Original Paper

# General Model for COVID-19 Spreading With Consideration of Intercity Migration, Insufficient Testing, and Active Intervention: Modeling Study of Pandemic Progression in Japan and the United States

Choujun Zhan[1], BSc, PhD; Chi Kong Tse[2], BEng, PhD; Zhikang Lai[3], BSc; Xiaoyun Chen[3], BSc; Mingshen Mo[3], BSc

[1]South China Normal University, Guangzhou, China
[2]City University of Hong Kong, Hong Kong, Hong Kong
[3]Sun Yat-sen University, Guangzhou, China

**Corresponding Author:**
Chi Kong Tse, BEng, PhD
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong, 0000
Hong Kong
Phone: 852 92701816
Email: cktse@ieee.org

## Abstract

**Background:**   The coronavirus disease (COVID-19) began to spread in mid-December 2019 from Wuhan, China, to most provinces in China and over 200 other countries through an active travel network. Limited by the ability of the country or city to perform tests, the officially reported number of confirmed cases is expected to be much smaller than the true number of infected cases.

**Objective:**   This study aims to develop a new susceptible-exposed-infected-confirmed-removed (SEICR) model for predicting the spreading progression of COVID-19 with consideration of intercity travel and the difference between the number of confirmed cases and actual infected cases, and to apply the model to provide a realistic prediction for the United States and Japan under different scenarios of active intervention.

**Methods:**   The model introduces a new state variable corresponding to the actual number of infected cases, integrates intercity travel data to track the movement of exposed and infected individuals among cities, and allows different levels of active intervention to be considered so that a realistic prediction of the number of infected individuals can be performed. Moreover, the model generates future progression profiles for different levels of intervention by setting the parameters relative to the values found from the data fitting.

**Results:**   By fitting the model with the data of the COVID-19 infection cases and the intercity travel data for Japan (January 15 to March 20, 2020) and the United States (February 20 to March 20, 2020), model parameters were found and then used to predict the pandemic progression in 47 regions of Japan and 50 states (plus a federal district) in the United States. The model revealed that, as of March 19, 2020, the number of infected individuals in Japan and the United States could be 20-fold and 5-fold as many as the number of confirmed cases, respectively. The results showed that, without tightening the implementation of active intervention, Japan and the United States will see about 6.55% and 18.2% of the population eventually infected, respectively, and with a drastic 10-fold elevated active intervention, the number of people eventually infected can be reduced by up to 95% in Japan and 70% in the United States.

**Conclusions:**   The new SEICR model has revealed the effectiveness of active intervention for controlling the spread of COVID-19. Stepping up active intervention would be more effective for Japan, and raising the level of public vigilance in maintaining personal hygiene and social distancing is comparatively more important for the United States.

XSL•FO
RenderX

## Introduction

### Background

The global spread of the coronavirus disease (COVID-19) has shown no sign of subsiding since its emergence in Wuhan, China, in December 2019 [1]. As of March 21, 2020, a total of 276,472 cases of COVID-19 infection have been confirmed in over 185 countries, with a death toll of 11,417 [2]. Different control strategies at different levels of stringency have been applied to slow the spread of the virus in different countries [3]. Although some countries have seen peaks of infected cases and have observed significant reductions in the number of new infections in the local communities [2,4], the spread has continued in many countries, and surges in infected cases have been observed in Europe, the United States, and Australia. Intercity travel has been found to be a contributing factor to the rapid spread of the virus [5,6]. Thus, effective models for describing the pandemic progression in different cities should take into consideration the volume of intercity travel [4,7]. Additionally, the virus spread from one country to another through the air transportation network [8-10]. Hence, population flow is expected to play an important role in the transmission of COVID-19, and travel restrictions would effectively slow the transmission of COVID-19 [11]. Furthermore, the rapid spread of the virus in a population has often been a result of delayed information or unawareness of the real situation in that population, despite the wide dissemination of information related to COVID-19 outbreaks in other parts of the world. The most notable information latency lies in the number of confirmed cases reported, which depends on the ability of the particular country or city to perform tests as well as the possible bureaucracy in the local system of reporting. Thus, the number of confirmed cases is almost certainly not the true number of infected individuals at any given time [12], and an improved model for predicting the spreading progression should incorporate the latency associated with the reporting system as well as the possible missing cases leading to delay and loss of information. The traditional susceptible-exposed-infectious-recovered (SEIR) model [13,14] thus has obvious shortfalls in describing the spreading dynamics of the COVID-19 pandemic. In this work, we attempt to fill the main gap between the number of confirmed cases and the actual number of infected cases. Specifically, in the proposed model, an infected individual may become a confirmed case and then recovered or removed. Moreover, an infected individual may also be recovered or removed without being confirmed as infected. In other words, the basic model proposed here is a susceptible exposed infected confirmed removed (SEICR) model, which has an additional state corresponding to an individual having been confirmed by the authority as being infected.

On the basis of an SEICR model, we developed a model incorporating intercity travel data that accounts for any increase or decrease in the number of exposed and infected individuals in a city due to intercity migration. Furthermore, the level of intervention in the form of travel restriction, regional lockdown, or other active control measures would profoundly influence the rapidity of the virus spread and the eventual number of infected cases. The model should, therefore, allow the level of active intervention to be included as a control parameter and produce the appropriate progression profile. A specific parameter was used to adjust the level of active intervention in the simulation of future progression profiles, which corresponds quantitatively to the increase in the number of individuals eventually infected due to an additional infected individual at any given time. In this work, we apply the model to study the COVID-19 spreading progression in Japan and the United States. Data of confirmed and recovered cases in 47 Japanese prefectures or regions (January 15 to March 20, 2020) and 50 US states plus Washington, DC (February 20 to March 20, 2020) were used for fitting with the model and retrieval of parameter values. The parameters found were then adjusted to produce future progression trajectories corresponding to the implementation of different levels of active intervention.

### Data

The World Health Organization has currently set the alert level of COVID-19 to the highest and has made data related to the pandemic available to the public in a series of situation reports as well as other formats [15]. Our data include the number of confirmed infected cases, the cumulative number of confirmed infected cases, the number of recovered cases, and death tolls for 47 individual prefectures and regions in Japan, from January 15 to March 20, 2020, and for 50 states and a federal district (Washington, DC) in the United States from February 20 to March 20, 2020. Data organized in convenient formats are also available elsewhere [12,16,17]. Moreover, the monthly intercity migration data for February 2020 are available from official statistics provided by the Japanese government [18] and are used as indicative migration strengths between prefectures or regions in Japan. For the United States, annual data for the volume of interstate travelers are available from the Census Bureau [19] and the Bureau of Transportation Statistics [20].

## Methods

### Migration-Data Augmented SEICR Model

In the proposed SEICR model, every individual would assume one of five possible states at any time, namely, susceptible ($S$), exposed ($E$), infected ($I$), confirmed ($C$), and recovered or removed ($R$). Compared to the traditional SEIR model [13,14], the new SEICR model has an additional $C$ state, corresponding to an individual that has been confirmed by the authority as infected. Thus, not all infected individuals will become confirmed, and some infected individuals will transit to the recovered state without going through the confirmed state. For city or region $j$, the number of individuals in the five states are $S_j(t)$, $E_j(t)$, $I_j(t)$, $C_j(t)$, and $R_j(t)$ at time $t$. The transitions of the five states are illustrated in Figure 1.

