Estimating the time-varying reproduction number of COVID-19 with a state-space method

Shinsuke Koyama,[1,*] Taiki Horie,[2,†] and Shigeru Shinomoto[2,3,‡]

[1] *The Institute of Statistical Mathematics, Tokyo 190-8562, Japan*
[2] *Department of Physics, Kyoto University, Kyoto 606-8502, Japan*
[3] *Brain Information Communication Research Laboratory Group, ATR Institute International, Kyoto 619-0288, Japan*

After slowing down the spread of the novel coronavirus COVID-19, many countries have started to relax their severe confinement measures in the face of critical damage to socioeconomic structures. At this point, it is desirable to monitor the degree to which political measures or social affairs have exerted influence on the spread of disease; however, tracing back individual transmission of infections whose incubation periods are long and highly variable seems to be difficult. Nevertheless, it may be possible to estimate the changes that may have occurred in the past, if we can suitably fit a proper model to daily event-occurrences. We have devised a state-space method for fitting the Hawkes process to a given dataset of daily confirmed cases. This method detects changes occurring in the spread of the contagion in each country. Furthermore, this method can assess the impact of social events in terms of the temporally varying reproduction number representing the average number of cases directly caused by a single infected case. This information might serve as a reference for the behavioral guidelines that should be adopted according to the varying risk of infection.

I. INTRODUCTION

In response to the steep increase in infections and deaths due to the novel coronavirus COVID-19, almost every country has taken severe confinement measures, such as lockdowns and border closures. The interventions that enforce social distancing have succeeded in slowing down the contagion, but they have also caused severe damage to socioeconomic structures, including an economic crisis, mental health deterioration, and a lag in education. There is no universal principle of adopting a specific intervention, and accordingly, there was a wide variety of initial intervention strategies, which was later followed by liberalization and relaxed behaviors of individuals, across different countries.

Nevertheless, it is desirable to capture the degree to which individual political interventions or social affairs in each country have exerted influence on the spread of the disease. A fundamental metric representing the degree of the spread is the reproduction number, R, which is defined as the average number of cases directly caused by a single case [1, 2]. One can easily infer the current status solely by glancing at the daily new cases; $R > 1$ if the daily cases are expanding on average and $R \leq 1$ otherwise. However, it is difficult to trace concrete processes by which infections are transmitted among individual persons, particularly in light of protecting private information. Thus, a statistical analytical method is needed to infer the underlying process from the available data, considering the possibility that they were obtained with partial observation and are accompanied by errors.

Mathematical epidemiological studies using ordinary differential equation models, such as the susceptible-infectious-recovered (SIR) model, have contributed to our understanding of causal factor dynamics, and these were further used in suggesting control measures needed in given situations [3–5]. While the original study of Kermack and McKendrick in 1927 [6] considered integro-differential equations taking into account the delay in the transmission of disease, the majority of later studies used ordinary differential equations in favor of analytical treatment. COVID-19 cases were also discussed using the susceptible-exposed-infectious-recovered (SEIR) model [7]. Though the ordinary differential equation models also assume the transmission delay, they cannot take account of the detailed distribution of delays, which are widely dispersed from 2 to 14 days [8–11]. To evaluate how individual political measures have exerted influence on the transmission of disease, it is necessary to incorporate the delay distribution explicitly in the analysis, as has been done previously in an analysis using the semi-mechanistic Bayesian hierarchical model [12].

Here, we construct a state-space method for estimating R that is changing in time, by fitting the Hawkes process [13], which adopts the delay distribution explicitly. While the semi-mechanistic Bayesian hierarchical model [12] requires the change-points be assigned manually, our method automatically detects the change-points solely from a given series of the number of daily cases. We first apply the method to synthetic data obtained by the simulation to confirm that the method properly detects the change-points in the reproduction number. We then apply our method to the real data and examine if the detected changes were consistent with the times at which political measures had been taken in each country. Importantly, our method assesses the impact of individual political measures accurately in terms of the time-varying reproduction number.

* skoyama@ism.ac.jp
† horie.taiki.55x@st.kyoto-u.ac.jp
‡ shinomoto.shigeru.6e@kyoto-u.ac.jp

II. METHODS

We have developed a state-space method of estimating the temporally changing reproduction number from a given series of the number of daily new infections, by adopting the Hawkes process [13] as a basic model describing the transmission of disease.

A. The rate process on a daily basis

The Hawkes process describes a self-excitation process in terms of the instantaneous occurrence rate $\lambda(t)$ as

$$\lambda(t) = \mu + R \sum_{t_k < t} \phi(t - t_k), \tag{1}$$

where μ is the spontaneous occurrence rate, and the second term represents a self-excitation effect such that the occurrence of an event adds the probability of future events (Fig. 1). R is the reproduction number representing the average number of events induced by a single event, t_k is the occurrence time of a past (kth) event, and $\phi(t)$ is a kernel representing the distribution of the transmission delays, satisfying the normalization $\int_0^\infty \phi(t)dt = 1$.

