# Comparative Study of COVID-19 Pandemic Progressions in 175 Regions in Australia, Canada, Italy, Japan, Spain, UK and USA Using a Novel Model That Considers Testing Capacity and Deficiency in Confirming Infected Cases

Choujun Zhan, *Member, IEEE*, Chi K. Tse, *Fellow, IEEE*, Ying Gao, and Tianyong Hao, *Senior Member, IEEE*

*Abstract*—Not identified as being exposed or infected, the group of asymptomatic and presymptomatic patients has become the key source of infectious hosts for the COVID-19 pandemic, triggering the re-emergence of outbreaks. Acknowledging the impacts of movement of unidentified patients and the limited testing capacity on understanding the spread of the virus, an augmented Susceptible-Exposed-Infectious-Confirmed-Recovered (SEICR) model integrating intercity migration data and testing capacity is developed to probe into the number of unidentified COVID-19 infected patients. This model allows evaluation of the effectiveness of active interventions, and more accurate prediction of the pandemic progression in a country, region or city. A pseudo-coevolutionary algorithm is adopted in the model fitting to provide an effective estimation of high-dimensional unknown parameter sets using a limited amount of historical data. The model is applied to 175 regions in Australia, Canada, Italy, Japan, Spain, the UK and USA to estimate the number of unconfirmed cases using limited historical data. Results showed that the actual number of infected cases could be 4.309 times as many as the official confirmed number. By implementing mass COVID-19 testing, the number of infected cases could be reduced by about 50%.

*Index Terms*—COVID-19 pandemic, model, human migration, test capacity.

## I. INTRODUCTION

THE Novel Coronavirus Disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1] which has a long evolutionary history in bats [2], began to spread in China since December 2019 and was found to be highly contagious with a reproductive number as high as 6.49 [3]. Studies have revealed the mean incubation period of COVID-19 being 5 to 14 days [4], with high transmissibility of the virus from asymptomatic patients [5], [6], [7] as well as possible transmission from presymptomatic patients [8], [9]. While infected individuals who have been tested positive and confirmed of being infected are mostly isolated [10], presymptomatic and asymptomatic patients have formed a large group of individuals who are mobile in the community and able to infect others [11]. In addition, some unidentified infected patients are still travelling [12], [13], [14] and continue to spread the virus through the transportation networks, restaurants, hotels, business venues, and other venues where they carry out their usual activities and interact with people [15], [16], [17]. Thus, the group of presymptomatic and asymptomatic patients have become the key source of infectious hosts that may trigger new waves of outbreaks. Up to the end of July 2020, the COVID-19 pandemic has affected 213 countries and caused a total of 17,187,414 infections, with a death toll of 670,202 [18], and second and third waves of outbreaks have begun to emerge in countries or cities where the pandemic had apparently come under control.

The number of confirmed cases is known to represent only a fraction of the infected patients, and any assessment of the extent of an outbreak in a country or city can only be as accurate as the truthfulness of information obtained in relation to the number of exposed, infected and recovered individuals in that country or city. Thus, depending on the number of tests the authorities manage to administer, there will be considerable discrepancy between the number of actual infections and the official confirmed cases [19]. Epidemiological studies indicated that 40% to 70% of the population could become infected unless preventive measures are taken, which include early and prompt testing to identify the infected population and prevent continuous onward transmission of the virus [20]. Healthcare professionals around the globe have advocated for improving testing capacity for SARS-CoV-2 to ascertain infection numbers, estimate reproductive numbers, and evaluate epidemiological risks [21]. More and faster tests generally improve the probability of identifying infected but unconfirmed cases, which can help isolate "invisible" COVID-

19 patients and disconnect transmission chains effectively. Different countries are at different stages of pandemic progression. Gearing up for mass testing is the only way to assess the scope of infection and the phase of the progression trajectory in order to inform policy makers to formulate and implement effective public health responses. It is thus important to develop disease transmission models that are able to estimate the number of infected but unconfirmed individuals in a certain population to help contain the spread of the virus [22].

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 this work, the basic model proposed is a Susceptible-Exposed-Infectious-Confirmed-Recovered (SEICR) model, which has an additional state corresponding to an individual having been confirmed by the authority as being infected [22]. In addition, the SEICR model, integrated with the daily infection data, intercity migration data and test data, can estimate the size of the unidentified infected population. Also, a specific parameter is included to adjust the level of active intervention in the simulation of 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. The proposed model can evaluate the effect of active control measures, and can provide realistic prediction for the spread of COVID-19 in a country or city under different scenarios. The resulting augmented SEICR model is a high-dimension nonlinear dynamic model. In this study, a pseudo-coevolutionary algorithm is adopted for model fitting to estimate the set of unknown parameters, which can then be used for generating future progression profiles.

