

## INFERRING KEY EPIDEMIOLOGICAL PARAMETERS AND TRANSMISSION DYNAMICS OF COVID-19 BASED ON A MODIFIED SEIR MODEL

XIAOYAN WANG<sup>1</sup>, TIANJIAO TANG<sup>1</sup>, LANG CAO<sup>2</sup>, KAZUYUKI AIHARA<sup>3</sup>  
AND QIAN GUO<sup>1,\*</sup>

**Abstract.** This study aims to establish a model-based framework for inferring key transmission characteristics of the newly emerging outbreak of the coronavirus disease 2019 (COVID-19), especially the epidemic dynamics under quarantine conditions. Inspired by the shifting therapeutic levels and capacity at different stages of the COVID-19 pandemic, we propose a modified SEIR model with a two-phase removal rate of quarantined hosts undergoing continuously tunable transition. We employ the Markov Chain Monte Carlo (MCMC) approach for inferring and forecasting the epidemiological dynamics from the publicly available surveillance reports. The effectiveness of a short-term prediction is illustrated by adopting the data sets from 10 demographic regions including Chinese mainland and South Korea. In the surveillance period, the average  $\mathcal{R}_0$  ranges from 1.74 to 3.28, and the median of the mean latent period does not exceed 10 days across the surveillance regions.

**Mathematics Subject Classification.** 92D30.

Received May 6, 2020. Accepted November 23, 2020.

### 1. INTRODUCTION

An outbreak of a novel coronavirus disease that causes severe acute respiratory syndrome (SARS-CoV) emerged in December 2019, and was subsequently declared a public health emergency of international concern by the World Health Organization (WHO) on January 30, 2020. Until now, the COVID-19 pandemic has led

---

*Keywords and phrases:* COVID-19, SEIR model, effective reproduction number, asymptomatic ratio, mean latent period.

<sup>1</sup> Department of Mathematics, Shanghai Normal University, Shanghai 200234, P.R. China.

<sup>2</sup> School of Information Engineering, Zhengzhou University, Zhengzhou 450052, P.R. China.

<sup>3</sup> International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan.

\* Corresponding author: [qguo@shnu.edu.cn](mailto:qguo@shnu.edu.cn)

to more than 16 million confirmed cases and 0.6 million deaths worldwide, which constitutes an unprecedented challenge to the global public health community. Under the increasingly severe situations of epidemic prevention and control in many countries over the world, the epidemiological modeling using surveillance data provides useful mathematical tools for identifying and predicting the disease outbreaks. Particularly important, among others, are the ‘‘Susceptible-Exposed-Infectious-Recovered’’ (SEIR) type models with consideration of the pre-symptomatic period of COVID-19. Wu *et al.* [15] employed the SEIR model to estimate the basic reproduction number of the outbreak in Wuhan, China. Prem *et al.* [8] considered a deterministic age-structured SEIR model with quarantine control included as a prevention strategy of contagion. Walker *et al.* [12] used an age-structured stochastic SEIR model to fit the key parameters of the observed transmission dynamics of COVID-19. Wan *et al.* [13] analyzed the spread dynamics and trend of COVID-19 in Wuhan using the data since the city’s closure to February 12, 2020. A series of papers [9, 10, 14] developed the quarantined SEIR-type models by imposing restrictions on mobility of patients, with model parameters fitted to the data reported by China CDC using the Markov Chain Monte Carlo (MCMC) method. Under the SEIR framework, Lin *et al.* [4] adopted a similar compartmental model to capture individual behavioural reaction and governmental actions against COVID-19. Li *et al.* [3] introduced a networked SEIR-type model with migration data within China, and employed Bayesian inference to infer critical epidemiological characteristics.

Recent months have witnessed the role and importance of mathematical modeling in the decision-making of governments in response to the current pandemic. However, detecting and forecasting the spread of infectious diseases at early stages remain a challenging task due to the lack of model selection criteria or the high computational cost for parameter tuning. Here, we aim at providing a modified SEIR model with asymptomatic and quarantined classes for adaption to the newly emerged COVID-19, and propose a MCMC based method for inferring the key parameters and related transmission characteristics of the COVID-19 pandemic.

## 2. METHODS

The data used to fit parameters in this paper are information for confirmed and removed cases in Chinese mainland (23 January to 12 February 2020), France (7 March to 12 April 2020), Germany (7 March to 1 April 2020), Iran (7 March to 31 March 2020), Italy (7 March to 12 April 2020), Japan (20 March to 12 April 2020), South Korea (16 February to 9 March 2020), Spain (7 March to 12 April 2020), United Kingdom (7 March to 12 April 2020), and the United States (7 March to 16 June 2020). Only the Chinese mainland and South Korea datasets are used in the main section to illustrate the results, but the other results are available in Table 2.