XSL•FO

**RenderX**

In addition, $P_j(t)$ stands for the population of region $j$. Furthermore, to account for intercity movement, we introduce a migration strength, $m_{ij}(t)$, which represents an indicative volume of people moving from region $i$ to region $j$ at time $t$ [4]. The augmented SEICR model is given as follows:

$$\Delta I_j(t) = (1 - \lambda_j - \gamma_j^{(I)})I_j(t) + \kappa E_j(t) + k_I \left( \sum_{i=1}^{N} \left( \frac{I_i(t)m_{ij}(t)}{P_i(t)} \right) - \frac{I_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} \right)$$

$$\Delta E_j(t) = \frac{\beta_j I_j(t) + \rho_j C_j(t)}{N_j^s(t)} S_j(t) + \frac{\alpha_j E_j(t)}{N_j^s(t)} S_j(t) - \kappa E_j(t) + \sum_{i=1}^{N} \left( \frac{E_i(t)m_{ij}(t)}{P_i(t)} \right)$$
$$- \frac{E_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$

$$\Delta S_j(t) = -\frac{\beta_j I_j(t) + \rho_j C_j(t)}{N_j^s(t)} S_j(t) - \frac{\alpha_j E_j(t)}{N_j^s(t)} S_j(t) + \sum_{i=1}^{N} \left( \frac{S_i(t)m_{ij}(t)}{P_i(t)} \right)$$
$$- \frac{S_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$

$$\Delta C_j(t) = \lambda_j I_j(t) - \gamma_j^{(C)} C_j(t)$$

$$\Delta R_j(t) = \gamma_j^{(I)} I_j(t) + \gamma_j^{(C)} C_j(t)$$

$$\Delta P_j(t) = \sum_{i=1}^{N} m_{ij}(t) - \sum_{i=1}^{N} m_{ji}(t)$$

$$\Delta N_j^s(t) = k_I \left( \sum_{i=1}^{N} \left( \frac{I_i(t)m_{ij}(t)}{P_i(t)} \right) - \frac{I_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} \right) + \sum_{i=1}^{N} \left( \frac{E_i(t)m_{ij}(t)}{P_i(t)} \right)$$
$$- \frac{E_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} + \sum_{i=1}^{N} \left( \frac{S_i(t)m_{ij}(t)}{P_i(t)} \right) - \frac{S_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$

**(1)**

$\Delta I_j(t) = I_j(t + 1) - I_j(t)$, $\Delta E_j(t) = E_j(t + 1) - E_j(t)$, $\Delta S_j(t) = S_j(t + 1) - S_j(t)$, $\Delta C_j(t) = C_j(t + 1) - C_j(t)$, $\Delta R_j(t) = R_j(t + 1) - R_j(t)$, $\Delta N_j^s(t) = N_j^s(t + 1) - N_j^s(t)$, and $\Delta P_j(t) = P_j(t + 1) - P_j(t)$.

The meaning of each parameter is given in Textbox 1. In addition, we assumed that the recovered and confirmed individuals would stay in region $j$.

In this SEICR model, the number of individuals eventually infected is set initially at $N_j^s(t_0) = \delta_j P_j$ ($\delta_j$ being constant), implying that some effective measures have been taken by the authorities to limit the upper bound of the susceptible population. Moreover, in the case of inactive or less effective intervention, or even unchecked spread, the growth in the number of infected cases will add to the eventual infected number. Hence, the number of eventually infected individuals should increase for each additional infected or exposed individual at time $t$. This is equivalent to adding an extra term (the boxed term below) to $\Delta S_j(t)$ and $\Delta N_j^s(t)$. Furthermore, as the number of infected cases increases and approaches a saturating percentage $k_h$ (such as a herd-immunity condition),

the spreading is expected to slow down significantly (ie, $\alpha_j$ and $\beta_j$ will drop as $N_j^s$ approaches $k_h P_j$, where $0 < k_h < 1$). Thus, we have:

$$\Delta S_j(t) = -\frac{\beta_j}{N_j^s(t)} I_j(t) S_j(t) - \frac{\alpha_j}{N_j^s(t)} E_j(t) S_j(t) + \sum_{i=1}^{N} \left( \frac{S_i(t)m_{ij}(t)}{P_i(t)} \right)$$
$$- \frac{S_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} + \boxed{k_j^{(c)} \left( \Delta I_j(t) + \Delta E_j(t) \right) \left( 1 - \frac{N_j^s(t)}{k_h P_j(t)} \right)}$$

$$\Delta N_j^s(t) = k_I \left( \sum_{i=1}^{N} \left( \frac{I_i(t)m_{ij}(t)}{P_i(t)} \right) - \frac{I_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} \right) + \sum_{i=1}^{N} \left( \frac{E_i(t)m_{ij}(t)}{P_i(t)} \right)$$
$$- \frac{E_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)} + \sum_{i=1}^{N} \left( \frac{S_i(t)m_{ij}(t)}{P_i(t)} \right) - \frac{S_j(t)\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}$$
$$+ \boxed{k_j^{(c)} \left( \Delta I_j(t) + \Delta E_j(t) \right) \left( 1 - \frac{N_j^s(t)}{k_h P_j(t)} \right)}$$

**(2)**

$k_j^{(c)}$ is an inverse indicator of the level of active intervention implemented, and corresponds quantitatively to an increase in the number of eventual infected individuals for each additional infected or exposed individual in region $j$, and the added term in $\Delta S_j(t)$ and $\Delta N_j^s(t)$ will approach zero as $N_j^s \rightarrow k_h P_j$. The meanings of other parameters are given in Textbox 1. Again, the recovered and confirmed individuals are assumed to stay in region $j$.

The model given in (1) and (2) is general in the sense that it applies to populations with varied effectiveness levels of active intervention during the outbreak. To further facilitate the assessment of control measures implemented in region $j$, we defined the level of active intervention as:

$$\psi_j = \frac{1}{k_j^{(c)}}$$

**(3)**

Thus, if $\psi_j > 1$, the control measures are effective and the progression is limited such that $k_j^{(c)} < 1$. The total number of eventually infected individuals is equal to $\sum_{j=1}^{N} P_j(t_0)\delta_j$. However, in the case of less effective or ineffective control (ie, $\psi_j < 1$), infected and exposed individuals continue to spread the disease, and for each additional infected individual, there will be $k_j^{(c)}$ more eventual infected individuals, and the pandemic progresses until the number of infected cases reaches $k_h P_j$.

**Figure 1.** State transition flow chart. C: confirmed; E: exposed; I: infected; R: removed; S: susceptible.

XSL•FO

**RenderX**
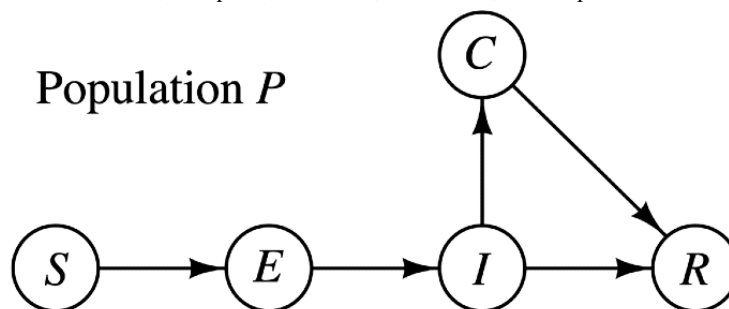
**Textbox 1.** Parameters of the migration-data augmented susceptible-exposed-infected-confirmed-removed model.