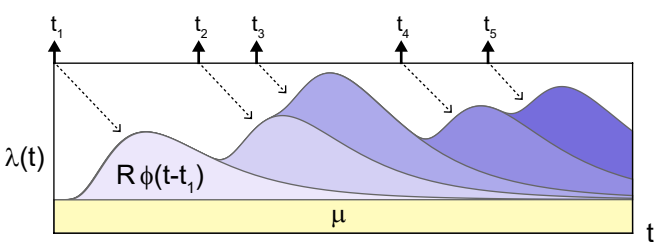

FIG. 1. Schematic description of the Hawkes process Eq. (1). The occurrence rate $\lambda(t)$ is increased according to past events occurred at times $t = t_k$ ($k = 1, 2, 3, \cdots$) with the transmission delays $t - t_k$ distributed with $\phi(t - t_k)$. R is the reproduction number that represents the average number of events induced by a single event.

We convert the original instantaneous rate Eq. (1) into the expected number of events on a daily basis:

$$\lambda_j = \mu' + \sum_{i=1}^{j-1} \nu_i R_i \phi_{j-i}, \tag{2}$$

where λ_j is the expected number of events on jth day. The first term μ' on the right-hand-side referes to the expected number of spontaneous occurrences on a daily basis. The second term represents the self-excitation process, the manner in which ν_i events that have occurred on a day i exerted influence with the delay of $j-i$ days. Here, we assume that the reproduction number R may change and represent the daily dependence as R_i. ϕ_{j-i} represents a distribution of the transmission delays $d = j - i$, satisfying the normalization $\sum_{d=1}^{\infty} \phi_d = 1$.

Given the mean rate λ_i or λ, the number of events ν_i or ν would be derived from the Poisson distribution $p(\nu|\lambda) = \frac{\lambda^\nu}{\nu!} e^{-\lambda}$. However, real data are subject to erroneous observation and accordingly they tend to be over-dispersed, or the sample variance exceeds the sample mean. Here, we incorporate over-dispersed data using the negative binomial distribution [14]:

$$p(\nu|\lambda, \rho) = \frac{\Gamma(\nu + \frac{\lambda}{\rho})}{\Gamma(\nu+1)\Gamma(\frac{\lambda}{\rho})} \left(\frac{\rho}{1+\rho}\right)^\nu \left(\frac{1}{1+\rho}\right)^{\frac{\lambda}{\rho}}, \tag{3}$$

where $\rho\,(> 0)$ represents the degree of over-dispersion, or the variance is $\mathrm{Var}(\nu) = (1+\rho)\lambda$. The Poisson distribution is in the limit of $\rho \to 0$.

The COVID-19 model parameters were chosen as follows: the spontaneous occurrence of infection is absent, $\mu' = 0$, because there is no spontaneous occurrence for COVID-19 except at the initial occurrence in China. The virus is transmitted between individuals during close contact, and each individual is determined to have an episode of infection. The transmission delay $d = j - i$ is an interval between the day i that an individual was found to be infected and the day j that another individual who caught infection was found. The duration between symptom onsets of successive cases is referred to as the serial interval [15]; it is considered to be slightly different from the incubation period [16, 17]. It is reported that the distribution of the serial intervals is suitably approximated with the log-normal distribution function of the mean 4.7 days and SD 2.9 days for COVID-19 [11]. We have adopted this distribution as the transmission delay kernel ϕ_d. The distribution of transmission delays on a daily basis is given as the difference of a cumulative distribution function of the log-normal distribution, $\phi_d = \Phi_d - \Phi_{d-1}$, where

$$\Phi_d = \frac{1}{2}\mathrm{erfc}\left(-\frac{\log d - \mu}{\sqrt{2\sigma^2}}\right), \tag{4}$$

where the parameters μ and σ are given in terms of the mean $m = 4.7$ and the SD $s = 2.9$ as $\mu = \log(m^2/\sqrt{s^2+m^2})$ and $\sigma = \sqrt{\log(1+s^2/m^2)}$.

B. The system equation

To detect change-points in the reproduction number R_i, we introduce a method of estimating stepwise dynamics [18]. We assume that system's state x_i obeys the evolution

$$x_i = x_{i-1} + \xi_i, \tag{5}$$

with the Cauchy random number ξ:

$$p(\xi|\gamma) = \frac{\gamma}{\pi(\xi^2 + \gamma^2)}. \tag{6}$$

We assume that the reproduction number is given as $R_i = f(x_i)$ with the non-negative function. Here we adopted a ramp function $f(x) = \max(0, x)$.

C. State inference

We constructed a state-space model for estimating the temporally changing reproduction number R_i from a given dataset of daily confirmed cases $\{\nu_1, \ldots, \nu_T\}$. The basic procedure of constructing the state-space method is similar to the one we developed for estimating exogenous and endogenous factors in a chain of point events [19].