The migration-and-test-capacity augmented SEICR model is applied to study the dynamics of the spread of COVID-19 in Australia, Canada, Italy, Japan, Spain, UK and USA. Daily COVID-19 data, intercity migration data, and COVID-19 test data for 175 states or regions (prefectures in the case of Japan) have been collected from official sources [23], [24], [25], [26], [27]. By fitting the model with data of the COVID-19 infected cases for Australia (January 30 to June 16, 2020), Canada (January 31 to June 17, 2020), Italy (February 4 to June 17, 2020), Japan (January 31 to May 30, 2020), Spain (February 6 to July 5, 2020), the UK (January 30 to June 7, 2020) and the USA (January 29 to June 7, 2020), the corresponding sets of model parameters are computed and then used to estimate the number of infected but unconfirmed cases and evaluate the pandemic progression in the 175 states or regions. Our results revealed that a large percentage of infected patients had not been confirmed, specifically, 75.19% for the USA, 81.48% for Japan, 71.24% for Canada, 78.75% for Australia, 75.36% for Italy, 42.64% for Spain, and 85.17% for the UK. The actual total number of infected individuals could be 4.309 times the official confirmed number. Results also showed that by increasing COVID-19 testing capacity, the number of infected cases could be reduced by half, and
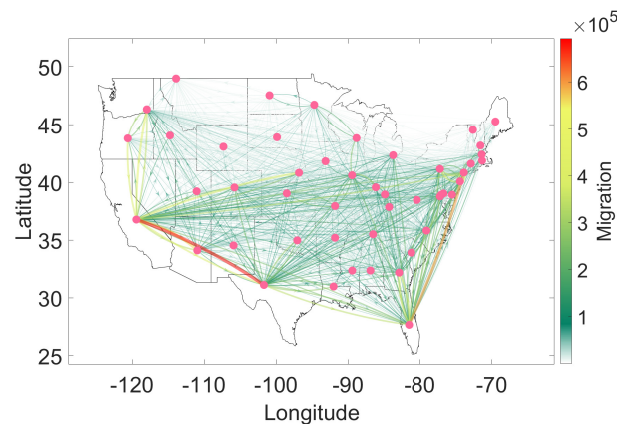


Fig. 1. Interstate travel network of 52 regions in USA. Color and width of lines indicates migration strength (color online), while arrows shows direction.

specifically, the total number of infected cases would reduce by 4.65% for Australia, 58.63% for Canada, 37.51% for Italy, 52.81% for Japan, 10.88% for Spain, 52.15% for the UK and 57.03% for the USA.

## II. DATA

The World Health Organization (WHO) announced the COVID-19 outbreak a pandemic on March 12, 2020 and set COVID-19 to the highest alert level [28]. Many government facilities release the pandemic data of COVID-19 to the public [29]. In this study, we have investigated the spread of COVID-19 based on epidemic data, migration data and COVID-19 tests data of 175 states or regions in seven countries.

### A. COVID-19 Data

The available datasets include the cumulative number of confirmed cases, recovered cases, and death tolls, for 8 states in Australia from January 24, 2020; 13 states in Canada from January 25, 2020; 22 regions in Italy from January 30, 2020; 47 prefectures in Japan from January 22, 2020; 19 regions in Spain from February 1, 2020; 14 regions in the UK from January 30, 2020; and 52 regions in the USA (50 states, a federal district and Puerto Rico) from January 22, 2020. In general, on day $t$, the official pandemic data released by region $j$ includes the number of cumulative confirmed cases $C_j(t)$, cumulative recovered cases $R_{j,c}(t)$, and confirmed death cases $D_{j,c}(t)$, where subscript $c$ means "confirmed". However, some authorities provide detailed data, while others only release crude information. In these seven countries, each state or region has generally released the detailed data of confirmed cases, recovered cases and death tolls. However, prefectures in Japan have only provided confirmed cases and death tolls. For Canada, some states have released recovered data. For the UK, no recovered data were available from January 22 to May 17, 2020, and no death tolls were available after May 17, 2020. Table I summarizes the available information.

### B. Migration Data

We have collected intercity travel data of the seven countries mentioned in Table I. The data contains the indicative number

TABLE I
SUMMARY OF COVID-19 DATA OF 175 REGIONS UNDER STUDY.

| Country | Number of states or regions (prefectures for Japan) | Confirmed cases | Recovered cases | Death toll | Start date |
|---|---|---|---|---|---|
| Australia | 8 | Yes | Yes | Yes | Jan 24, 2020 |
| Canada | 13 | Yes | Part | Yes | Jan 25, 2020 |
| Italy | 22 | Yes | Yes | Yes | Jan 30, 2020 |
| Japan | 47 | Yes | No | Yes | Jan 22, 2020 |
| Spain | 19 | Yes | Yes | Yes | Feb 01, 2020 |
| UK | 14 | Yes | Part | Part | Jan 30, 2020 |
| USA | 52 | Yes | Yes | Yes | Jan 22, 2020 |