$\alpha_j$

Rate of infecting a susceptible individual by an exposed individual in region $j$

$\beta_j$

rate of infecting a susceptible individual by an infected individual in region $j$

$\rho_j$

rate of infecting a susceptible individual by a confirmed individual in region $j$

$\lambda_j$

confirmed rate of infected individuals in region $j$

$\kappa_j$

rate of an exposed individual becoming infected

$\gamma_j^{(I)}$

recovery rate of an infected but not confirmed individual in region $j$

$\gamma_j^{(C)}$

recovery rate of a confirmed individual in region $j$

$k_I$

possibility of an infected individual moving from one region to another

$k_j^{(c)}$

increase in number of individuals eventually infected for each additional infected or exposed individual in region $j$

$\psi_j$

level of active intervention, $\psi_j = 1/k_j^{(c)}$

$k_h$

proportion of population infected achieving no further spreading (ie, absolute upper bound for $N_j^s$ for all $j$)

$\delta_j$

initial percentage of eventual infected individuals in region $j$

$I_{j\,0}$

initial number of infected individuals in region $j$

$E_{j\,0}$

initial number of individuals in region $j$

$C_{j\,0}$

initial number of confirmed infected individuals in region $j$

## Parameter Identification

The model represented by (1) and (2) describes the dynamics of the pandemic propagation with consideration of human migration dynamics and the reality of insufficient testing that leads to confirmed infected cases being fewer than the actual infected cases. The parameters in (1) and (2) are unknown and to be estimated from historical data of $C$ and $R$. We solve this parameter identification problem via constrained nonlinear programming with the objective of finding an estimated growth trajectory that fits the data. An estimated number of infected cases of each city can be generated from (1) and (2) with unknown set $\theta_j$ given by:

$$\theta_j = \{\alpha_j, \beta_j, \gamma_j, \delta_j, \lambda_j, \gamma_j^{(I)}, \gamma_j^{(C)}, k_j^{(c)}, I_{j,0}, E_{j,0}\}$$

**(4)**

$I_{j,0}(t) = I_j(t_0)$ and $E_{j,0}(t) = E_j(t_0)$ are the initial numbers of infected and exposed individuals in region $j$, and $\{\alpha_j, \beta_j, \gamma_j, \delta_j, \lambda_j, \gamma_j^{(I)}, \gamma_j^{(C)}, k_j^{(c)}\}$ are the model parameters of region $j$. Here, we assume that all confirmed cases are either quarantined or hospitalized and, hence, not infectious (ie, $\rho_j=0$). The unknown set is then $\Theta = \{\theta_1, \theta_2, \ldots, \theta_K, \kappa, k_I, k_h\}$, which essentially has $8K + 3$ unknowns, where $K$ is the number of regions in the entire

population under study. The identification of unknown parameters would require a considerable effort of computation.

Specifically, the parameter estimation problem can be formulated as the following constrained nonlinear optimization problem:

$$P_0: \min_{\Theta} \sum_{j=0}^{N} \left\| w_j^{(C)}(C(t_j) - \hat{C}(t_j)) \right\|_l + \left\| w_j^{(R)}(R(t_j) - \hat{R}(t_j)) \right\|_l$$

$$s.t. \begin{cases} \text{(i)} & \hat{x}(t+1) = \hat{x}(t) + F(\hat{x}(t)), \\ \text{(ii)} & \Theta_L \leq \Theta \leq \Theta_U, \end{cases}$$

**(5)**

$F(.)$ represents the model given by (1) and (2), $\omega_j^{(C)}$ and $\omega_j^{(R)}$ are the weighting coefficients, and $\hat{x}(t) = [\hat{C}(t), \hat{R}(t), I(t), E(t), S(t), P(t), \hat{N}^s(t)]$ is the set of estimated variables, with unknown set $\Theta$ being bounded between $\Theta_L$ and $\Theta_U$. In this work, an inverse approach is taken to find the unknown parameters and states by solving (5).

**Prediction**

The model parameters characterize the spreading dynamics, and once the set of parameters has been identified using the previously mentioned optimization procedure, we may generate future progression profiles by using the same set of parameters. Moreover, we may also adjust some of the parameters to examine different possible scenarios, corresponding to varying levels of active intervention $\psi_j = 1/k_j^{(c)}$, including travel restriction, mandatory quarantine, and other control measures. If the level of active intervention stays with the status quo, we will use the same value of $k_j^{(c)}$ for generating future progression profiles. Future paths under more active intervention can be predicted by reducing the value of $k_j^{(c)}$. In our study, by extending each simulation run to the forthcoming 200 days, we obtain a set of predicted progression profiles for each region in Japan and the United States. Moreover, different levels of active intervention can be assessed by adjusting parameter $k_j^{(c)}$ relative to the values found in each candidate set. For instance, by reducing $k_j^{(c)}$ and rerunning the simulation, we may assess the effect of tightening the control measures. In particular, we will consider three levels of active intervention: (1) staying with the status quo, corresponding to the same value of $\psi_j$ or $k_j^{(c)}$; (2) 2-fold step-up of active intervention, corresponding to $2\psi_j$ or $0.5k_j^{(c)}$; and (3) 4-fold step-up of active intervention, corresponding to $4\psi_j$ or $0.25k_j^{(c)}$

The pandemic progression profiles of 47 Japanese prefectures or regions were examined. We perform data fitting of the model, described by (1) and (2), using historical daily data of confirmed and recovered cases. For the United States, the pandemic progression profiles of 50 states and a federal district were examined. We again performed data fitting of the model using historical daily data of confirmed and recovered cases from February 20 to March 20, 2020, and obtained 100 candidate sets of parameters that satisfy the fitting criteria.

The level of public vigilance in exercising personal protective measures can also be incorporated in our model through adjusting infection rates $\alpha_j$ and $\beta_j$. We can, therefore, assess the combined effectiveness of active intervention and practicing protective measures in controlling the pandemic. Here, we varied $\alpha_j$, $\beta_j$, and $k_j^{(c)}$ from 10% to 100% of the originally identified values in 10 intervals, corresponding to 10 different levels of public vigilance and active intervention by the authorities. In particular, we assess $\alpha_j$ and $\beta_j$ as one property and $k_j^{(c)}$ as another (ie, varying $\alpha_j$ and $\beta_j$ in synchrony). Specifically, for each candidate parameter set, we perform 100 simulation runs for each combination of $\alpha_j$, $\beta_j$, and $k_j^{(c)}$, where $\alpha_j$, $\beta_j$, and $k_j^{(c)}$ vary from 10% to 100% of the original values in 10 steps. We then investigate the percentage of the population eventually infected in Japan and the United States.

## Results

### Parameters and Prediction for Japan

A typical candidate set of parameter values that fit well with the data from January 15, 2020, to March 20, 2020, is as follows: $1.3941 < k_j^{(c)} < 1.5979$; $0.0982 < \alpha_j < 0.1158$; $0.3895$ $\beta_j < 0.5163$; $0.0098 < \gamma_j^{(I)} < 0.0128$; $0.0027 < \gamma_j^{(C)} < 0.0047$; $0.0019 < \lambda_j < 0.0052$; $\kappa = 0.1861$; $k_h = 0.6514$. This set of parameters reflects an inadequate level of control to slow the spread of the disease, as indicated by the value of $k_j^{(c)}$ being larger than 1. Specifically, for each additional infected or exposed individual, the number of eventual infected individuals would increase by around 1.5 on average. The number of individuals eventually infected will approach a saturating percentage $k_h$.

We have identified 100 candidate sets of parameters that satisfy the fitting criteria, and for each set of parameters, we perform a separate simulation run. Figure 2 shows one particular simulation run of a well-fitted candidate set of parameters for 8 selected prefectures in Japan. The averaged results of all simulation runs are consolidated in the charts shown in Figure 2. Based on the data up to March 20, 2020, our model estimates that less than 3% of the infected cases are confirmed, with Hokkaido having the highest percentage (6.9%) and Hyogo-ken the least (1.5%), as shown in Figure 3(a). In other words, the actual number of infected individuals could be 20 times as many as the official confirmed number. Statistics of percentages for the population of confirmed and infected with the disease up to March 20, 2020, are shown in Figure 3(b).