We first illustrate the SEIR-type dynamic model under quarantined measures that is based on the previous works [5, 9, 10, 14], as shown in Figure 1. In particular, the infectious population is divided into four compartments: the unquarantined infectious individuals with symptoms ( $I$ ), the undetected asymptomatic carriers ( $A$ ), the quarantined infectious individuals with symptoms ( $I_q$ ), and the detected asymptomatic carriers ( $A_q$ ). Given the evidence showing that COVID-19 is contagious even during the incubation period of infection, we modify the previous SEIR models [5, 9, 10, 14] as follows:

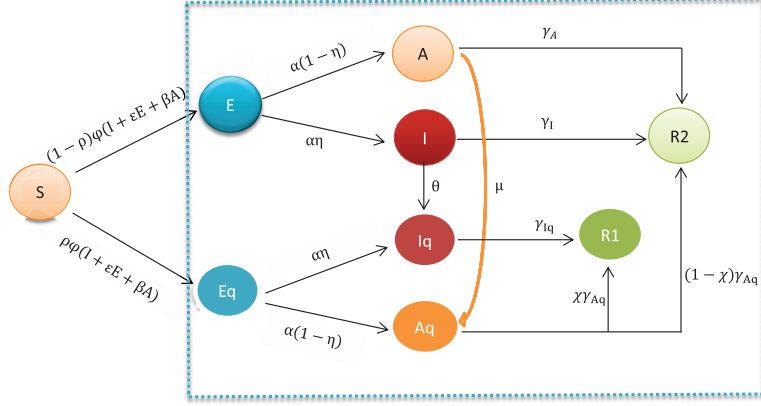


FIGURE 1. SEIR model with quarantined measures.

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -(1-\rho)\varphi(I + \varepsilon E + \beta A)S - \rho\varphi(I + \varepsilon E + \beta A)S, \\ \frac{dE}{dt} = (1-\rho)\varphi(I + \varepsilon E + \beta A)S - \alpha E, \\ \frac{dE_q}{dt} = \rho\varphi(I + \varepsilon E + \beta A)S - \alpha E_q, \\ \frac{dI}{dt} = \alpha\eta E - \theta I - \gamma_I I, \\ \frac{dI_q}{dt} = \alpha\eta E_q + \theta I - \gamma_{I_q}(t)I_q, \\ \frac{dA}{dt} = \alpha(1-\eta)E - \mu A - \gamma_A A, \\ \frac{dA_q}{dt} = \alpha(1-\eta)E_q + \mu A - \gamma_{A_q} A_q, \\ \frac{dR_1}{dt} = \gamma_{I_q}(t)I_q + \chi\gamma_{A_q} A_q, \\ \frac{dR_2}{dt} = \gamma_A A + \gamma_I I + (1-\chi)\gamma_{A_q} A_q, \end{array} \right. \quad (2.1)$$

where the notation is summarized in Table 1.

Note that we exclude, without loss of generality, the quarantined susceptible individuals ( $S_q$ ) from our SEIR model by assuming that they constitute a subpopulation isolated from the reaction mixture during the epidemic spread. Therefore we divide the total population into two parts: the quarantined susceptible population and an effective population. Here, we regard  $N_e$  as the effective population size, which can be obtained using the fraction  $p$  of the total population during the epidemic outbreak,  $N_e = pN_{\text{total}}$ , where  $N_{\text{total}}$  is the total population size. In our numerical simulations, the effective population size  $N_e$  appears as a model parameter to be determined.

TABLE 1. Full list of notations used in the SEIR model.

Notation	Definition	Surveillance	Value	Inferred
Variables				
$N_e$	Effective population size	No	$0 \sim 10^7$	Yes
$S$	Population size of susceptible individuals	No	$0 \sim 10^7$	Yes
$E$	Population size of exposed individuals	No	$0 \sim 10^4$	Yes
$A$	Population size of undetected asymptomatic carriers and undetected mild carries	No	$0 \sim 10^4$	Yes
$I$	Population size of obviously symptomatic infectious individuals without quarantine	No	$0 \sim 10^4$	Yes
$E_q$	Population size of quarantined exposed individuals	No	$0 \sim 10^4$	Yes
$A_q$	Population size of detected asymptomatic carriers and detected mild carries	No	$0 \sim 10^4$	Yes
$I_q$	Population size of quarantined infectious individuals with obvious symptom	Yes	—	No
$R_1$	Population size of quarantined removed individuals	Yes	—	No
$R_2$	Population size of removed individuals without quarantine	No	$0 \sim 10^4$	Yes
Parameters				
$\rho$	Fraction of quarantined susceptible individuals		$0.1 \sim 0.95$	Yes
$\varphi$	Transmission probability of infectious individuals by contact		$10^{-6} \sim 10^{-3}$	Yes
$\beta$	Propagation coefficient of asymptomatic infectious individuals relative to infectious individuals		$0.1 \sim 0.9$	Yes
$\varepsilon$	Propagation coefficient of exposed individuals relative to infectious individuals		$0.1 \sim 0.9$	Yes
$\alpha$	Rate of exposed individuals progressing to infectious state		$0.1 \sim 1$	Yes
$\eta$	Fraction of exposed individuals progressing to symptomatic infectious state		$0.2 \sim 0.95$	Yes
$\theta$	Quarantine rate of symptomatic infectious individuals		$0.1 \sim 0.95$	Yes
$\mu$	Detection rate of asymptomatic infectious individuals		$0 \sim 1$	Yes
$\gamma_I$	Removal rate of (unquarantined and obviously symptomatic) infectious individuals		$0 \sim 1$	Yes
$\gamma_A$	Removal rate of unquarantined asymptomatic carriers and undetected mild carries		$0 \sim 1$	Yes
$\gamma_{I_q}(t)$	Removal rate of quarantined infectious individuals with obvious symptom ( $= z_1 + z_2 \tanh(\frac{t-a}{b})$ )		$0 \sim 1$	Yes
$\gamma_{A_q}$	Removal rate of detected asymptomatic carriers and detected mild carries		$0 \sim 1$	Yes
$\chi$	Binary indicator whether detected asymptomatic carriers are counted as confirmed cases		$\{0, 1\}$	No