of travellers from one city to another within a country [23], [24], [25], [26], [27]. Note that these migration data do not include international, intracity or intraregion travel. Based on the intercity migration data, we can construct the *migration matrix*, which is given as

$$
M(t) = \begin{bmatrix}
m_{11}(t) & m_{12}(t) & \cdots & m_{1N}(t) \\
m_{21}(t) & m_{22}(t) & \cdots & m_{2N}(t) \\
\vdots & \vdots & \ddots & \vdots \\
m_{N1}(t) & m_{N2}(t) & \cdots & m_{NN}(t)
\end{bmatrix}, \quad (1)
$$

where $N$ is the number of regions in a country, and $m_{ij}(t)$ is the migrant volume from region $i$ to region $j$ at time $t$. There exists $m_{ij} \geq 0$ ($i \neq j$) and $m_{ii} = 0$. Migration matrix $M$ thus effectively describes the network of regions with human movement constituting the links of the network. Fig. 1 shows migration network of the USA, where the width of edge indicates the volume of migration population and arrows stands for direction. Hence, $\sum_{i=1}^{N} m_{ij}(t)$ represents the volume of population moving in region $j$ from other regions, while $\sum_{i=1}^{N} m_{ji}(t)$ stands for the volume of individuals moving out of region $j$ to other regions. Then, the dynamic change of population in region $j$ is

$$
\Delta P_j(t) = \sum_{i=1}^{N} m_{ij}(t) - \sum_{i=1}^{N} m_{ji}(t). \quad (2)
$$

It is worth noting that the contribution of human migration to the pandemic progression was less significant after travel restrictions and local area lockdowns were implemented, though it played a crucial role in the early spread of the epidemic in China from late December 2019 to mid January 2020 [17].

### C. COVID-19 Tests Data

Data related to testing capacities of the seven countries have been collected since February 2020, and graphically shown in Fig. 2. In particular, Fig. 2(b) shows the daily number of tests performed per thousand people (i.e., $\frac{N_T(t)}{N_p} \times 1000$) in each country. The periodic spikes displayed in these data can be attributed to limited test availability during weekends and holidays, and the possible cumulative reporting due to administrative delays. Let $N_T(t)$ be the daily number of people tested in a region on day $t$, and $N_p$ be the size of population. From the data collected, the USA has the highest total testing volume, which is 72 times higher than Japan's. We see that the



(a) Daily tests performed



(b) Daily tests per 1000 people
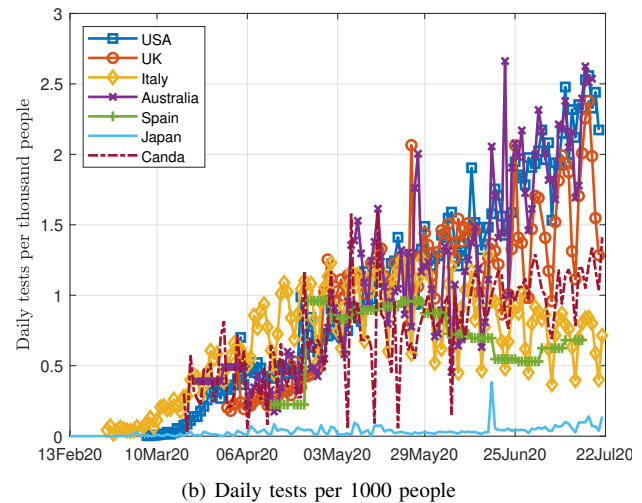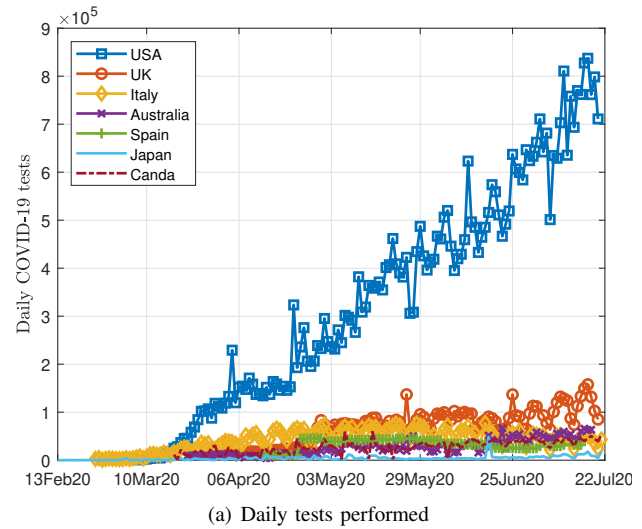
Fig. 2. COVID-19 testing capacity. (a) Number of daily COVID-19 tests; (b) number of daily COVID-19 tests per 1,000 people.

USA, UK and Australia have performed more than 2 tests per thousand people each day until July 25, 2020, whereas Canada, Spain and Italy have conducted about 1 test per thousand people. Japan has the lowest testing capacity among the seven countries under study, and has performed less than 0.1 test per thousand people. Table II summarizes the testing capacities of the seven countries.