We examine three cases corresponding to the level of active intervention being unchanged, 2-fold elevated, and 4-fold elevated. First, staying with the status quo ($k_j^{(c)}$ unchanged), if there is no further tightening of control aiming to slow the spread, all parameters of the candidate sets will remain unchanged. The total number of individuals eventually infected until September 23, 2020, in each region is shown in Figure 3(c). In this case, the number of infected individuals in Osaka-fu and Tokyo-to will reach about 2,300,000 and 600,000 (12% and 4.2% of the population), respectively, while most other regions will have around 5% of the population eventually

infected by September 23, 2020, as shown in Figure 3(d). In total, about 6.55% of the population in Japan will be infected.

Second, with two-fold elevated active intervention ($k_j^{(c)} \rightarrow 0.5k_j^{(c)}$), if active intervention is stepped up to twice the current level (ie, the value of $k_j^{(c)}$ is set to half of the original value in each simulation run), we observe a significant drop in the number of individuals eventually infected, as given in Figure 3(c). Specifically, the percentage of the population eventually infected by September 23, 2020, in Osaka-fu and Tokyo-to would drop to about 6.8% and 2.3%, respectively, while most other regions would drop to less than 2%, as shown in Figure 3(d). In total, about 4.14% of the population in Japan will be infected.

Third, with 4-fold elevated active intervention ($k_j^{(c)} \rightarrow 0.25k_j^{(c)}$), if active intervention is stepped up to four times the current level (ie, the value of $k_j^{(c)}$ set to a quarter of the original value in each simulation run), we observe a drastic drop in the number of individuals eventually infected, as given in Figure 3(c). Specifically, the percentage of the population eventually infected by September 23, 2020, in Osaka-fu and Tokyo-to would drop to about 4.1% and 2.3%, respectively, while most other regions would drop to less than 1%, as shown in Figure 3(d). In total, about 1.54% of the population in Japan will be infected.

In addition, our model estimates that the number of infected individuals could be 20 times as many as the currently confirmed number due to various reasons such as insufficient testing. Based on the data collected so far and assuming no further tightening of control, our model estimates about 6.65% of the population will be eventually infected, and a 4-fold elevation in control efforts may bring it down to 1.54% (about a 75% reduction) and end the pandemic sooner.

**Figure 2.** Official and estimated number of infected individuals in 8 selected prefectures in Japan (upper), the estimated number of infected individuals (not confirmed; middle), and the estimated number of exposed individuals (lower).
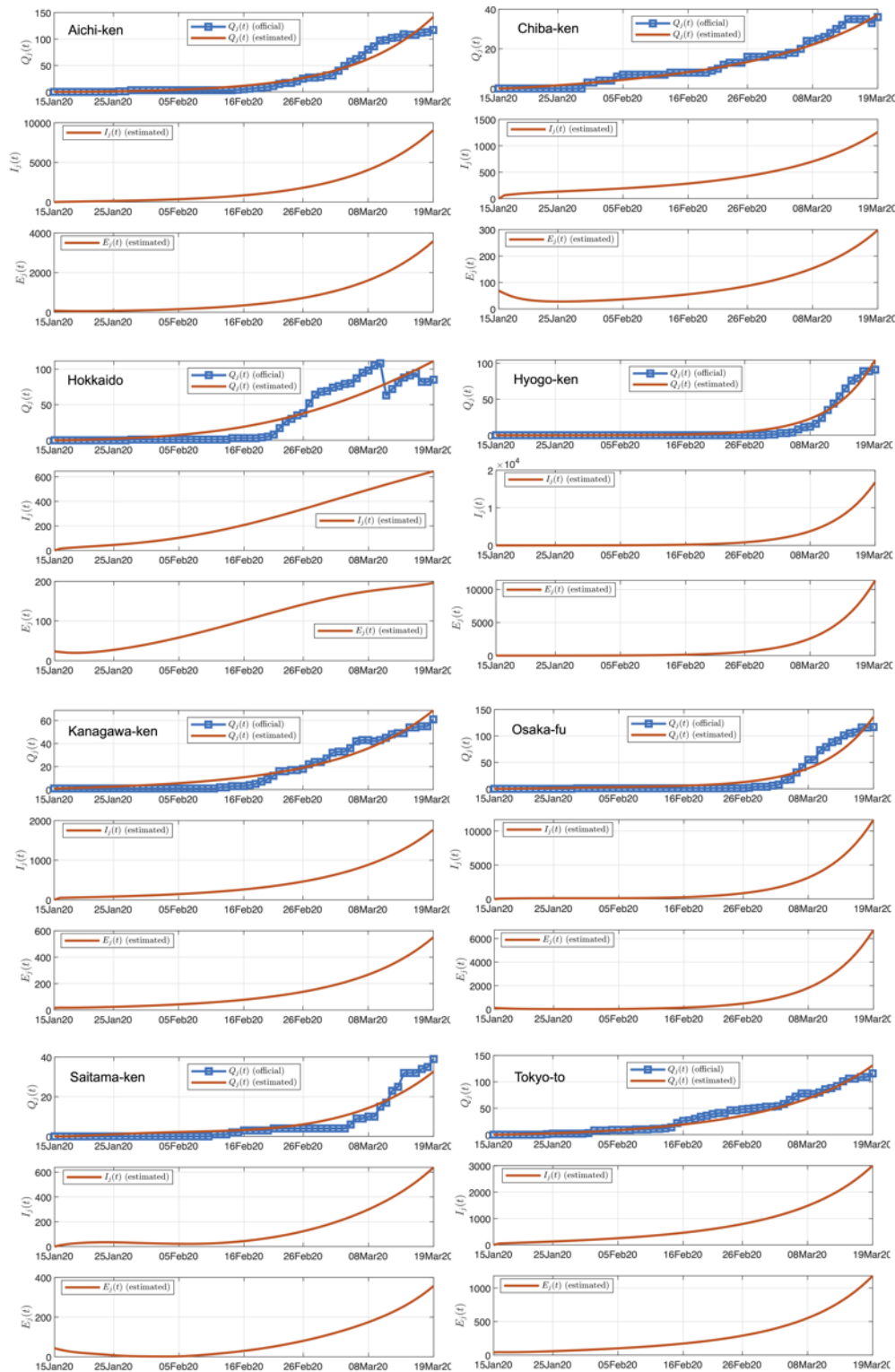
**Figure 3.** Statistics of data and predicted results for Japan. (a) Proportion of infected cases that are confirmed as of March 20, 2020; (b) number of confirmed cases and estimated number of infected cases as of March 20, 2020; (c) number of individuals eventually infected under three levels of active intervention; (d) proportion of population eventually infected under three levels of active intervention.
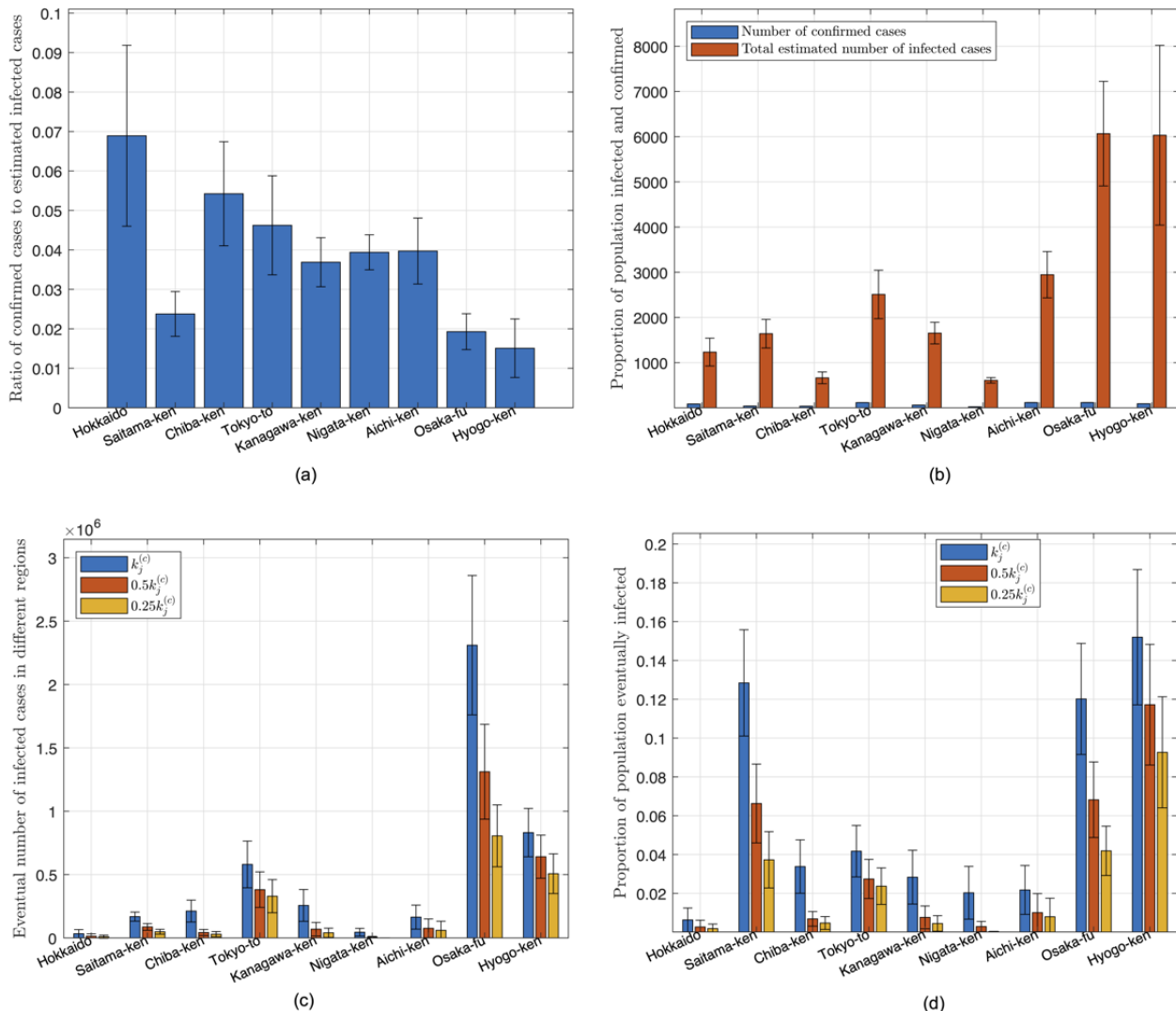