By substituting  $S = N_e - E - E_q - I - I_q - A - A_q - R_1 - R_2$ , the model (2.1) can be written as

$$\left\{ \begin{array}{l} \frac{dE}{dt} = (1 - \rho)\varphi(I + \varepsilon E + \beta A)(N_e - E - E_q - I - I_q - A - A_q - R_1 - R_2) - \alpha E, \\ \frac{dE_q}{dt} = \rho\varphi(I + \varepsilon E + \beta A)(N_e - E - E_q - I - I_q - A - A_q - R_1 - R_2) - \alpha E_q, \\ \frac{dI}{dt} = \alpha\eta E - \theta I - \gamma_I I, \\ \frac{dI_q}{dt} = \alpha\eta E_q + \theta I - \gamma_{I_q}(t)I_q, \\ \frac{dA}{dt} = \alpha(1 - \eta)E - \mu A - \gamma_A A, \\ \frac{dA_q}{dt} = \alpha(1 - \eta)E_q + \mu A - \gamma_{A_q} A_q, \\ \frac{dR_1}{dt} = \gamma_{I_q}(t)I_q + \chi\gamma_{A_q} A_q, \\ \frac{dR_2}{dt} = \gamma_A A + \gamma_I I + (1 - \chi)\gamma_{A_q} A_q. \end{array} \right. \quad (2.2)$$

Specifically, we adopted a two-phase removal rate  $\gamma_{I_q}(t)$  for the quarantined symptomatic population as follows:

$$\gamma_{I_q}(t) = z_1 + z_2 \tanh\left(\frac{t - a}{b}\right), \quad (2.3)$$

where  $z_1$  and  $z_2$  are parameters controlling both low and high levels of the removal rate, and  $a$  and  $b$  are parameters controlling the position and steepness of the transition between the low and high rates. An intuition behind this choice is as follows. The cure rate could be generally low at the early stage due to the novelty of the COVID-19 disease, gradually improved after a buffering period, and eventually saturated as the pandemic crisis unfolds. Our numerical results show that this heuristic removal rate significantly improves the reproduction of observed surveillance data and enhances the prediction of the COVID-19 outbreaks. We note that the prediction accuracy does not sensibly depend on the specific form of  $\gamma_{I_q}(t)$ , and other conceivable protocols (e.g., sigmoid function) also generate similar results.

To infer the key transmission characteristics of COVID-19 including the effective reproduction number, the asymptomatic ratio, the mean latent period, and the inflection points, we fit the parameters of model (2.2) using the surveillance data collected from China CDC (<http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm>) and Johns Hopkins University (<https://github.com/CSSEGISandData/COVID-19>). Specifically, denote by  $\tilde{I}_q(t)$  and  $\tilde{R}_1(t)$  the number of confirmed and removed cases of COVID-19 on the day  $t$  in the surveillance period, and the least squares loss function for optimization of our SEIR model is then written as follows:

$$C(\tilde{I}_q, \tilde{R}_1; \Theta) = \sum_{t \in T} |\tilde{I}_q(t) - \hat{I}_q(t; \Theta) - \chi \hat{A}_q(t; \Theta)|^2 + |\tilde{R}_1(t) - \hat{R}_1(t; \Theta)|^2, \quad (2.4)$$

where  $T$  is the surveillance period available for inference of our model parameters,  $\hat{I}_q(t; \Theta)$ ,  $\hat{A}_q(t; \Theta)$  and  $\hat{R}_1(t; \Theta)$  denotes the corresponding numerical solution to equation (2.2) during the surveillance period using the fourth-order Runge-Kutta method with the epidemic parameters and initial condition collectively denoted by  $\Theta = [\alpha, \beta, \varepsilon, \eta, \gamma_I, \gamma_A, z_1, z_2, a, b, \mu, \rho, \theta, \varphi, \gamma_{A_q}, N_e, E_0, A_0, I_0, E_{q0}, A_{q0}, R_{20}]$ . See Table 1 for more detailed

description. To solve the high-dimensional optimization problem in an effective manner, we perform the MCMC sampling [9, 10] of the parameter space using the MATLAB toolbox MCMCSTAT [2], as shown in Algorithm 1. In our optimization procedure, we randomly generate 20 sets of initial values in a reasonable range around the parameters, and the proposed algorithm selects the best initial value leading to the least sum of squared deviations from the surveillance data. We further numerically examine the initial-value sensitivity (IVS) of our model predictions, finding that a considerable fraction of parameters share almost the same optimal initial values (which we called “generic initial estimation”, see Table 4) across different demographic regions, possibly and partially due to the common epidemiological properties of the COVID-19 dynamics. This robust IVS of our model significantly alleviates the difficulty of the inference problem and renders the estimation process of (non-generic) parameters computationally tractable. It is noteworthy that our results derived from local optima of a loss function subject to specific initial values of the (non-generic) parameters only provide suboptimal, rather than an optimal, predictions, but numerical experiments demonstrate the potential of our method in forecasting of the COVID-19 outbreaks in practice.