To put the model in the state-space form, we take the summation in Eq. (2) over the last L days,

$$\lambda_j = \mu' + \sum_{i=j-L}^{j-1} \nu_i R_i \phi_{j-i}, \tag{7}$$

and introduce a concatenated state vector,

$$X_i := (x_{i-1}, \ldots, x_{i-L})^T, \tag{8}$$

so that the rate process (7) depends only on the current state X_i. Accordingly, the state X_i obeys the evolution

$$X_i = FX_{i-1} + G\xi_{i-1}, \tag{9}$$

where

$$F = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{10}$$

We have chosen $L = 30$ in the following analysis.

The posterior distribution of system's state X_i, given a set of daily new cases until ith day $Y_i := \{\nu_1, \ldots, \nu_i\}$ is obtained using Bayes' theorem as

$$p(X_i|Y_i) = \frac{p(\nu_i|X_i)p(X_i|Y_{i-1})}{p(\nu_i|Y_{i-1})}. \tag{11}$$

Here, $p(X_i|Y_{i-1})$ may be obtained using a system model $p(X_i|X_{i-1})$ and the posterior distribution on day $i-1$, $p(X_{i-1}|Y_{i-1})$, as

$$p(X_i|Y_{i-1}) = \int p(X_i|X_{i-1})p(X_{i-1}|Y_{i-1})dX_{i-1}. \tag{12}$$

Starting from the initial distribution $p(X_1|Y_0)$, we iterate Eqs. (11) and (12) to compute $p(X_i|Y_{i-1})$ and $p(X_i|Y_i)$ for $i = 1, 2, \ldots, T$.

Then, we compute the distribution of system's states $\{X_1, \ldots, X_T\}$, given an entire set of occurrences $Y_T := \{\nu_1, \ldots, \nu_T\}$ with

$$p(X_i|Y_T) = p(X_i|Y_i) \int \frac{p(X_{i+1}|Y_T)p(X_{i+1}|X_i)}{p(X_{i+1}|Y_i)} dX_{i+1} \tag{13}$$

in reverse order as $i = T-1, T-2, \ldots, 1$, using the distribution functions $p(X_i|Y_i)$ and $p(X_i|Y_{i-1})$, which were obtained with Eqs. (11) and (12).

We then take the median of the posterior distribution $p(X_{i+1}|Y_T)$ for the estimate of the state $\hat{X}_{i+1}$. The estimate of the reproduction number, $\hat{R}_i$, is then given by the first element of $f(\hat{X}_{i+1})$. With the estimated reproduction number, we obtain the estimated total rate as

$$\hat{\lambda}_j = \mu' + \sum_{i=1}^{j-1} \nu_i \hat{R}_i \phi_{j-i}. \tag{14}$$

We devised an algorithm that performs the integrations in Eqs. (11), (12), and (13) numerically using a sequential Monte Carlo method [20, 21].

To avoid bias in estimating the state, which is caused by outliers in the data, we may discard the preassigned outliers and treat them as "missing observations" [21], for which the posterior distribution of X_i, conditional on Y_i, is set to $p(X_i|Y_i) = p(X_i|Y_{i-1})$ without applying the Bayesian update (11).

D. Variation by day of the week

A large variation has been observed by day of the week in the number of reported infections, as reported infections tend to be fewer on the weekend than on the weekdays. There might have been variations in the original infectious activity due to human behavior, but it is more likely that this variation was caused by the delay in confirming infections and compiling the results at the weekend. The variation by day of the week is commonly observed, but there are large differences between countries, presumably due to the cultural difference in weekly activities (Fig. 2).

Before analyzing a sequence of daily cases of a given country, we process the data as follows; we first obtained the gross daily variation β_i in a week by averaging over the entire infection record (from March 1 to the present), so that the average over a week is normalized as

$$\frac{1}{7} \sum_{\text{Sunday}}^{\text{Saturday}} \beta_i = 1. \tag{15}$$

Then, we convert the original data of daily infections $\{n_1, \ldots, n_T\}$ to an adjusted dataset $\{\nu_1, \ldots, \nu_T\}$ by

$$\nu_i = n_i/\beta_i, \tag{16}$$

to which we apply the state-space method.

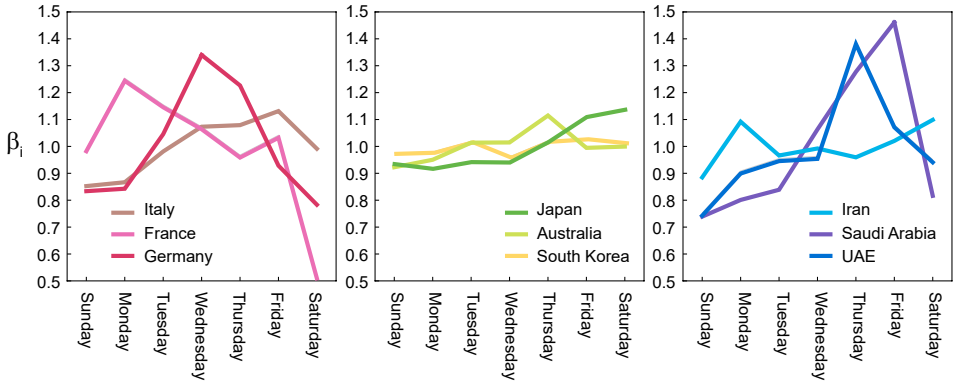

FIG. 2. Variations by day of the week in the number of reported infections $\{\beta_i\}$ computed for several countries.