## Parameters and Prediction for United States

We present here the results for eight selected states having significant numbers of infected individuals as of March 20, 2020. Figure 4 shows one typical simulation run, showing the number of confirmed cases, the estimated number of infected individuals (not confirmed), and the estimated number of exposed individuals.

As of March 19, 2020, our model showed that less than 20% of the infected cases are confirmed, with Washington, DC having the highest percentage (36%) and Michigan state the least (0.7%), as shown in Figure 5(a). In other words, the actual number of infected individuals in the United States could be 5 times as many as the confirmed number. Statistics of percentages for the population of confirmed and infected with the disease up to March 19, 2020, are shown in Figure 5(b).

The key results for the three cases corresponding to three different levels of active intervention are as follows. First, staying with the status quo ($k_j^{(c)}$ unchanged), if there is no further tightening of control aiming to slow the spread, all parameters

of the candidate sets will remain unchanged. The total number of individuals eventually infected until September 23, 2020, in each state is shown in Figure 5(c). In this case, the number of infected individuals in California and New York State will reach about 5,800,000 and 7,300,000 (15% and 37.5% of population), respectively, while most other states will have less than 20% of the population eventually infected by September 23, 2020, as shown in Figure 5(d). In total, about 18.2% of the population in the United States will be infected.

Second, with 2-fold elevated active intervention ($k_j^{(c)} \rightarrow 0.5k_j^{(c)}$), if active intervention is stepped up to twice the current level (ie, the value of $k_j^{(c)}$ set to half of the original value in each simulation run), we observe a significant drop in the number of individuals eventually infected, as given in Figure 5(c). Specifically, the percentage of the population eventually infected by September 23, 2020, in California and New York State would drop to about 4.5% and 29.5%, respectively, while most other states would drop to less than 10%, as shown in Figure 5(d). In total, about 14% of the population in the United States will be infected.

Third, with 4-fold elevated active intervention ($k_j^{(c)} \to 0.5k_j^{(c)}$), if active intervention is stepped up to four times the current level (ie, the value of $k_j^{(c)}$ set to a quarter of the original value in each simulation run), we observe further reduction in the number of individuals eventually infected, as given in Figure 5(c). Specifically, the percentage of the population eventually infected by September 23, 2020, in California and New York State would drop to about 2.5% and 23%, respectively, while most other states would drop to less than 3%, as shown in Figure 5(d). In total, about 9.32% of the population in the United States will be infected.

The results of assessing the combined effectiveness of active intervention and practicing protective measures in controlling the pandemic through adjusting parameters $\alpha_j$, $\beta_j$, and $k_j^{(c)}$ are shown in Figure 6(a) and 6(b).

**Figure 4.** Official and estimated number of infected individuals in 8 selected states in the United States (upper), the estimated number of infected individuals (not confirmed; middle), and the estimated number of exposed individuals (lower).
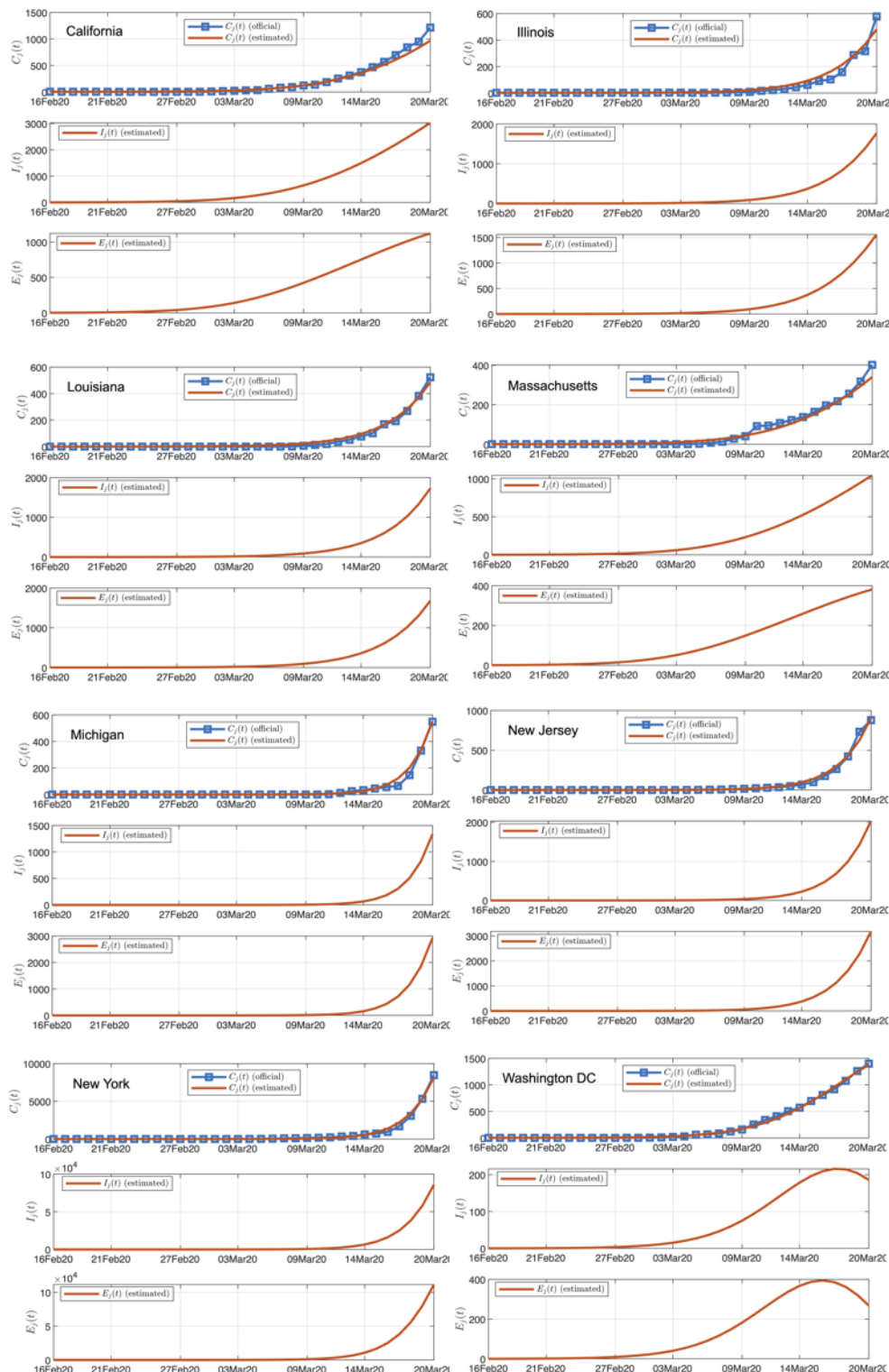
**Figure 5.** Statistics of data and predicted results for the United States. (a) Proportion of infected cases that are confirmed as of March 19, 2020; (b) number of confirmed cases and estimated number of infected cases as of March 19, 2020; (c) number of individuals eventually infected under three levels of active intervention; (d) proportion of population eventually infected under three levels of active intervention.