---

**Algorithm 1:** Inference algorithm for the modified SEIR model (see Appendix B for details)

---

**Input:** Real surveillance data  $\tilde{I}_q$  and  $\tilde{R}_1$ , and initial values of undetermined parameters (see Table 4);

- 1 **foreach**  $k \in \{1, \dots, k_{max}\}$  **do**
- 2     Refine the initial guess by using sequential quadratic programming or the function ‘fmincon.m’ in MATLAB;
- 3     Fit the parameter set  $\Theta$  by employing MATLAB toolbox MCMCSTAT with the simulation number 2000;
- 4 **end**
- 5 Calculate with  $\{\hat{\Theta}^{(k)}\}$  the 95% CIs of the inflection time and maximal  $I_q$ , the effective reproduction number  $\mathcal{R}_0$ , the asymptomatic ratio, and the mean latent period, respectively.

---

### 3. RESULTS

We first apply our method to predict the inflection time with the largest single-day number of confirmed cases during the outbreaks of COVID-19 in Chinese mainland and South Korea, respectively. Considering that the asymptomatic ratio of confirmed cases is very small in the surveillance period [1], we set  $\chi = 0$  when fitting with data reported by China CDC. But in other regions, we set  $\chi = 1$ . Figures 2 and 3 plot the simulated populations of quarantined infectious and quarantined removed individuals using the inferred parameters from the surveillance data (also see Tab. 2). Here we set  $k_{max} = 1000$ .

With the estimated parameters  $\hat{\Theta}$  and simulated time course of epidemics, one can readily obtain several key transmission characteristics of COVID-19, such as the inflection time  $\tau_{\text{infl}} = \arg \max_t \{I_q(t; \hat{\Theta}) + \chi A_q(t; \hat{\Theta})\}$ , the mean latent period  $\bar{\tau}_{\text{inc}} = 1/\hat{\alpha}$ , the asymptomatic ratio  $p_A(t) = \frac{A_q(t) + A(t)}{I_q(t) + I(t) + A_q(t) + A(t)}$  and the effective reproduction number defined as follows (see Appendix for detailed derivation):

$$\mathcal{R}_0 = (1 - \hat{\rho}) \hat{\varphi} \hat{N}_e \left\{ \frac{\hat{\varepsilon}}{\hat{\alpha}} + \frac{\hat{\beta}(1 - \hat{\eta})}{\hat{\mu} + \hat{\gamma}_A} + \frac{\hat{\eta}}{\hat{\gamma}_I + \hat{\theta}} \right\}. \quad (3.1)$$

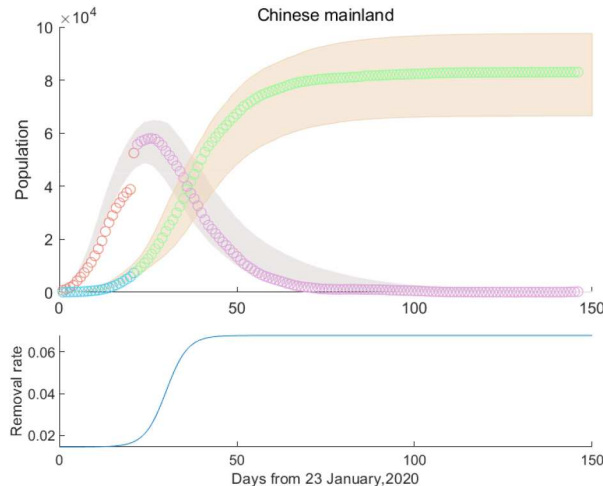


FIGURE 2. (a) Populations of the quarantined infectious and the quarantined removed individuals as functions of elapsed time since the COVID-19 outbreak in Chinese mainland. The circles represent the surveillance data by the China CDC from January 23 to June 16 with different colors indicating the reported data before and after February 12, and the colored area represents 95% CI of the model prediction using the parameters inferred by Algorithm 1. (b) Inferred removal rate  $\gamma_{I_q}(t)$  of the quarantined symptomatic population.

In the surveillance period, the average  $\mathcal{R}_0$  ranges from 1.74 to 3.28, and the mean latent period is relatively short in the Chinese mainland, Japan, and South Korea, while the average does not exceed ten days across the surveillance regions. Moreover, the real values of the inflection point and the number of existing confirmed cases at the inflection point fall within the confidence intervals. See Tables 2 for the detailed estimation results using the surveillance data across different demographic regions. To explore the proportion of cases with asymptomatic infections in the surveillance period, we create the box plots for the data of asymptomatic ratio  $p_A(t)$  at  $t = 7$ , see Figure 4. These plots show that the asymptomatic ratio of the United States exhibits a more obvious increment to the other ones, and the standard deviation of the prediction to the asymptomatic ratio of Japan is relatively high.