III. RESULTS

We first evaluated the functionality of the state-space method by applying it to synthetic data, which was generated by numerically simulating the self-exciting process. After confirming that the estimation method was able to detect the change-points and estimate the amplitude of the reproduction number properly, we applied the method to real data of daily confirmed cases in several countries; then we further examined whether the detected changes were consistent with social events that occurred in each country.

A. Analysis of synthetic data

We constructed simulations of the Hawkes process mimicking prototypical evolutions in several countries. We took $\mu' = 0$, as there has been no spontaneous occurrence of COVID-19 except at the initial occurrence in China. We began with a few infections as initial seeds, mimicking those who introduced the disease into each country. With an initial reproduction number $R > 1$, the daily cases initially grew exponentially. To reproduce a variety of evolutions in the number of infected cases in different countries, we have evaluated several schedules of the reproduction number R_i.

Figure 3 depicts three prototypical cases: the rapid increase is followed by a slow decrease, as in Italy and France (type A); the rapid increase is followed by a rapid decrease, and then it started to increase slowly, as in Japan, and Australia (type B); and the increase is followed by a decrease, and then another large increase, as in Iran and Saudi Arabia (type C). For each type of time-varying reproduction number R_i, the Hawkes process was simulated over an interval of length $T = 120$ days, in order to generate daily cases $\{\nu_1, \ldots, \nu_T\}$ (Fig. 3). The parameter of the negative binomial distribution (3) was set to $\rho = 50$.

We were able to estimate the reproduction number $\hat{R}_i$ by applying the state-space method to a series of the number of daily new cases. We then tested different values for the hyperparameter γ as 10^{-2}, 10^{-3}, and 10^{-4}, and confirmed that the free energy was lowest for the case of $\gamma = 10^{-3}$. Accordingly, we fixed the hyperparameter as $\gamma = 10^{-3}$ throughout the following analysis.

To estimate R_i from the synthetic data using the state-space method, we first estimated ρ using the equation

$$\hat{\rho} = \frac{1}{T} \sum_{i=1}^{T} \frac{(\nu_i - \bar{\lambda}_i)^2}{\bar{\lambda}_i} - 1, \tag{17}$$

where $\bar{\lambda}_i = \sum_{j=-3}^{3} \nu_{i+j}/7$ represents the mean daily cases averaged over a week. The parameter of the Cauchy distribution (6) was set to $\gamma = 10^{-3}$. With these parameters, we performed the sequential Monte Carlo algorithm with 10^6 particles to compute the posterior distributions of the reproduction number for each day. Figure 3 depicts the median (solid line) and 95% range (shaded areas) of the posterior distributions. We see that the amplitude of the reproduction number is estimated properly. In particular, $\hat{R}_i$ was able to detect change-points.

B. Analysis of real data

1. *Daily new cases*

The number of daily new cases in various countries is made available on websites hosted by public research centers such as Our World in Data (https://ourworldindata.org/coronavirus-source-data) and the Humanitarian Data Exchange (https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases). We used data from the former site in this analysis. For a sequence of daily cases $\{n_1, \ldots, n_T\}$ in each country, we obtained the gross daily variation β_t in a week and adjusted the real data using Eq. (16) to $\{\nu_1, \ldots, \nu_T\}$. The parameter of the negative binomial