XSL•FO
**RenderX**
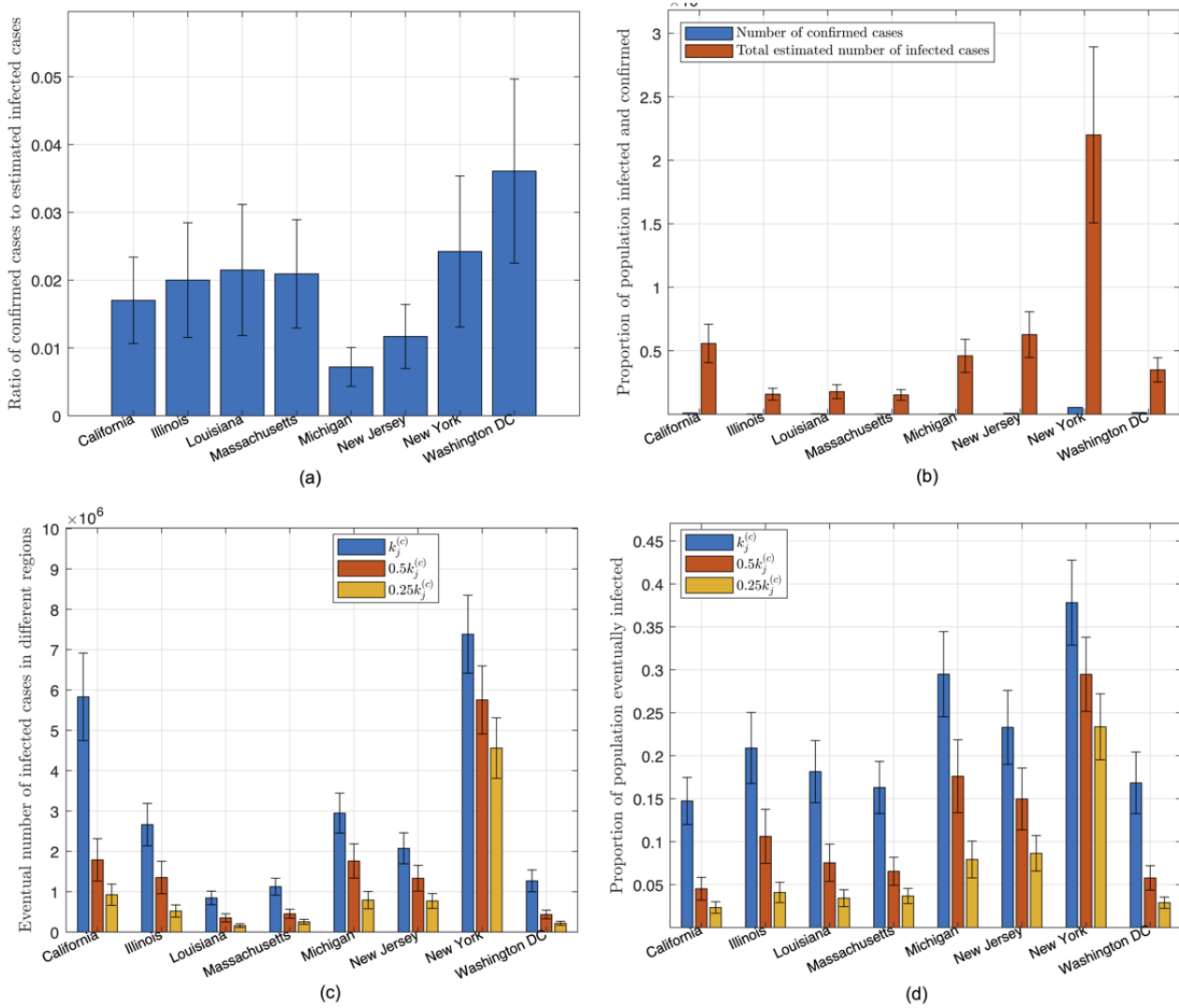
**Figure 6.** The proportion of the population eventually infected under different levels of the active intervention indicated by kj(c) (smaller the stronger) and maintaining personal hygiene and exercising protective measures indicated by αj, βj (smaller the stronger). (a) Japan; (b) the United States.



(a)



