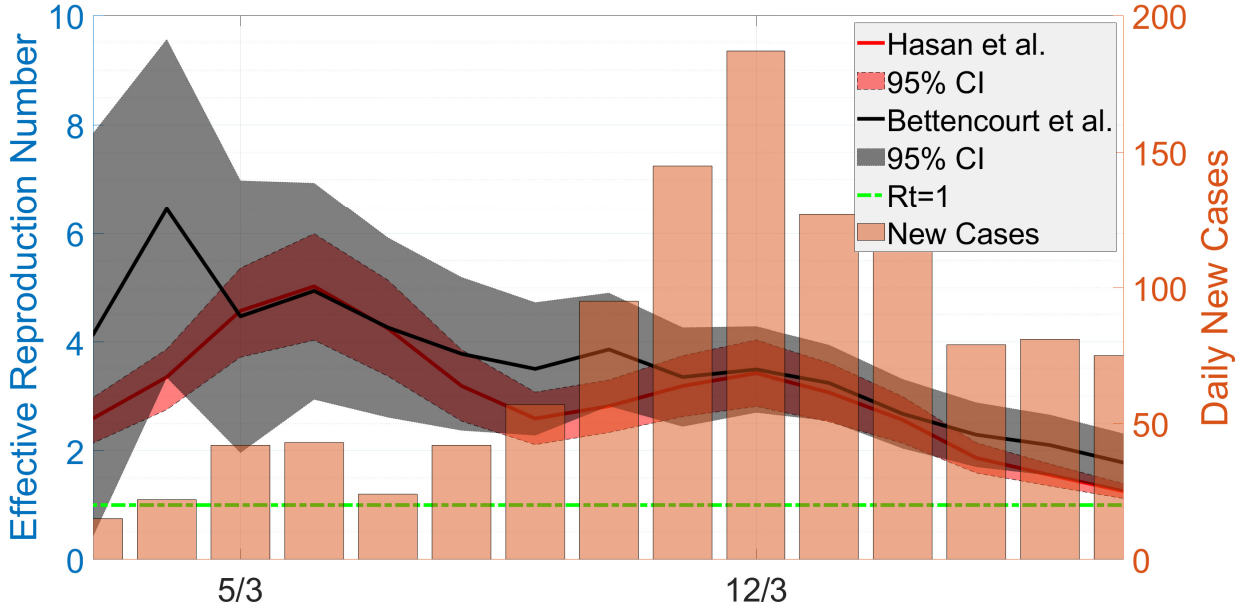
(c)

Figure 1. Active (I), Recovered (R), Death (D), and daily new cases (E) in (a) Denmark; (b) Sweden, and (c) Norway. Plotted are reported data and model fit estimated using our method.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



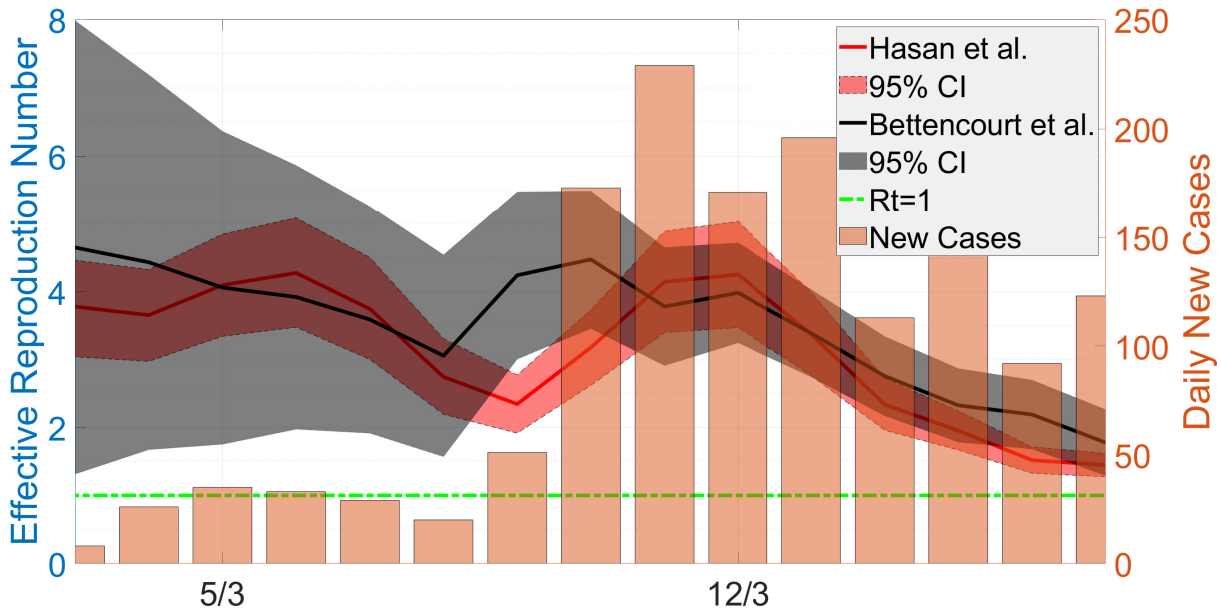
(a)