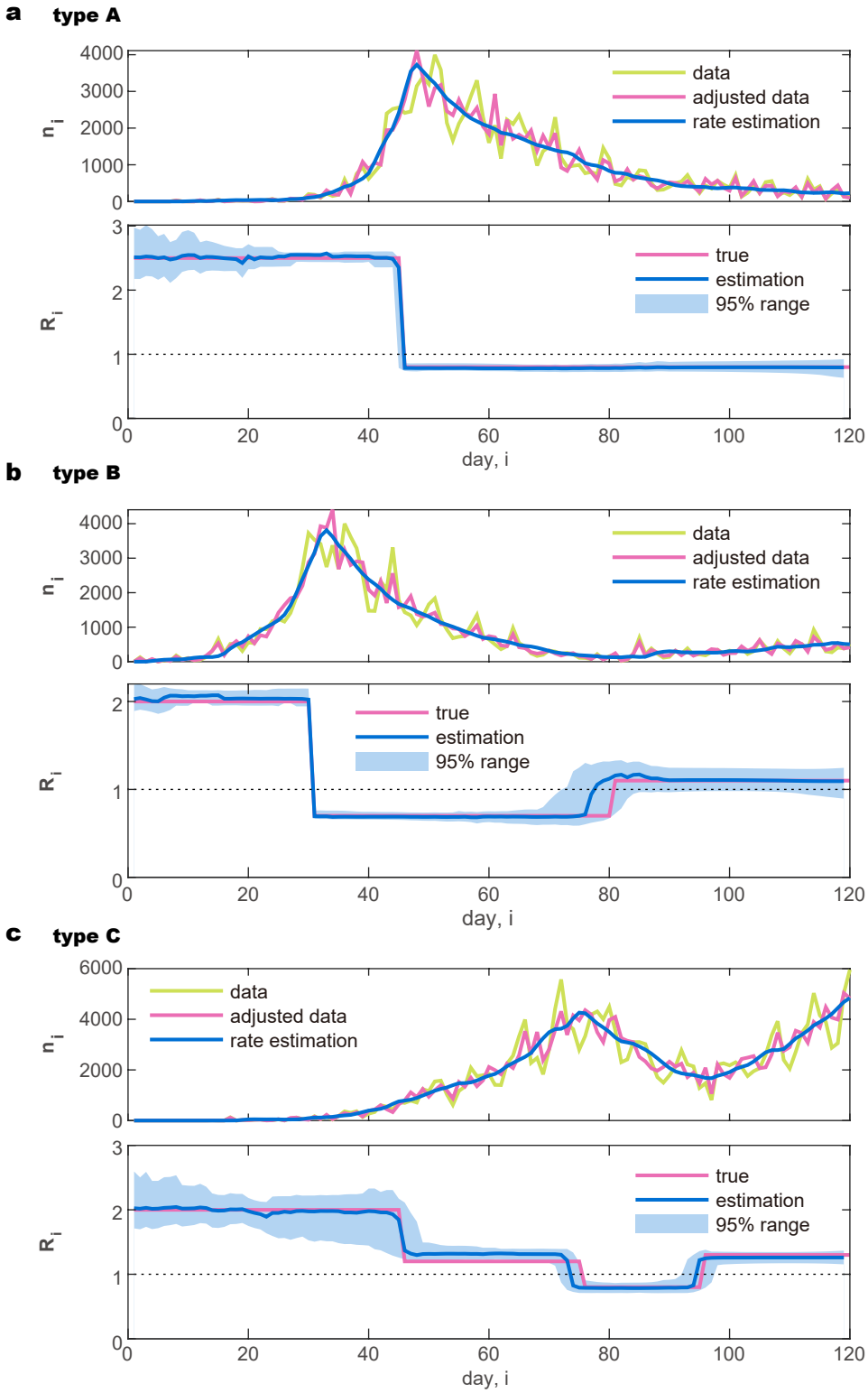

FIG. 3. Synthetic daily cases obtained by simulating the Hawkes process. The reproduction number $\hat{R}_i$ estimated with the state-space method is plotted below in reference to the true reproduction number R_i (blue line). The purple solid line and the shaded area represent the median and 95 % range of the posterior distribution, respectively. (a) Type A: rapid increase followed by a slow decrease. (b) Type B: increase followed by a rapid decrease, and then slow increase. (c) Type C: slow increase followed by a decrease, and then another large increase.