TABLE II
COVID-19 TEST DATA IN SEVEN COUNTRIES UNDER STUDY UNTIL JULY 22, 2020.

| Country | Total COVID-19 tests | Percentage of population tested | Confirmed cases | Percentage of positive cases per 100 tests |
|---|---|---|---|---|
| Australia | 3,280,519 | 13.1273% | 13,595 | 0.4144% |
| Canada | 3,448,189 | 9.1732% | 114,385 | 3.3172% |
| Italy | 6,248,997 | 10.3529% | 245,338 | 3.9260% |
| Japan | 651,685 | 0.5152% | 28,963 | 4.4443% |
| Spain | 2,943,846 | 6.2715% | 267,551 | 9.0885% |
| UK | 7,785,664 | 11.7484% | 298,731 | 3.8369% |
| USA | 46,472,212 | 14.2030% | 4,038,596 | 8.6903% |
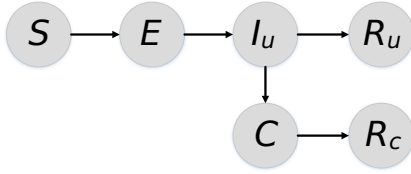


Fig. 3. State transition flow chart.

## III. METHOD

### A. Model

In the traditional Susceptible-Exposed-Infectious-Recovered (SEIR) model, each individual would assume one of four possible states: susceptible ($S$), exposed ($E$), infected ($I$), and recovered/removed ($R$) [30], [31]. In reality, however, an unconfirmed COVID-19 patient (in state $I_u$) would transit into the confirmed state only if he or she is tested positive. However, not all COVID-19 patients have a chance to be tested, and thus not all infected individuals will become confirmed. Hence, some of the COVID-19 patients would transit to the recovered (removed) state without going through the confirmed state. Based on the traditional SEIR model, we introduce a "confirmed state" $C$, corresponding to an individual who has been tested positive, giving a new SEICR model [22]. Moreover, acknowledging the influences of intercity migration and testing capacity of each region, we propose here a new augmented SEICR model, in which any individual would assume one of six states at any time, i.e., susceptible ($S$), exposed ($E$), unconfirmed infections ($I_u$), confirmed ($C$), confirmed recovered/removed ($R_c$), and unconfirmed recovered/removed ($R_u$). We also assume no reinfection occurs to the recovered patients. For region $j$ at time $t$, the number of individuals in each state is denoted by $S_j(t)$, $E_j(t)$, $I_{j,u}(t)$, $C_j(t)$, $R_{j,c}(t)$ and $R_{j,u}(t)$. The state transitions in this model are illustrated in Figure 3. The cumulative number of confirmed cases is $C(t) + R_c(t)$, and the cumulative number of unconfirmed cases is $I_u(t) + R_u(t)$.

*1) Intercity Migration:* Suppose region $j$ has a population $P_j(t)$. Let $m_{ij}(t)$ represent the number of people moving from region $i$ to region $j$ at time $t$. We assume that susceptible individuals are homogeneously mixed in the group of migrating individuals. Then, on day $t$, the number of susceptible individuals moving from region $i$ to region $j$ is

$$\Delta S_j^{\text{in}}(t) = \sum_{i=1}^{N} \left( \frac{S_i(t) m_{ij}(t)}{P_i(t)} \right). \tag{3}$$

Also, if the number of individuals migrating out of region $j$ is $\sum_{i=1}^{N} m_{ji}(t)$, the number of susceptible individuals moving out of region $j$ is

$$\Delta S_j^{\text{out}}(t) = S_j(t) \frac{\sum_{i=1}^{N} m_{ji}(t)}{P_j(t)}. \tag{4}$$

Similarly, groups of exposed and unconfirmed individuals follow the same migration rules. In addition, all confirmed individuals are assumed to stay in region $j$, i.e., confirmed cases are isolated.

*2) Testing Capacity:* In most countries, COVID-19 tests are provided mainly for individuals who have developed symptoms or have been in close contact with infected individuals [32], [33]. As the testing capacity is limited, only a portion of the unconfirmed cases will be confirmed within a short period of time. Let $\lambda_j(N_{T,j}(t))$ denote the confirmed rate of infected individuals in region $j$, which is a function of $N_{T,j}(t)$ (the number of tests performed on day $t$ in region $j$). Let $\delta C_j(t)$ represent the number of newly confirmed cases on day $t$ (note that $\delta C_j(t) \neq C_j(t+1) - C_j(t)$). Suppose, on day $t$, the number of active unconfirmed cases is $I_u(t)$. Obviously, larger $I_u(t)$ means higher probability of confirming a patient as infected by performing a test. Hence, in region $j$, the number of active unconfirmed cases $I_{u,j}(t)$ is proportional to $\delta C_j(t)$. Furthermore, a larger $N_{T,j}(t)$ means that a larger group of individuals with symptoms will be tested, resulting in more patients to be identified. Thus, we have, for region $j$,