(b)

Figure 2. (Continued)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



(c)

Figure 2. The estimated reproduction number at the early stage of the pandemic in (a) Denmark, (b) Sweden, and (c) Norway using our proposed method and[10].

	Current method	Bettencourt & Ribeiro[10]
Denmark	7.5 [95% CI: 6.1-8.9]	8.6 [95% CI: 6.7-10.5]
Sweden	5.0 [95% CI: 4.1-5.9]	6.5 [95% CI: 3.3-9.6]
Norway	4.3 [95% CI: 3.5-5.1]	4.6 [95% CI: 1.3-7.9]

Table 2. Estimates of the basic reproduction number $\mathscr{R}_0$ of the countries by our method and Bettencourt and Ribeiro[10].

and the instantaneous method by Fraser[5] that is implemented with a Bayesian analysis[11,15]. The former method exploits the new reported incidence, while the latter one uses the distribution of the serial interval.

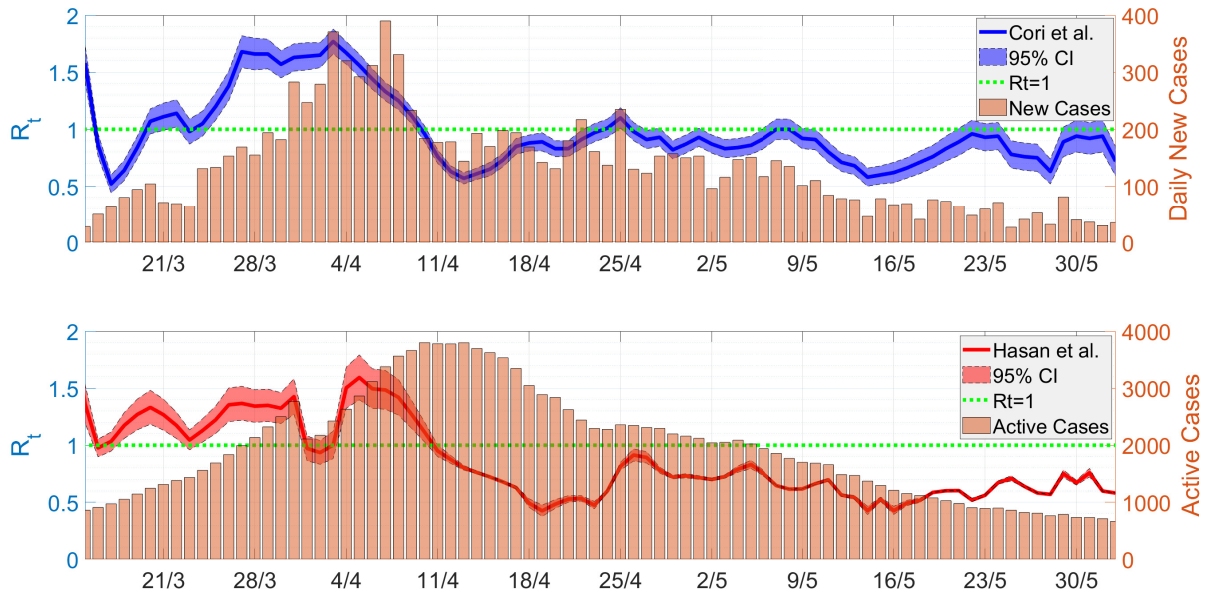
First, we compare our method with that of Bettencourt and Ribeiro[10], that allows sequential estimation of the basic reproduction number at the initial stage, i.e., when the growth is still exponential. While the two methods are based on the SIR model, Bettencourt and Ribeiro[10] rather use new incidence data, which in here need to be smoothed out using a five-day moving average filter. In Fig. 2, we plot the comparison and summarise the basic reproduction numbers that are taken to be the maximum of the curves in Table 2. It is interesting to note how the methods give rather similar estimations. This indicates that our method gives comparable results to those of[10].