#### 4. DISCUSSION

In this paper, we proposed a modified SEIR model with both quarantined and asymptomatic populations as well as infection by exposed individuals for fitting the observed transmission dynamics of COVID-19, thus enabling an accurate short-term prediction of the infectious disease and providing implications for outbreak control. Several important epidemiological parameters represented by the effective reproduction numbers  $\mathcal{R}_0$  across different demographic regions were also calculated, which are close to the result obtained in [16]. A recent review [6] reported that an average estimation for  $\mathcal{R}_0$  is around 3.28, which is consistent with our findings (see Tab. 2).

Our prediction of the asymptomatic ratio of COVID-19 also agrees basically with the medical evidence [7] based estimates of the pandemic. From Table 3, we obtain the mean value of  $\hat{\beta}$ , that is, the transmission

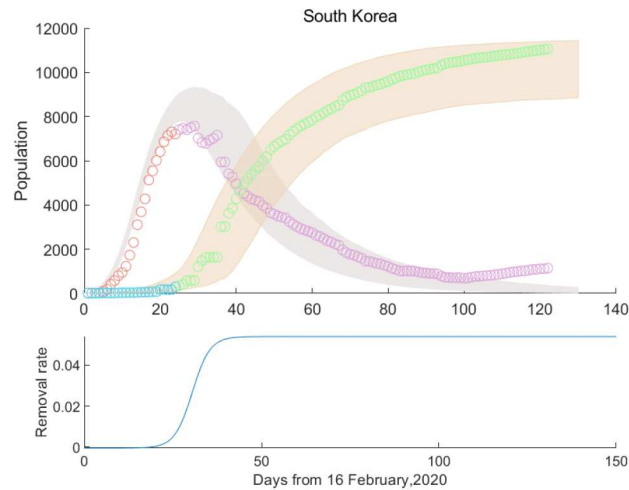


FIGURE 3. (a) Populations of the quarantined infectious and the quarantined removed individuals as functions of elapsed time since the COVID-19 outbreak in South Korea. The circles represent the surveillance data by Johns Hopkins University from February 16 to June 16 with different colors indicating the reported data before and after March 9, and the colored area represents 95% CI of the model prediction using the parameters inferred by Algorithm 1. (b) Inferred removal rate  $\gamma_{I_q}(t)$  of the quarantined symptomatic population.

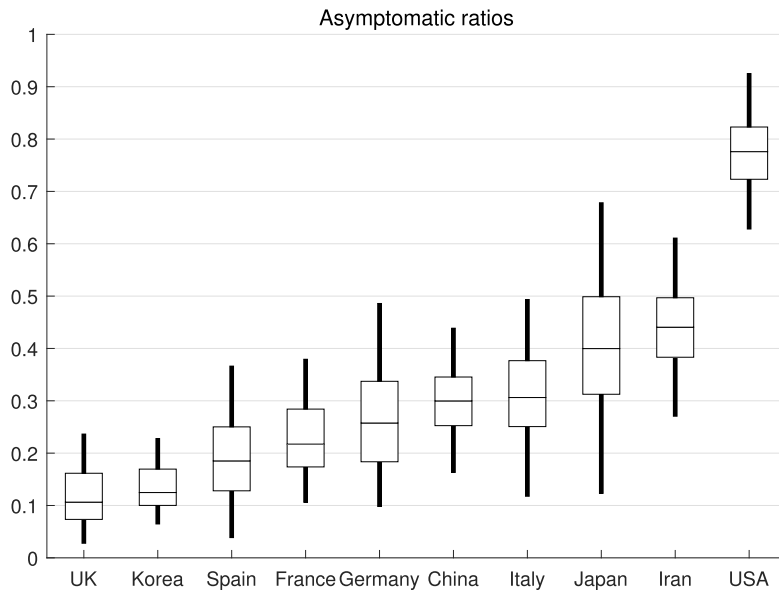


FIGURE 4. Box plots of asymptomatic ratios  $p_A(t)$  at  $t = 7$ .



TABLE 2. The estimation of key transmission characteristics across different demographic regions.

	Demographic region	Italy	UK	USA	Spain	South Korea
$\mathcal{R}_0$	95% CI	1.31–2.96	1.91–4.76	1.29–6.88	1.74–3.35	2.12–3.15
	Mean value	2.04	2.96	3.28	2.30	2.52
	Median	1.92	2.75	2.95	2.19	2.48
Mean latent period	95% CI	5.10–13.80	4.10–16.50	3.36–13.82	4.09–8.50	3.02–4.33
	Mean value	9.04	8.49	7.14	5.79	3.88
	Median	8.64	7.66	6.69	5.67	3.95
Inflection point	95% CI	Apr.9 – Apr.19	Apr.12 – Apr.29	Jul.24 – Feb.6,2021	Apr.7 – May.5	Mar.10 – Mar.18
	Real date	Apr.19	–	–	Apr.23	Mar.15
maximal $I_q + \chi A_q$ ( $\times 10^5$ )	95% CI	0.99–1.22	0.64–1.16	17.72–33.75	0.89–1.16	0.078–0.093
	Real data	1.08257	–	–	1.01617	0.07577