$$\delta C_j(t) = I_{u,j}(t) \lambda_j(N_{T,j}(t)), \tag{5}$$

where $\lambda_j(N_{T,j}(t))$ is a monotonic increasing function which may be conveniently approximated by a simple linear relationship, i.e.,

$$\lambda_j(N_{T,j}(t)) \approx \lambda_{0,j} + k_\lambda N_{T,j}(t), \tag{6}$$

with $\lambda_{0,j}$ and $k_{\lambda,j}$ being constant parameters for region $j$. Note that the number of newly confirmed cases must not exceed the number of active unconfirmed cases, i.e., $\lambda_j(N_{T,j}(t)) \leq 1$. Hence, there exists an upper threshold $N_{T,\text{thr}}$ of testing capacity which guarantees all of the unconfirmed cases be identified (confirmed). We may therefore write the rate of confirmation as

$$\lambda_j(N_{T,j}(t)) = \begin{cases} \lambda_{0,j} + k_{\lambda,j} N_{T,j}(t), & \text{if } N_{T,j}(t) \leq N_{T,\text{thr}} \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

where $N_{T,\text{thr}} = (1 - \lambda_{j,0})/k_{j,\lambda}$. Then, the number of newly confirmed COVID-19 patients in region $j$ on day $t$ is

$$\delta C_j(t) = \begin{cases} I_{j,u}(t)(\lambda_{j,0} + k_{j,\lambda} N_{T,j}(t)), & \text{if } N_T(t) \leq N_{T,\text{thr}} \\ I_{j,u}(t), & \text{otherwise} \end{cases} \tag{8}$$
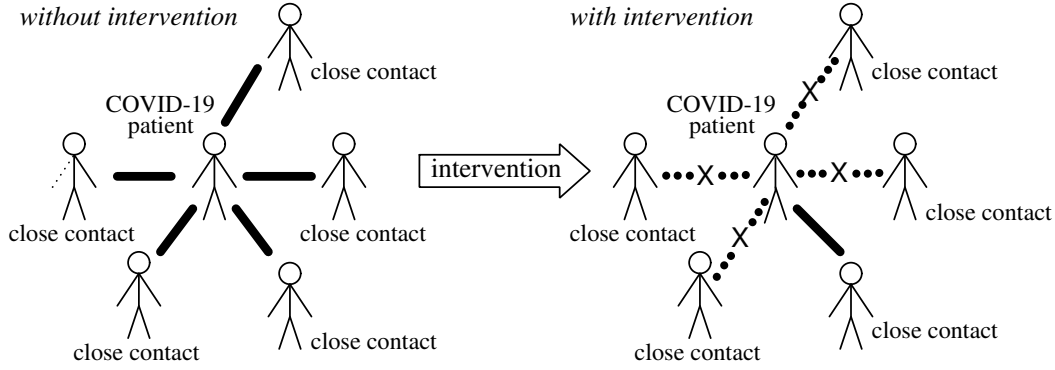
Fig. 4. Illustrative examples of epidemic spreading in a contact network without intervention (left), where a COVID-19 patient can freely contact with $k^{(c)} = 5$ individuals; and with intervention (right) such as quarantine and social distancing, where a COVID-19 patient can hardly contact with others. In the latter case, most of the transmission paths are disconnected and only $k^{(c)} = 1$ could be infected.

TABLE III
PARAMETERS OF THE NEW AUGMENTED SEICR 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 $j$ |
| $\gamma_j^{(u)}$ | recovery (removal) rate of an infected but not confirmed individual in region $j$ |
| $\gamma_j^{(c)}$ | recovery (removal) rate of a confirmed individual in region $j$ |
| $\kappa$ | rate of an exposed individual becoming infected |
| $\lambda_j(N_{T,j}(t))$ | $\lambda_j((N_{T,j}t))) = \lambda_{0,j} + k_{\lambda,j}N_{T,j}(t)$ represents confirmation rate of infected individuals in region $j$ |
| $N_{T,j}(t)$ | daily test capacity of 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 $j$ |
| $k_h$ | proportion of population infected achieving no further spreading, i.e., absolute upper bound for $N_j^s$ for all $j$ |
| $\delta_j$ | initial percentage of susceptible, infected, and recovered individuals in region $j$ |
| $I_{j,u,0}$ | initial number of infected but not confirmed 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 $j$ |

*3) Active Intervention:* The role of active intervention is to limit or remove the contact paths connecting an infected individual with the group of susceptible individuals whom the infected individual may be in close contact with. The aim of implementing active intervention is thus to lower the number of effective contact paths, which can be represented quantitatively by $k^{(c)}$. An illustration is given in Fig. 4, where active intervention reduces $k^{(c)}$ from 5 to 1. It can be readily conceived that when more people are infected, the number of effective contact paths would increase, and so would the size of the susceptible group. In quantitative terms, 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 to $\Delta S_i(t)$ and $\Delta N_j^s(t)$ in

the model. This extra intervention term, dependent upon $k^{(c)}$, can be derived as follows.