Finally, we plot the time-varying effective reproduction number in Fig. 3. Here, we compare our results with those using Cori et al.[11]. The method of[11] utilises the disease serial interval, which we approximate using a shifted Gamma distribution[11] with mean 4.7 and standard deviation 2.9[16]. The prior belief for the value of $\mathscr{R}_t$ is taken to be Gamma function with mean and standard deviation 5. We do not average out the data of daily new cases, but instead take the likelihood estimation of a new case at one day to depend also on the estimation of the previous three days.

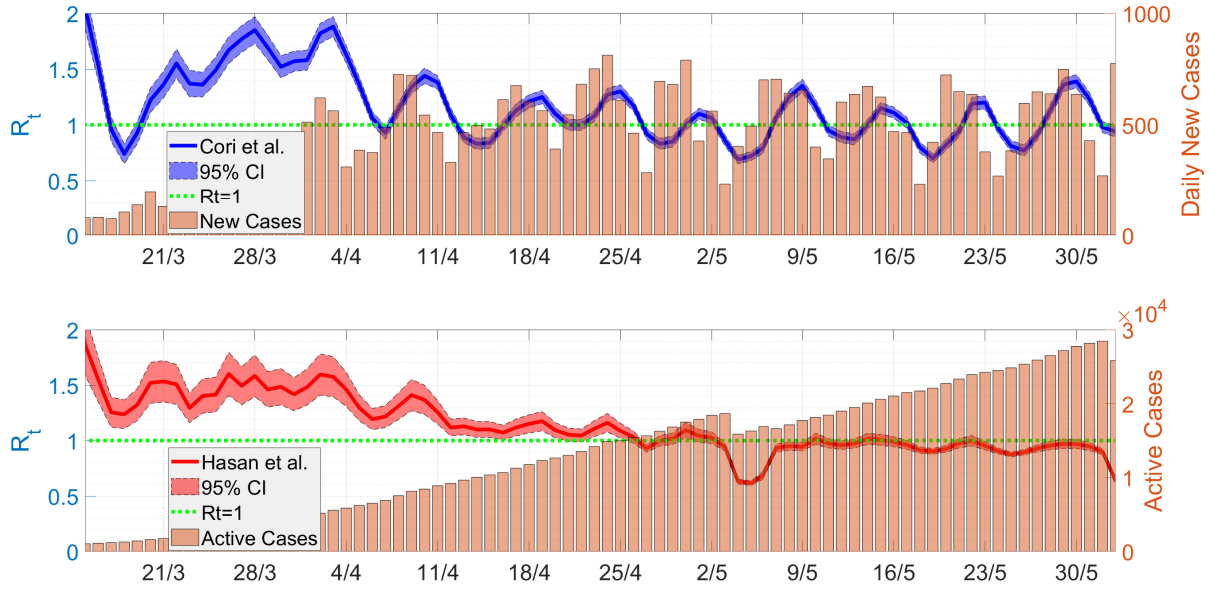
In Fig. 3 we obtain that the two methods give the plot of $\mathscr{R}_t$ with the same trend, indicating that our method is also comparable with[11]. There is a delay of about four days in the trend, especially with the time when the reproduction number curve crossed the horizontal axis. The delay is caused by the peaks of new daily cases and active ones that also differ by about the same days.