---

	Demographic region	Iran	France	Germany	Japan	Chinese mainland
$\mathcal{R}_0$	95% CI	1.07–3.63	1.29–2.68	1.77–3.10	1.59–5.16	1.70–2.63
	Mean value	2.07	1.74	2.30	2.44	2.08
	Median	1.98	1.65	2.25	2.14	2.05
Mean latent period	95% CI	4.60–8.27	4.14–9.85	3.45–10.11	3.43–4.89	2.79–4.03
	Mean value	6.63	6.14	5.96	4.49	3.54
	Median	6.73	5.81	5.68	4.61	3.59
Inflection point	95% CI	Mar.28 – Apr.10	Apr.9 – Jun.14	Apr.1 – Apr.12	Apr.24 – May.7	Feb.14 – Feb.19
	Real date	Apr.5	Apr.21	Apr.6	Apr.27	Feb.17
maximal $I_q + \chi A_q$ ( $\times 10^5$ )	95% CI	0.21–0.54	0.58–1.13	0.61–0.81	0.07–0.15	0.49–0.65
	Real data	0.34887	0.91763	0.72864	0.11869	0.58016

TABLE 3. The average estimation of several parameters across different demographic regions.

Parameters	Chinese mainland	South Korea	Japan	UK	USA
$\hat{\beta}$	0.4542	0.4492	0.4625	0.3873	0.3879
$\hat{\theta}$	0.6851	0.8074	0.6611	0.7667	0.6832
$\hat{\mu}$	0.2989	0.3985	0.1827	0.2034	0.0960

---

Parameters	France	Italy	Spain	Iran	Germany
$\hat{\beta}$	0.3624	0.4062	0.4759	0.3873	0.4956
$\hat{\theta}$	0.6535	0.7750	0.7494	0.6073	0.7811
$\hat{\mu}$	0.3780	0.1975	0.1503	0.3601	0.2291

efficiency of asymptomatic cases is average less than 50% that of symptomatic individuals. Reported by [3], the transmission rate of every undocumented infectious individual was 55% the transmission rate of documented infectious case. Further, thanks to the high quarantine rate and detection rate (see  $\hat{\theta}$  and  $\hat{\mu}$  in Tab. 3), which benefits from the efficiency of contacts tracing, Chinese mainland and South Korea have already brought the

TABLE 4. The reference value of the initial parameter estimation to the optimization problem.

Parameter	Chinese mainland	South Korea	Iran	Japan	Italy	generic initial estimation
$\eta$	–	–	–	–	–	0.75
$\theta$	–	–	–	–	–	0.75
$\rho$	0.85	0.9	0.75	0.5	0.85	–
$\beta$	–	–	–	–	–	0.4
$\varphi$	$3.696 \times 10^{-5}$	$9.696 \times 10^{-4}$	$2.7 \times 10^{-5}$	$2.7 \times 10^{-5}$	$1.1068 \times 10^{-5}$	–
$\varepsilon$	–	–	–	–	–	0.5
$\alpha$	–	–	–	–	–	0.2
$\gamma_A$	–	–	–	–	–	$10^{-4}$
$\gamma_I$	–	–	–	–	–	$7 \times 10^{-4}$
$p$	$10^{-4}$	$2.15 \times 10^{-4}$	$1.7 \times 10^{-3}$	$1.5 \times 10^{-4}$	$5.15 \times 10^{-3}$	–
$\mu$	–	–	–	–	–	0.2
$\gamma_{A_q}$	0.03	0.04	0.017	0.017	0.01	–
$z_1$	0.045	0.03	0.06	0.065	0.035	–
$z_2$	0.026	0.025	0.0306	0.045	0.02	–
$a$	28	29	22	50	45	–
$b$	–	–	–	–	–	5
$E_0$	–	–	–	–	–	$I_{q_0} \times 1.1$
$A_0$	–	–	–	–	–	$I_{q_0} \times 0.29$
$I_0$	–	–	–	–	–	$I_{q_0} \times 0.2$
$E_{q_0}$	–	–	–	–	–	$I_{q_0} \times 0.26$
$R_{2_0}$	–	–	–	–	–	$R_{1_0} \times 0.33$
$A_{q_0}$	–	–	–	–	–	$I_{q_0} \times 1/3$