Suppose, in region $j$, the average number of close contacts $k_j^{(c)}$ of each COVID-19 patient is highly influenced by the containment strategy implemented by the authority. We assume that each confirmed case will be isolated and cease to infect others. Thus, the intervention term should take the following form:

$$k_j^{(c)}\left(\Delta I_{u,j}(t) + \Delta E_j(t)\right) \qquad (9)$$

where $k_j^{(c)}$ corresponds quantitatively to an increase in the number of eventual infected individuals for each additional infected or exposed individual in region $j$. Note that if strict control measures are adopted, $k_j^{(c)}(t)$ will become very small, implying that most contact paths have been disconnected and any newly infected or exposed individual would unlikely infect others. Thus, by setting $k_j^{(c)}(t)$, the effect of active intervention can be incorporated in the model.

*4) Augmented SEICR Model:* Incorporating the intercity migration, testing capacity and active intervention terms, the final form of the augmented SEICR model is given as follows:

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

$$\Delta I_{j,u}(t) = (1 - \lambda_j(N_{T,j}(t)) - \gamma_j^{(u)})I_{j,u}(t) + \kappa E_i(t)$$
$$+ k_I\left(\sum_{i=1}^{N}\left(\frac{I_{u,i}(t)m_{ij}(t)}{P_i(t)}\right) - I_{j,u}(t)\frac{\sum_{i=1}^{N}m_{ji}(t)}{P_j(t)}\right)$$
$$\Delta E_j(t) = \frac{\beta_j I_{j,u}(t) + \rho_j C_j(t)}{N_j^s(t)}S_j(t) + \frac{\alpha_j E_j(t)}{N_j^s(t)}S_j(t)$$
$$- \kappa E_i(t) + \sum_{i=1}^{N}\left(\frac{E_i(t)m_{ij}(t)}{P_i(t)}\right)$$
$$\qquad (10)$$

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

$$\Delta C_j(t) = \lambda_j(N_{T,j}(t))I_{j,u}(t) - \gamma_j^{(c)}C_j(t)$$

$$\Delta R_{j,u}(t) = \gamma_j^{(u)}I_{j,u}(t)$$

$$\Delta R_{j,c}(t) = \gamma_j^{(c)}C_j(t)$$

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

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

where $\Delta I_{j,u}(t) = I_{j,u}(t+1) - I_{j,u}(t)$, $\Delta E_j(t) = E_j(t+1) - E_j(t)$, $\Delta S_j(t) = S_j(t+1) - S_j(t)$, $\Delta C_j(t) = C_j(t+1) - C_j(t)$, $\Delta R_{j,u}(t) = R_{j,u}(t+1) - R_{j,u}(t)$, $\Delta R_{j,c}(t) = R_{j,c}(t+1) - R_{j,c}(t)$, $\Delta N_j^s(t) = N_j^s(t+1) - N_j^s(t)$, and $\Delta P_j(t) = P_j(t+1) - P_j(t)$. The physical meanings of all parameters are given in Table III. In particular, the following parameters, which play special epidemiological or control roles, are worthnoting:

- $\alpha_j$ represents the rate of infecting a susceptible individual by an exposed individual in region $j$. Note that the infection rate is influenced by local factors and control measures, such as lockdown, social distancing, contact tracing and quarantine. Hence, $\alpha$ actually varies from region to region and is time-varying. However, in a short period of time, control measures may not change dramatically. Thus, we can assume that $\alpha_j$ is constant for region $j$. Similarly, $\{\beta_j, \rho_j, \gamma_j^{(u)}, \gamma_j^{(Q)}\}$ can be assumed constant.
- $\kappa$ represents the rate of an exposed individual transiting into an infected individual. This transition rate is a natural property of COVID-19, which can be considered as a constant rate irrespective of local factors and control measures.
- $k_I$ represents the possibility of an unconfirmed infected individual moving from one region to another. It can be assumed constant when migration control is in place.
- $k_j^{(c)}$ represents the effectiveness of the implemented active intervention. Specifically, $k_j^c < 1$ corresponds to effective control measures being applied and the spread is being slowed down, whereas $k_j^c > 1$ corresponds to less effective control and the spread of the disease being severe. For each additional infected individual, there will be $k_j^{(c)}$ more eventual infected individuals.

*Remarks:* The model proposed here highlights the crucial role test capacities play in the progression of the pandemic.

The augmented SEICR model differs from most existing models (standard SEIR model and its variants) [34], [35], [36], [37] in its ability to identify unconfirmed infected cases. The parameter identification, as will be explained in the next subsection, will involve testing data and hence differs significantly from those used for most existing models. Results generated from our model will reflect the influence of test capacities, which is not provided in other models.