A different trend especially at later times between the methods appears in Fig. 3(c) for Norway. The curve from our method is quite smooth, while it is rather fluctuating in that using Cori et al.[11]. The discrepancy is caused by the active and recovered cases that apparently were not updated regularly, in contrast to the new positive cases needed by the method of[11]. The unreported recovery cases were all released at once on May 22, 2020, see Fig. 1(c).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



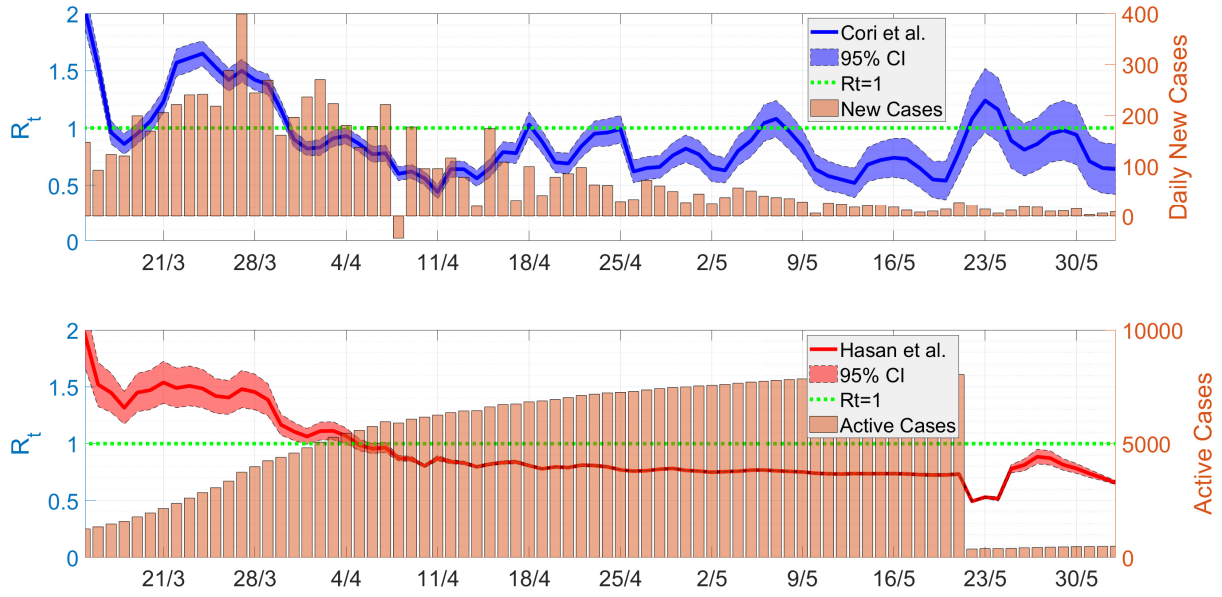
(a)



(b)

Figure 3. (Continued)

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



(c)

Figure 3. Estimated time-varying effective reproduction number of (a) Denmark, (b) Sweden, and (c) Norway by our method and[11].

Conclusion and Future Work

Many mathematical models have been developed to estimate several types of reproduction numbers during epidemic outbreaks. We here provide a novel method exploiting reported active, recovered and death cases using the SIR model as an underlying approach. This new method offers several advantages, from the modeling point of view, the resulting $\mathscr{R}_t$ value can follow the dynamics of the model suggested, so it is possible to develop it further if the model chosen has a higher complexity, besides that the estimation approach used can still be expanded in terms of statistical view. In the case that the data provided in time series do not change much or instead have drastic changes, such as accumulating at a certain time, the resulting $\mathscr{R}_t$ value will show the same spikes and serrations. As a result, the latest information from data dynamics can be more elaborated.