---

Parameter	Spain	Germany	France	UK	USA	generic initial estimation
$\eta$	–	–	–	–	–	0.75
$\theta$	–	–	–	–	–	0.75
$\rho$	0.89	0.9	0.916	0.85	0.5	–
$\beta$	–	–	–	–	–	0.4
$\varphi$	$9.9135 \times 10^{-6}$	$1.9068 \times 10^{-5}$	$2.81068 \times 10^{-5}$	$2.6695 \times 10^{-5}$	$3.21 \times 10^{-6}$	–
$\varepsilon$	–	–	–	–	–	0.5
$\alpha$	–	–	–	–	–	0.2
$\gamma_A$	–	–	–	–	–	$10^{-4}$
$\gamma_I$	–	–	–	–	–	$7 \times 10^{-4}$
$p$	$7.8 \times 10^{-3}$	$2.5 \times 10^{-3}$	$4.47 \times 10^{-4}$	$2.66 \times 10^{-3}$	$1.13 \times 10^{-3}$	–
$\mu$	–	–	–	–	–	0.2
$\gamma_{A_q}$	0.098	0.035	0.022	0.2	0.00307	–
$z_1$	0.025	0.05	0.0187	0.035	0.31	–
$z_2$	0.02	0.045	0.0144	0.025	0.123	–
$a$	64	20	48	20	44	–
$b$	–	–	–	–	–	5
$E_0$	–	–	–	–	–	$I_{q_0} \times 1.1$
$A_0$	–	–	–	–	–	$I_{q_0} \times 0.29$
$I_0$	–	–	–	–	–	$I_{q_0} \times 0.2$
$E_{q_0}$	–	–	–	–	–	$I_{q_0} \times 0.26$
$R_{2_0}$	–	–	–	–	–	$R_{1_0} \times 0.33$
$A_{q_0}$	–	–	–	–	–	$I_{q_0} \times 1/3$

epidemic under control. In contrast to these two regions, the detection rate of asymptomatic carriers  $\hat{\mu}$  was much low, and the epidemic situation of the United States seems to have been more severe.

On the other hand, since the early recovery rate in these high-prevalence areas is generally low because of the long incubation period and high incidence of COVID-19 within aged population, it is therefore imminent to continuously strengthen prevention and control measures and improve medical standards. According

to the current result, strengthening the prevention and control measures can slow the development of the epidemic.

To generalize this work, we need more time-dependent parameters instead of constant ones. In the present situation, possibly unsteady infection prevention and control measures may induce that our prediction lacks precision. Besides, due to the local convergence of the algorithm, we need to find the appropriate initial estimation to accelerate the process of fitting the reported data. In future studies, it would be important to design a more effective algorithm for this kind of high-dimensional optimization problem.

### APPENDIX A. DERIVATION OF EQUATION (3.1)

In this appendix we provide detailed derivation of the basic production number  $\mathcal{R}_0$  using the next-generation matrix approach [11]. At the disease-free steady state  $S = N_e$ , we consider the following closed linearized infection subsystem of the model (2.1):

$$\begin{cases} \frac{dE}{dt} = (1 - \rho)\varphi N_e(I + \varepsilon E + \beta A) - \alpha E, \\ \frac{dI}{dt} = \alpha\eta E - \gamma_I I - \theta I, \\ \frac{dA}{dt} = \alpha(1 - \eta)E - \mu A - \gamma_A A. \end{cases}$$

If we set  $\mathbf{x} = (E, I, A)^\top$ , where the symbol  $\top$  denotes transpose, the subsystem can be rewritten in a vector form as follows:

$$\frac{d\mathbf{x}}{dt} = (\mathbf{F} - \mathbf{V})\mathbf{x},$$

where  $\mathbf{F}$  and  $\mathbf{V}$  represent the transmission and transition matrices of the linearized infection subsystem, respectively. More specifically,

$$\mathbf{F} = \begin{pmatrix} \varepsilon(1 - \rho)\varphi N_e & (1 - \rho)\varphi N_e & (1 - \rho)\varphi\beta N_e \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and

$$\mathbf{V} = \begin{pmatrix} \alpha & 0 & 0 \\ -\alpha\eta & \gamma_I + \theta & 0 \\ -\alpha(1 - \eta) & 0 & \mu + \gamma_A \end{pmatrix}.$$

Then from the next-generation matrix given by

$$\mathbf{FV}^{-1} = \begin{pmatrix} (1 - \rho)\varphi N_e \left\{ \frac{\varepsilon}{\alpha} + \frac{\beta(1 - \eta)}{\mu + \gamma_A} + \frac{\eta}{\gamma_I + \theta} \right\} & (1 - \rho)\varphi N_e \frac{1}{\gamma_I + \theta} & (1 - \rho)\varphi\beta N_e \frac{1}{\mu + \gamma_A} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

we have the basic reproduction number  $\mathcal{R}_0$  as follows:

$$\mathcal{R}_0 = \|\mathbf{FV}^{-1}\| = (1 - \rho)\varphi N_e \left\{ \frac{\varepsilon}{\alpha} + \frac{\beta(1 - \eta)}{\mu + \gamma_A} + \frac{\eta}{\gamma_I + \theta} \right\},$$

where  $\|\cdot\|$  denotes the matrix spectral norm.

## APPENDIX B. DETAILED DESCRIPTION OF ALGORITHM 1

---

**Algorithm 2:** Adaptive MCMC sampling algorithm for estimation of parameters in the SEIR model

---

**Input:**  $k_{max}, T, L$  # MCMC sample size, surveillance period, and Markov Chain State Size  
 $\{\tilde{I}_q(t), \tilde{R}_1(t)\}_{t \in [1, \dots, T]}$  # Surveillance data of confirmed and removed cases  
 $\Theta_{hyper}^0, \Theta_{hyper}^{upper}, \Theta_{hyper}^{lower}$  # Initial values, upper and lower bounds of hyperparameters (see Table 4)

**Output:**  $\{\hat{\Theta}^{(k)}\}_{k \in [1, \dots, k_{max}]}$  # MCMC sample of parameters of interest (see the text for detail)