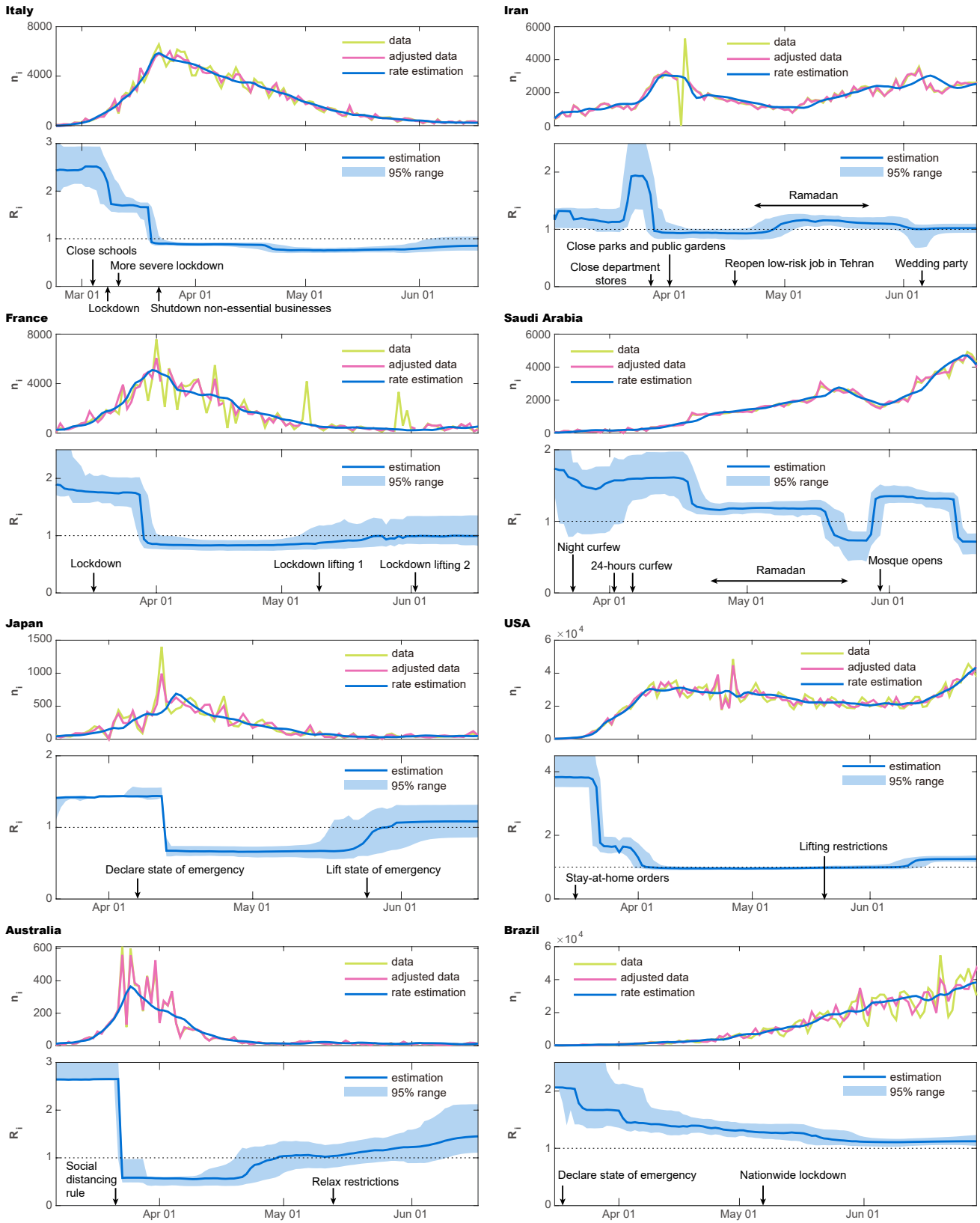

FIG. 4. Number of daily new cases and the reproduction number $\hat{R}_i$ estimated using the state-space method. (left) Italy, France, Japan, and Australia; (right) Iran, Saudi Arabia, and the USA, and Brazil.

distribution ρ was estimated using Eq. (17). The parameter of the Cauchy distribution (6) was set to $\gamma = 10^{-3}$, as in the analysis of synthetic data. With these parameters, the reproduction number was estimated using the sequential Monte Carlo method with 10^6 particles.

Figure 4 depicts the daily new cases in several countries and the reproduction number $\hat{R}_i$ estimated with our state-space method:

Italy and France:
A rapid increase in new cases was followed by a slow decrease. The estimated reproduction number was $\hat{R} > 1$ at the outset and dropped to $\hat{R} < 1$. It is interesting to note that the drop in the reproduction number is temporally similar to the time at which political measures, such as lockdown and border closure, were enforced.

Japan and Australia:
The number of cases in countries of Asia and Oceania is found to be relatively small compared to those in Europe. An increase in the number of cases was followed by a rapid decrease, and then by a second slow increase. Accordingly, the reproduction number exhibited a drop from $\hat{R} > 1$ to $\hat{R} < 1$, and then it increased slightly.

Iran and Saudi Arabia:
The number of new cases in these countries alternately moved up and down, and so accordingly, the estimated reproduction number $\hat{R}$ repeatedly crossed the value unity. It can be observed that Ramadan has promoted increased reproduction number, as it may have facilitated human contact.

The United States:
A rapid increase of new cases is followed by a very slow decrease, and then another growth. The estimated reproduction number $\hat{R}$ was higher than unity at the beginning, dropped off to near unity due to the confinement measures taken, but then it exceeded unity again. The political measures taken were found to vary by state, making it difficult to interpret the data from this country as a whole in our model.

Brazil:
The number of new cases keeps growing. The estimated reproduction number $\hat{R}$ has remained greater than unity.

2. *The reproduction numbers at the initial phase*

There have been debates about why infection rate and mortality rate change by orders of magnitude across different countries. Though these numbers likely reflect the confinement measures taken in individual countries, there might also have been differences across nations in susceptibility to COVID-19, reflecting not only genetic resistance but also lifestyle and cultural differences, such as shaking hands or hugging.

Because most governments did not implement serious confinement measures at the initial phase, the initial exponential increase of infections might reflect the natural susceptibility of citizens of each country. We realized that the estimated reproduction number was stable in a certain period before each country took confinement measures such as a lockdown or social distancing. Figure 5a depicts the reproduction numbers estimated with our state-space method for 10 days until the day before the confinement measures of each country. The initial variation in the numbers of daily new cases is depicted in Fig. 5b, indicating that the estimated reproduction number is correlated to the slope in the log plot. Here we have selected the period shifted by 5 days by taking account of the typical transmission delays. We can observe that countries in different regions tend to cluster, indicating that the susceptibility tended to be similar between nations in the same region.