### B. Parameter Identification

The model represented by (10) has a large number of unknown parameters to be estimated from historical data through fitting. Specifically, for region $j$, the set of unknown parameters are

$$\theta_j = \{\alpha_j, \beta_j, \rho_j, \lambda_{j,0}, k_{j,\lambda}, \gamma_j^{(u)}, \gamma_j^{(Q)}, k_j^{(c)}\}. \quad (11)$$

The procedure for parameter estimation involves minimizing the distance between generated growth trajectories and corresponding historical trajectories. The parameter estimation problem can be formulated as an optimization problem. The initial values of susceptible, exposed and unconfirmed cases are unknown and to be estimated, while the initial numbers of confirmed and removed cases are equal to the official initial numbers. Here, we define the initial number of susceptible, exposed, infected, and confirmed individuals in region $j$ as

$$X_{j,0} = \{S_{j,0}, E_{j,0}, I_{j,u,0}\}. \quad (12)$$

Note that $N_j(t_0) = S_{j,0} + E_{j,0} + I_{j,u,0} + C_{j,0} + R_{j,c,0} + R_{j,u,0}$ and $\delta_j = N_j(t_0)/P_j(t_0)$. Then, $\delta_j$ would not be a parameter to be estimated. Hence, the unknown parameter set is $\Theta = \{X_{1,0}, X_{2,0}, \cdots, X_{N,0}, \theta_1, \theta_2, \cdots, \theta_N, \kappa, k_I\}$. The total number of parameters is thus $11N+2$, where $N$ is the number of regions.

The SEICR model has a set of extended state variables, i.e.,

$$X_j(t) = [S_j^T(t), E_j^T(t), I_{j,u}^T(t), C_j(t), R_{j,c}^T(t),$$
$$R_{j,u}^T(t), P_j^T(t), (N_j^s(t))^T]^T. \quad (13)$$

Overall, for a total of $N$ regions, we have

$$X(t) = [X_1^T(t)\ X_2^T(t)\ \cdots\ X_N^T(t)]^T. \quad (14)$$

Model (10) can be written as

$$\Delta X_j(t_i) = f(X_j(t_i), X(t_i), \Theta), \quad (15)$$

where $f(x)$ is the right side of (10), and $\Theta$ is the set of unknown parameters. Writing $\Delta X_j(t) = X(t_i + 1) - X(t_i)$, we have the following model in discrete-time form:

$$X_j(t_i + 1) = X_j(t_i) + f(X_j(t_i), X(t), \Theta). \quad (16)$$

Then, the parameter estimation problem can be formulated as the following constrained nonlinear optimization problem (CNOP) [38]:

$$\mathrm{P_0:} \min_{\Theta} \sum_{j=1}^{N} \sum_{i=1}^{K} w_{ij}^{(c)}\left(C_j(t_i) - \hat{C}_j(t_i|\Theta)\right)^2$$
$$+ w_{ij}^{(r)}\left(R_{j,c}(t_i) - \hat{R}_{j,c}(t_i|\Theta)\right)^2$$

---

**Algorithm 1** Algorithm for calculating time series $C_j(t_i)$ and $R_{j,c}(t_i)$

---

**Input:** Parameter set and initial number of infected and exposed individuals of each region $\Theta = \{X_{1,0}, X_{2,0}, \cdots, X_{N,0}, \theta_1, \theta_2, \cdots, \theta_N, \kappa, k_I\}$, where $\theta_j$ and $X_{j,0}$ are defined in equations (11) and (12).

**Output:** Time series $\{C_j(t_0), C_j(t_1), \cdots, C_j(t_K)\}$ and $\{R_{j,c}(t_0), R_{j,c}(t_1), \cdots, R_{j,c}(t_K)\}$.

Initialization:
Set $S_j(t_0) = S_{j,0}$, $E_j(t_0) = E_{j,0}$, and $I_{j,u}(t_0) = I_{j,u,0}$, while $R_j(t_0)$ and $C_j(t_0)$ equal the official initial numbers; $N_j(t_0) = S_{j,0} + E_{j,0} + I_{j,u,0} + C_{j,0} + R_{j,0}$; $X(t_0) = [X_1^T(t_0), X_2^T(t_0), \cdots, X_K^T(t_0)]^T$ and $X_j(t_0) = [S_j^T(t_0), E_j^T(t_0), I_{j,u}^T(t_0), R_{j,u}^T(t_0), R_{j,c}^T(t_0), P_j^T(t_0), (N_j^s(t_0))^T]^T$.