- 1 **foreach**  $k \in [1, \dots, k_{max}]$  **do**
- 2      $\Theta_0 \leftarrow \text{fmincon}(\tilde{I}_q, \tilde{R}_1, \Theta_{hyper}^0)$  # First guess of  $\Theta^{(k)}$  using sequential quadratic programming
- 3     # implemented by MATLAB function `fmincon.m`
- 4      $\sigma^2 \leftarrow C(\tilde{I}_q, \tilde{R}_1; \Theta_0), p(\Theta) \leftarrow \text{Gauss}(\Theta_0, \sigma^2)$
- 5     **foreach**  $j \in [1, \dots, L]$  **do**
- 6          $Q(\Theta | \Theta_{j-1}) \leftarrow \text{Inv-Gamma}(\lambda, \beta), \lambda = C(\tilde{I}_q, \tilde{R}_1; \Theta_{j-1}), \beta = 1$
- 7         # Conjugate prior (inverse gamma distribution) with shape parameter  $\lambda$  and scale parameter  $\beta$
- 8          $\Theta^\dagger \sim Q(\Theta | \Theta_{j-1})$  # Generate a proposal by sampling from  $Q(\Theta | \Theta_{j-1})$
- 9         **if**  $C(\tilde{I}_q, \tilde{R}_1; \Theta^\dagger) < C(\tilde{I}_q, \tilde{R}_1; \Theta_{j-1})$  **then**
- 10              $\Theta_j \leftarrow \Theta^\dagger$
- 11         **else**
- 12              $\alpha(\Theta^\dagger, \Theta_{j-1}) \leftarrow \min\left\{\frac{p(\Theta^\dagger)Q(\Theta_{j-1}|\Theta)}{p(\Theta_{j-1})Q(\Theta|\Theta_{j-1})}, 1\right\}$  # Calculate the acceptance probability
- 13              $r \leftarrow \text{uniform}(0, 1);$  # Draw a random number uniformly distributed within  $[0, 1]$
- 14             **if**  $r < \alpha(\Theta^\dagger, \Theta_{j-1})$  **then**  $\Theta_j \leftarrow \Theta^\dagger$  **else**  $\Theta_j \leftarrow \Theta_{j-1}$
- 15         **end**
- 16     **end**
- 17      $\hat{\Theta}^{(k)} \leftarrow \frac{1}{L} \sum_{j \in [1, \dots, L]} \Theta_j$  # Calculate the  $k$ th MCMC sample of the parameters
- 18 **end**
- 19 **return**  $\{\hat{\Theta}^{(k)}\}$

---

*Acknowledgements.* This work was supported by the National Natural Science Foundation of China (11871343) and JSPS KAKENHI (15H05707). QG thanks D. Gao and X. Pan for their valuable advice. KA thanks H. Toshiyoshi for his valuable discussion.

## REFERENCES

- [1] Epidemiology Working Group for NCIP Epidemic Response, The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chin. J. Epidemiol.* **41** (2020) 145–151.
- [2] H. Haario, M. Laine, A. Mira and E. Saksman, DRAM: efficient adaptive MCMC. *Stat. Comput.* **16** (2006) 339–354.

- [3] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang and J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368** (2020) 489–493.
- [4] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S.S. Musa *et al.*, A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int. J. Infect. Dis.* **93** (2020) 211–216.
- [5] M. Lipsitch, Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300** (2003) 1966–1970.
- [6] Y. Liu, A.A. Gayle, A. Wilder-Smith and J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27** (2020).
- [7] K. Mizumoto, K. Kagaya, A. Zarebski and G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **25** (2020) 2000180.
- [8] K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies *et al.*, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5** (2020) E261–E270.
- [9] B. Tang, N.L. Bragazzi, Q. Li, S. Tang, Y. Xiao and J. Wu, An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). *Infect. Dis. Modell.* **5** (2020) 248–255.
- [10] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao and J. Wu, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.* **9** (2020) 462.
- [11] P. van den Driessche and J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math. Biosci.* **180** (2002) 29–48.
- [12] P.G. Walker, C. Whittaker, O. Watson, M. Baguelin, K.E.C. Ainslie, S. Bhatia *et al.*, Report 12: The global impact of COVID-19 and strategies for mitigation and suppression. Technical report, Imperial College London (2020). <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gidafellowships/Imperial-College-COVID19-Global-Impact-26-03-2020v2.pdf>.
- [13] K. Wan, J. Chen, C. Lu, L. Dong, Z. Wu and L. Zhang, When will the battle against novel coronavirus end in Wuhan: A SEIR modeling analysis. *J. Glob. Health* **10** (2020) 011002.
- [14] Y. Wei, Z. Lu, Z. Du, Z. Zhang, Y. Zhao, S. Shen *et al.*, Fitting and forecasting the trend of COVID-19 by seir (+ caq) dynamic model. *Chin. J. Epidemiol.* **41** (2020) 470–475.
- [15] J.T. Wu, K. Leung and G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395** (2020) 689–697.
- [16] S. Zhao, Q. Lin, J. Ran, S.S. Musa, G. Yang, W. Wang *et al.*, Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *Int. J. Infect. Dis.* **92** (2020) 214–217.