By applying the method to COVID-19 cases in the Scandinavian countries and comparing the results to commonly used methods due to[10] and[11], we showed that our model is comparable, which expectedly will allow for fast assessment of the reproduction number in new outbreaks. Using the method to forecast and critically assess incidence data in countries with high under-reporting, such as Indonesia, is addressed for future work.

Data and Code Availability

Data and codes can be found here: `https://github.com/agusisma/covid19`

References

1. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team. An outbreak of NCIP (2019-nCoV) infection in China - Wuhan, Hubei province. *China CDC Wkly.* **2**, 79–80, DOI: 10.46234/ccdcw2020.022 (2020).

2. Liu, Y., Gayle, A., Wilder-Smith, A. & Rocklov, J. The reproductive number of COVID-19 is higher compared to SARS Coronavirus. *J. Travel. Medicine* **27**, 1–4, DOI: 10.1093/jtm/taaa021 (2020).

3. Delamater, P., Street, E., Leslie, T., Yang, Y. & Jacobsen, K. Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.* **25**, 1–4, DOI: 10.3201/eid2501.171901 (2019).

4. Sanche, S. *et al.* High contagiousness and rapid spread of severe acute respiratory syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1–8, DOI: 10.3201/eid2607.200282 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.28.20142158; this version posted January 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One* **2**, 1–12, DOI: 10.1371/journal.pone.0000758 (2007).
6. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516, DOI: 10.1093/aje/kwh255 (2004).
7. Cauchemez, S. *et al.* Real-time estimates in early detection of SARS. *Emerg. Infect. Dis.* **12**, 1–4, DOI: 10.3201/eid1201.050593 (2006).
8. White, L. & Pagano, M. Transmissibility of the Influenza virus in the 1918 pandemic. *PLoS One* **3**, 1–6, DOI: 10.1371/journal.pone.0001498 (2008).
9. Nishiura, H., Chowell, G., Safan, M. & Castillo-Chavez, C. Pros and cons of estimating the reproduction number from early epidemic growth rate of Influenza a (H1N1) 2009. *Theor. Biol. Med. Model.* **7**, 1–13, DOI: 10.1186/1742-4682-7-1 (2010).
10. Bettencourt, L. & Ribeiro, R. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* **3**, 1–9, DOI: 10.1371/journal.pone.0002185 (2008).
11. Cori, A., Ferguson, N., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512, DOI: 10.1093/aje/kwt133 (2013).
12. Kai-Wang To, K. *et al.* Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infect. Dis.* **20**, 565–574, DOI: 10.1016/S1473-3099(20)30196-1 (2020).
13. Masreliez, C. & Martin, R. Robust bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Transactions on Autom. Control.* **22**, 361–371, DOI: 10.1109/TAC.1977.1101538 (1977).
14. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (Wiley-Interscience, 2006), 1 edn.
15. Chowell, G., Hyman, J., Bettencourt, L. & Castillo-Chavez, C. *Mathematical and Statistical Estimation Approaches in Epidemiology* (Springer, 2009), 1 edn.
16. Nishiura, H., Linton, N. & Akhmetzhanov, A. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **93**, 284–286, DOI: 10.1016/j.ijid.2020.02.060 (2020).

Acknowledgements

This work was supported by National Research and Innovation Agency (BRIN), project number 133/FI/P-KCOVID-19 2B3/IX/2020, 835/IT1 B07/KS.00.00/2020. RK gratefully acknowledgement financial support from Riset Peningkatan Kapasitas ITB 2020.

Author contributions statement

A.H. proposed the method and written the codes, H.S., V.T., R.K., E.P., P.H., and N.N. analysed the data, compared the method, and discussed the results. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.