(b)
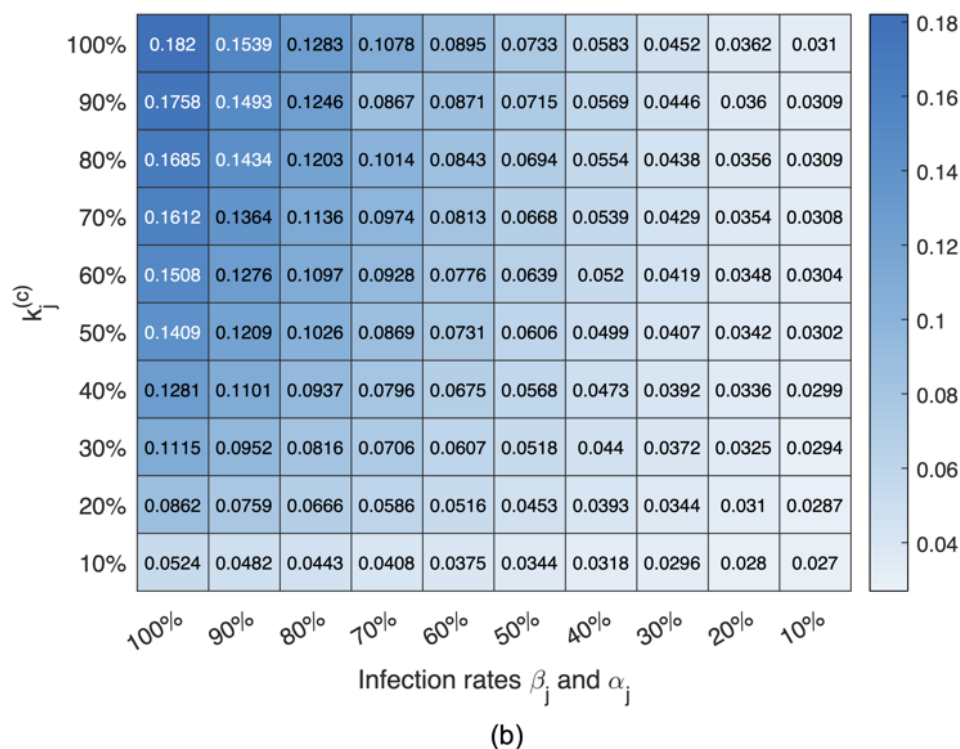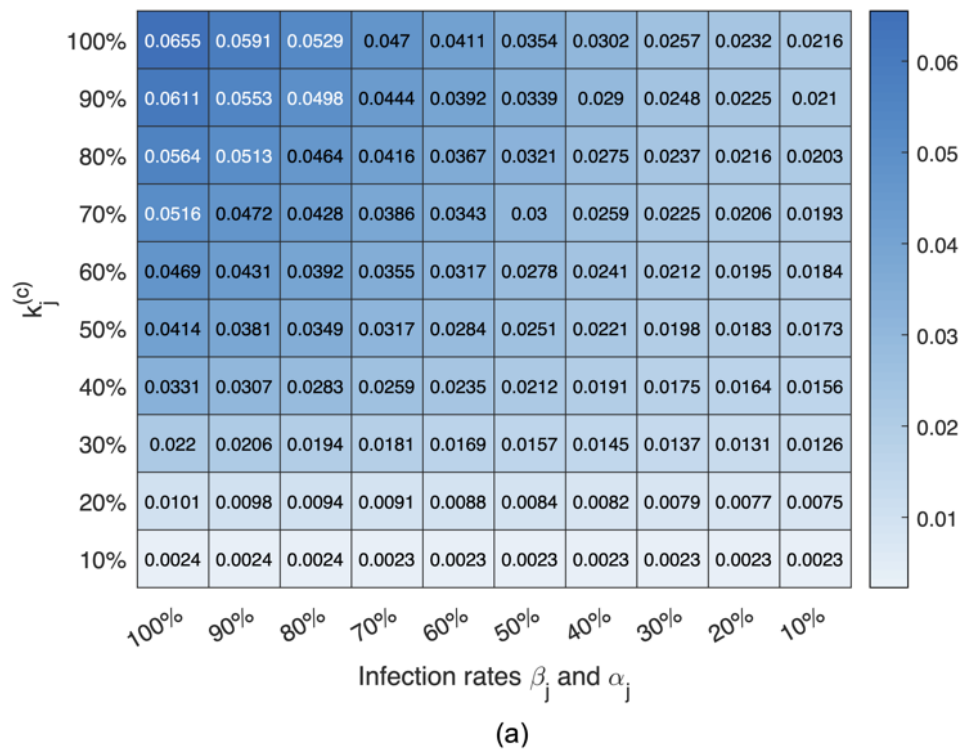
## Discussion

### Principal Findings

A significant step-up in the level of active intervention is necessary to slow the spread of the virus, especially for the United States. Based on the data collected up to March 20, 2020, and assuming no further tightening of the governments' control, our model estimates that about 6.55% and 18.2% of the population would eventually be infected in Japan and the United States, respectively, and a drastic 10-fold elevated active control may bring it down further to 0.24% and 5.24% for Japan and the United States, respectively.

Our results have highlighted the ability of the model in assessing the impact of active intervention through adjusting one of the parameters, namely, $\psi_j = 1/k_j^{(c)}$. Moreover, it has been widely

disseminated that maintaining personal hygiene is equally important in curbing the spread of the virus. The World Health Organization recommends several specific protective measures to be practiced by the public including frequent hand washing; maintaining social distancing, avoiding touching one's eyes, nose, and mouth; and practicing respiratory hygiene [21]. Recent studies have also shown that wearing surgical masks would help in some cases [22,23]. The level of public vigilance in exercising personal protective measures can also be incorporated in our model through adjusting infection rates $\alpha_j$ and $\beta_j$. We can, therefore, assess the combined effectiveness of active intervention and practicing protective measures in controlling the pandemic. Here, we varied $\alpha_j$, $\beta_j$, and $k_j^{(c)}$ from 10% to 100% of the originally identified values in 10 intervals, corresponding to 10 different levels of public vigilance and active intervention by the authorities. In particular, we assess $\alpha_j$ and $\beta_j$ as one property and $k_j^{(c)}$ as another (ie, varying $\alpha_j$ and $\beta_j$ in synchrony). Specifically, for each candidate parameter set, we performed 100 simulation runs for each combination of $\alpha_j$, $\beta_j$, and $k_j^{(c)}$, where $\alpha_j$, $\beta_j$, and $k_j^{(c)}$ varied from 10% to 100% of the original values in 10 steps. We then investigated the percentage of the population eventually infected in Japan and the United States. The results are shown in Figure 6(a) and (b).

The mean percentage of the population eventually infected under different combinations of parameter values for Japan and the United States are given in the charts shown in Figure 6(a) and (b), respectively. For instance, suppose the level of public vigilance has dramatically raised and the level of active intervention has been stepped up, resulting in a 90% reduction in the infected rates and a 90% reduction in $k_j^{(c)}$ (ie, parameters changed to $0.1\alpha_j$, $0.1\beta_j$, and $0.1k_j^{(c)}$). Referring to Figure 6(a) and (b), the percentage of the population eventually infected can be dramatically reduced to 0.23% for Japan and 2.7% for the United States. Similar interpretations can be taken for any other combination of public vigilance levels and active intervention.

Our results have highlighted an interesting difference between the effectiveness of government's active intervention and maintaining personal hygiene by the public for Japan and the United States. For Japan, we observed a 27-fold reduction (from 6.55% to 0.24%) in the percentage of individuals eventually infected upon a drastic 10-fold step-up of active intervention ($k_j^{(c)}$ changed to $0.1k_j^{(c)}$), whereas less than 3-fold reduction (from 6.55% to 2.16%) is observed in the percentage of individuals eventually infected upon the same 10-fold improvement in personal hygiene (values of $\alpha_j$ and $\beta_j$ reduced by a factor of 0.1). Thus, government's active intervention seems to be more important for Japan. Moreover, for the United States, we see the opposite. Specifically, only about 4-fold reduction in the percentage of individuals eventually infected is observed upon a drastic 10-fold step-up of active intervention, whereas a 6-fold reduction is observed upon a 10-fold improvement in maintaining personal hygiene by the public. Thus, raising the level of public vigilance in exercising personal protective measures is comparatively more important for the United States.

The reason for the difference between Japan and the United States is that the United States has higher infection rates compared to Japan. Reducing $k_j^{(c)}$ for the US case is thus less effective at such high infection rates (ie, a smaller eventual infected number per additional infected individual would not help too much). In contrast, the parameter sets for Japan already have relatively lower infection rates, and further improvement by reducing the infection rates would be limited. As a final remark, combining a very high level of pubic vigilance in exercising strict protective measures and a drastic step-up of government intervention, the percentage of the population getting infected can be reduced to 0.23% in Japan and 2.7% in the United States.

As this study was conducted during the early phase of the pandemic for Japan and the United States, the amount of data used was moderate, though adequate in generating sets of parameter values that fit the data with sufficiently small errors. With more data available, the accuracy of the parameters obtained is expected to improve, and the prediction henceforth would also be more accurate. However, in predicting the pandemic progression, especially during the early phase of an outbreak, we are often confronted with limited data, and the results in this study did demonstrate the application of the proposed model and data fitting method in offering highly consistent prediction of the extent of the pandemic (eg, percentage of the population infected) for Japan and the United States.