IV. DISCUSSION

Society and the media currently alternate between hope and despair in response to the temporary decrease or increase of daily new COVID-19 infections, which came out after a long latency period. To make an objective assessment of the current status, we have developed a state-space method for estimating the control status, in particular quantifying the time-dependent reproduction number R.

Future prediction

Using our method, it is also possible to predict the number of new cases in the future. This can be done by simulating the converted Hawkes process Eq. (2) with the negative binomial distribution Eq. (3), whose parameter ρ is chosen at the value estimated from the data using Eq. (17). One may adopt the reproduction number R_i as constant at the value of an endpoint of estimation if the current conditions are assumed to be maintained. Alternatively, one may also examine various time schedules of R_i, by assuming possible relaxation or confinement.

Pros and cons

We adopted the Hawkes process, or the self-exciting model, in describing the variables underlying the transmission of disease. In contrast to ordinary differential equation models such as SIR or SEIR models, the Hawkes process is advantageous in that it explicitly specifies the distribution of transmission delays. However, the Hawkes

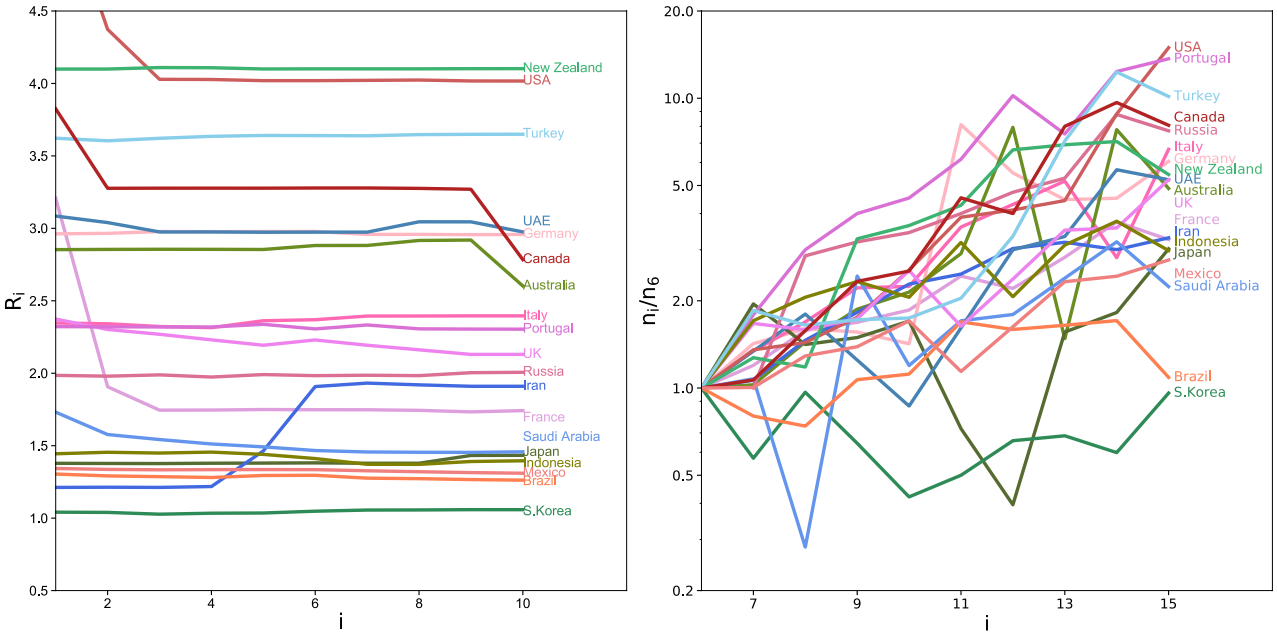

FIG. 5. The difference in the reproduction number at the initial phase. (a) The reproduction numbers estimated with our state-space method for 10 days until the day before the lockdown measures of each country. Days are counted from 12 days before the confinement measures. (b) Initial variation in the numbers of daily new cases; n_i divided by n_6. The period is shifted by 5 days, by taking account of the typical transmission delay.

process does not account for the finite size effect, in which infected and recovered people represent a finite fraction among the entire population. There have been some models that incorporate the finite population size effect into the Hawkes process, as has been done with the SIR or SEIR models [22, 23]. To analyze the current status of COVID-19, however, we do not take the finite size effect into account, as the fraction of the recovered or removed people is still less than a few % of the entire population.

While preparing this manuscript, the authors discovered another interesting approach that applied the Hawkes process for modeling COVID-19 transmission [24]. In that study, the authors combined the Hawkes process with spatial and temporal covariates, such as demographic features and Google mobility indices, to explain the variability of the reproduction number, and to forecast future cases and deaths in the USA. By contrast, our state-space method estimates the change-points in the reproduction number solely from a series of daily cases. Though these methods apply the Hawkes process differently, both methods are seen to demonstrate the effectiveness of this model in analyzing epidemics.

We introduced the Cauchy distribution, Eq. (6), into our analysis, assuming the stepwise changes in the reproduction number R_i. Accordingly, we were able to infer change-points from the posterior distribution taking on stepwise characteristics. It should be noted that, however, there may be an additional latency between the times at which political measures were conducted and the changes in the reproduction number, which may reflect the change in behavior. This delay may also be country-specific. Therefore, it could be interesting to investigate the delay in the change-points in the reproduction number following social events.

When inferring the transmission of disease from daily confirmed cases, we have considered potentially erroneous observations made in the real data. We took into account counting errors by assuming a negative binomial distribution that represents the over-dispersion. We also took into account the variation by day of the week and adjusted the data by compensating for the periodic dependency. Note that there may still be an underestimation of infection numbers, as asymptomatic cases may have been overlooked. Though this is unavoidable unless the inspection is enforced, it is reported that the infections caused by asymptomatic people are relatively small (about 6%) for COVID-19 [16].

We have assumed that the transmission delay is a serial interval defined as the duration between symptom onsets of successive cases and adopted the log-normal distribution with the mean 4.7 days and SD 2.9 days, as suggested in reference [11]. As our mathematical formulation is more general, it is possible to search for a more suitable transmission kernel ϕ_d without relying on such external knowledge, if the numbers of daily cases are accurately provided.

The most crucial assumption in the majority of mathe-

matical model studies, including this study, is the mean-field assumption, in which all individuals are assumed to interact uniformly. Though difficult to incorporate, it is desirable to consider the heterogeneity of the real-world community in analyzing the communicability of disease.

Despite these assumptions, our state-space method may be of worth in assessing the status of the disease systematically, based on reported daily confirmed cases. This method might serve as a reference for governments adopting variable regulations that should be changed according to current infection circumstances.

Code availability

The application program and example datasets are available at our site hosted publicly on GitHub, accessible via https://github.com/shigerushinomoto.

ACKNOWLEDGMENT

We thank Shin Takagi, Masaki Ogura, Ryota Kobayashi, Takaaki Aoki, and Hideaki Shimazaki for their constructive comments on this manuscript. S.S. is supported by the New Energy and Industrial Technology Development Organization (NEDO).

[1] G. N. Milligan and A. D. Barrett, *An Essential Guide* (Wiley Online Library, 2015).
[2] C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons, *et al.*, science **324**, 1557 (2009).
[3] H. W. Hethcote, SIAM review **42**, 599 (2000).
[4] F. Brauer, Infectious Disease Modelling **2**, 113 (2017).
[5] J. C. Blackwood and L. M. Childs, Letters in Biomathematics **5**, 195 (2018).
[6] W. O. Kermack and A. G. McKendrick, Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**, 700 (1927).
[7] J. T. Wu, K. Leung, and G. M. Leung, The Lancet **395**, 689 (2020).
[8] S.-m. Jung, A. R. Akhmetzhanov, K. Hayashi, N. M. Linton, Y. Yang, B. Yuan, T. Kobayashi, R. Kinoshita, and H. Nishiura, Journal of clinical medicine **9**, 523 (2020).
[9] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, Journal of clinical medicine **9**, 538 (2020).
[10] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T. T. Lam, J. T. Wu, G. F. Gao, B. J. Cowling, B. Yang, G. M. Leung, and Z. Feng, N. Engl. J. Med. **382**, 1199 (2020).
[11] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, medRxiv 2020.02.03.20019497 [preprint] (2020).
[12] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, H. Coupland, T. A. Mellan, H. Zhu, T. Berah, J. W. Eaton, P. N. Guzman, *et al.*, arXiv preprint arXiv:2004.11342 (2020).
[13] A. G. Hawkes, Biometrika **58**, 83 (1971).
[14] S. Koyama and Y. Fujiwara, Journal of Statistical Mechanics: Theory and Experiment **2019**, 023402 (2019).
[15] P. E. M. Fine, Am J Epidemiol **158**, 1039 (2003).
[16] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, Science **368** (2020), 10.1126/science.abb6936, https://science.sciencemag.org/content/368/6491/eabb6936.full.pdf.
[17] X. He, E. H. Y. Lau, and P. W. et al., Nat Med **26**, 672 (2020).
[18] G. Kitagawa, Journal of the American Statistical Association **82**, 1032 (1987).
[19] S. Koyama and S. Shinomoto, arXiv preprint arXiv:2003.00659 (2020).
[20] N. Kantas, A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, Statistical Science **30**, 328 (2015).
[21] G. Kitagawa, *Introduction to Time Series Modeling* (Chapman and Hall/CRC, 2010).
[22] T. Onaga and S. Shinomoto, Scientific reports **6**, 33321 (2016).
[23] M.-A. Rizoiu, S. Mishra, Q. Kong, M. Carman, and L. Xie, in *Proceedings of the 2018 World Wide Web Conference* (2018) pp. 419–428.
[24] W. H. Chiang, X. Liu, and G. Mohler, medRxiv 2020.06.06.20124149 [preprint] (2020).