Several limitations of the model presented here are worth noting. First, we observed that the actual epidemic trajectory deviates above or below the estimated trajectory due to the varying levels of public health measures applied at particular times, which cause parameters $k_j^{(c)}$, $\alpha_j$, and $\beta_j$ to vary with time. Thus, if a city or region has implemented highly successful public health measures, then the actual values of $k_j^{(c)}$, $\alpha_j$, and $\beta_j$ would be less than their estimated values. The number of confirmed cases would be less than that estimated by the model and vice versa. Furthermore, the number of confirmed cases is highly related to the number of patients who have been tested [23,24]. The value of $\lambda_j$ is thus also time varying as the test capacity varies in time. In our model, we take the parameters as constants for simplicity. Using constant parameters, the model can only give an average profile prediction. Second, expanding the parameter set would improve the ability of the model to isolate the different causes that contribute to the pandemic progression profile. For instance, we may introduce a parameter corresponding to the testing capacity of a city or region instead of integrating it with $\lambda_j$, which may blur the key factor affecting $\lambda_j$. However, with more parameters, the parameter extraction process will become more time-consuming and computationally more intensive. Thus, a right balance should be sought to achieve an adequate coverage of interpretation for physical causes by the parameter set while maintaining a reasonable computational efficiency. Finally, the model has a large set of parameters, and the relative importance of each parameter is not identical [25,26]. A detailed sensitivity analysis can be performed to identify the set of crucial parameters so that resources can be directed to specific kinds

of active measures to slow the pandemic progression more effectively.

## Conclusion

One of the key challenges in data-driven modelling and analysis is the delayed and missing information that makes fitting of models either difficult or unreliable, resulting in inconsistent or even erroneous dynamical profiles generated by a poorly parameterized model. The traditional SEIR model provides a general dynamical description of the disease spread in a population and involves a series of transitional processes that describe how a healthy individual becomes exposed, infected, and eventually recovered or removed from the population. However, the data of infected and recovered cases reported by different cities and regions have been found unreliable or incomplete, as they are subject to the availability of test facilities as well as other factors related to the bureaucracy of reporting and the operation mode of the medical systems. In this paper, we propose a new disease spreading model with consideration of the delayed and missing data of infected cases, intercity travel, and the level of active intervention. The model, which estimates the actual number of infected cases after identifying the best parameter sets, was applied to study the COVID-19 pandemic progression in Japan and the United States. Results reveal that the actual number of infected individuals could be up to 20-fold and 10-fold as many as the confirmed numbers in Japan and the United States, respectively, as of March 19, 2020. Our model also allows assessment of varying levels of active intervention implemented by the government, and the results showed that the current level of control by the Japanese and US governments may be inadequate, and a significant step-up in the level of active intervention is necessary to slow the aggressive progression trend in both countries. For Japan, based on the data collected so far and assuming no further tightening of control, our model estimates about 6.55% of the population eventually infected, and a 4-fold elevation in control efforts may bring it down to 1.54%. For the United States, our model estimates about 18.2% of population will eventually be infected if the government does not step up its control, and a 4-fold elevation in active intervention may bring it down to 9.32%. Finally, adjusting the infection rates permits assessing the effectiveness of practicing protective measures and maintaining personal hygiene. Our results show that stepping up government's active intervention would be more effective for Japan, while raising the level of public vigilance in maintaining personal hygiene and social distancing is comparatively more important for the United States.

## Conflicts of Interest

None declared.

## References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. Coronavirus cases by country, territory, or conveyance. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/
3. Pueyo T. Coronavirus: the hammer and the dance. Medium. 2020 Mar 19. URL: https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56
4. Zhan C, Tse CK, Fu Y, Lai Z, Zhang H. Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data [preprint]. medRxiv 2020 Feb 20. [doi: 10.1101/2020.02.18.20024570]
5. Du Z, Wang L, Cauchemez S, Xu X, Wang X, Cowling BJ, et al. Risk for transportation of coronavirus disease from Wuhan to other cities in China. Emerg Infect Dis 2020 May;26(5):1049-1052. [doi: 10.3201/eid2605.200146] [Medline: 32053479]
6. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
7. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 2020 Jun;582(7812):389-394. [doi: 10.1038/s41586-020-2284-y] [Medline: 32349120]
8. Wu JT, Leung KG, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb;395(10225):689-697. [doi: 10.1016/s0140-6736(20)30260-9]
9. Tian H, Liu Y, Li Y, Wu C, Chen B, Kraemer MUG, et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 2020 May 08;368(6491):638-642 [FREE Full text] [doi: 10.1126/science.abb6105] [Medline: 32234804]

XSL•FO

RenderX

10. Gilbert M, Pullano G, Pinotti F, Valdano E, Poletto C, Boëlle P, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet 2020 Mar;395(10227):871-877. [doi: 10.1016/s0140-6736(20)30411-6]

11. Peto J. COVID-19 mass testing facilities could end the epidemic rapidly. BMJ 2020 Mar 22;368:m1163. [doi: 10.1136/bmj.m1163] [Medline: 32201376]

12. Coronavirus testing: criteria and numbers by country. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/covid-19-testing

13. Diekmann O, Heesterbeek H, Britton T. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton, New Jersey: Princeton University Press; 2013.

14. Small M, Shi P, Tse CK. Plausible models for propagation of the SARS virus. IEICE Trans Fund Electron Comm Computer Sci 2004;87(9):2379-2386.

15. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

16. COVID-19 pandemic in Japan. Wikipedia. 2020 Mar. URL: https://en.wikipedia.org/wiki/COVID-19_pandemic_in_Japan

17. New coronavirus epidemic real-time tracking. Sina News. 2020 Mar. URL: https://news.sina.cn/zt/_d/yiqing0121

18. Japan Government e-Stat. URL: https://www.e-stat.go.jp/dbview?sid=0003280293 [accessed 2020-03-31]

19. State-to-state migration flows. United States Census Bureau. 2020 Mar. URL: https://www.census.gov/data/tables/time-series/demo/geographic-mobility/state-to-state-migration.html

20. Bureau of Transportation Statistics. URL: https://www.transtats.bts.gov/ [accessed 2020-03-31]

21. Coronavirus disease (COVID-19) advice for the public. World Health Organization. 2020 Mar. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

22. Milton DK, Fabian MP, Cowling BJ, Grantham ML, McDevitt J. Influenza virus aerosols in human exhaled breath: particle size, culturability, and effect of surgical masks. PLoS Pathog 2013 Mar;9(3):e1003205 [FREE Full text] [doi: 10.1371/journal.ppat.1003205] [Medline: 23505369]

23. Cherney K. Does wearing a mask protect you from the flu and other viruses? Healthline. 2020 Mar 18. URL: https://www.healthline.com/health/cold-flu/mask

24. Global leaders unite to ensure everyone everywhere can access new vaccines tests and treatment. World Health Organization. 2020 Apr 24. URL: https://www.who.int/news-room/detail/24-04-2020-global-leaders-unite-to-ensure-everyone-everywhere-can-access-new-vaccines-tests-and-treatments-for-covid-19

25. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 2009 Aug 01;25(15):1923-1929. [doi: 10.1093/bioinformatics/btp358] [Medline: 19505944]

26. Zhan C, Li BL, Yeung LF. Structural and practical identifiability analysis of S-system. IET Syst Biol 2015 Dec;9(6):285-293. [doi: 10.1049/iet-syb.2015.0014] [Medline: 26577163]

## Abbreviations

**C:** confirmed
**COVID-19:** coronavirus disease
**E:** exposed
**I:** infected
**R:** recovered or removed
**S:** susceptible
**SEICR:** susceptible-exposed-infected-confirmed-removed
**SEIR:** susceptible-exposed-infectious-recovered

XSL•FO
RenderX

XSL•FO

**RenderX**