LOOP Process:
**for** $(i = 0$ to $N - 1)$ **do**
  Calculate $X(t_i + 1)$ from model (10), namely,

$$X_j(t_i + 1) = X_j(t_i) + f(X_j(t_i), X(t), \Theta).$$

  Since $C_j(t_i), R_{j,c} \in X_j(t_i)$ and $X_j(t_i) \in X(t_i)$, find $C_j(t_i)$ and $R_{j,c}$ from $X(t_i)$.
**end for**
**return** $\{C_j(t_0), C_j(t_1), \cdots, C_j(t_N)\}$ and $\{R_{j,c}(t_0), R_{j,c}(t_1), \cdots, R_{j,c}(t_N)\}$

---

$$\text{s.t.} \begin{cases} \text{(i)} & X_j(t_i + 1) = X_j(t_i) + f(X_j(t_i), X(t_i), \Theta) \\ \text{(ii)} & \Theta_L \leq \Theta \leq \Theta_U \\ \text{(iii)} & j = 1, 2, \cdots, K \end{cases}$$

$$(17)$$

where $\hat{C}_j(t_i | \Theta)$ and $\hat{R}_{j,c}(t_i | \Theta)$ represent the estimated number of confirmed cases and confirmed removed cases at time $t_i$ with parameter set $\Theta$ and initial condition $X_0$; $w_{ij}^{(c)}$ and $w_{ij}^{(u)}$ stand for the weighted coefficients, which are constant. The unknown parameter set is bounded between $\Theta_L$ and $\Theta_U$. In this work, an inverse approach is taken to find the unknown parameters and states by solving (17). Moreover, the values of the confirmed and recovered cases can be derived directly from the available data, as outlined in Algorithm 1.

With $11N + 2$ parameters to be found, the optimization procedure associated with parameter estimation is computationally intensive. In addition, the search space in the optimization procedure may be highly nonlinear and may contain many local minima. Evolutionary algorithms have been extensively used in solving such nonlinear optimization problems [39], [40]. In this work, a new pseudo-coevolutionary algorithm is adopted to solve this problem, as outlined in Algorithm 2. Specifically, the procedure involves two co-adaptive simulated-annealing-based optimization processes:

1) In the main procedure, we tune all the $11N + 2$ parameters. Then, the Root Mean Square Percentage Error (RMSPE$_j$) is utilized to measure the difference between the real values of confirmed cases and the estimated values generated by the model with an optimal parameter

---

**Algorithm 2** Psuedo-coevolutionary algorithm for estimating the parameter set $\Theta^*$

---

**Input:** The set of unknown parameters and initial number of infected and exposed individuals of each region $\Theta = \{X_{1,0}, X_{2,0}, \cdots, X_{N,0}, \theta_1, \theta_2, \cdots, \theta_N, \kappa, k_I\}$, where $\theta_j = \{\alpha_j, \beta_j, \rho_j, \lambda_{j,0}, k_{j,\lambda}, \gamma_j^{(u)}, \gamma_j^{(c)}, k_j^{(c)}\}$ and $j = 1, 2, \cdots N$.

**Output:** Optimal parameter set $\Theta^*$.

Initialization:
Initialize temperature $T$, and take a random starting point.

$$\Theta_0 = \Theta_L + k_{rand} * (\Theta_U - \Theta_L)$$

The index of adopted region is $\Phi = 1, 2, \cdots, K$.
Apply evolutionary algorithm to optimize parameter set to achieve $\bar{\Theta}^*$, i.e., $\Theta_0 \leftarrow \bar{\Theta}^*$

LOOP Process:
**for** $i = 0$ to $M$ **do**
  **for** $j = 1$ to $N$ **do**
    Set the model parameter as $\Theta_0$ and apply Algorithm 1 to derive $\{\hat{C}(t_0 | \Theta_0), \hat{C}(t_1 | \Theta_0), \cdots, \hat{C}(t_K | \Theta_0)\}$.
    Using evaluation criterion (18) to find RMSPE$_j$ for each region.
  **end for**
  Find the index of the $m$ largest RMSPE$_j$ and set $\Phi = \{a_1, a_2, \cdots, a_m\}$,
  where $a_i$ represents the index of the $i$th largest RMSPE$_j$.
  Apply evolutionary algorithm to optimize parameter set to achieve $\bar{\Theta}^*$, i.e., $\Theta_0 \leftarrow \bar{\Theta}^*$.
**return** Optimal parameter set $\Theta^*$.

---

set $\Theta$, i.e.,

$$\text{RMSPE}_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( \frac{\hat{C}_j(t_i | \Theta) - C_j(t_i)}{C_j(t_i)} \right)^2}. \quad (18)$$

2) In the sub-procedure, we identify the regions (values of $j$) corresponding to the $m$ largest RMSPE$_j$, and tune the parameters for these regions.

It is worth noting that the model, in the form of a set of differential equations, is expected to generate solutions that are the progression profiles. Here, we actually used real data as the solutions to be matched by the model, and the numerical optimization procedure adopted will guarantee that the model does give a solution which resembles the real data when all parameter values are correctly found. Thus, when we get the candidate parameters, it is automatically guaranteed that the parameter values can indeed produce the same data within a small tolerance. In other words, our approach is founded on a rigorous data fitting procedure that generates the parameter values, and we did not "estimate" the parameter values based on some vague rationale or heuristic method.
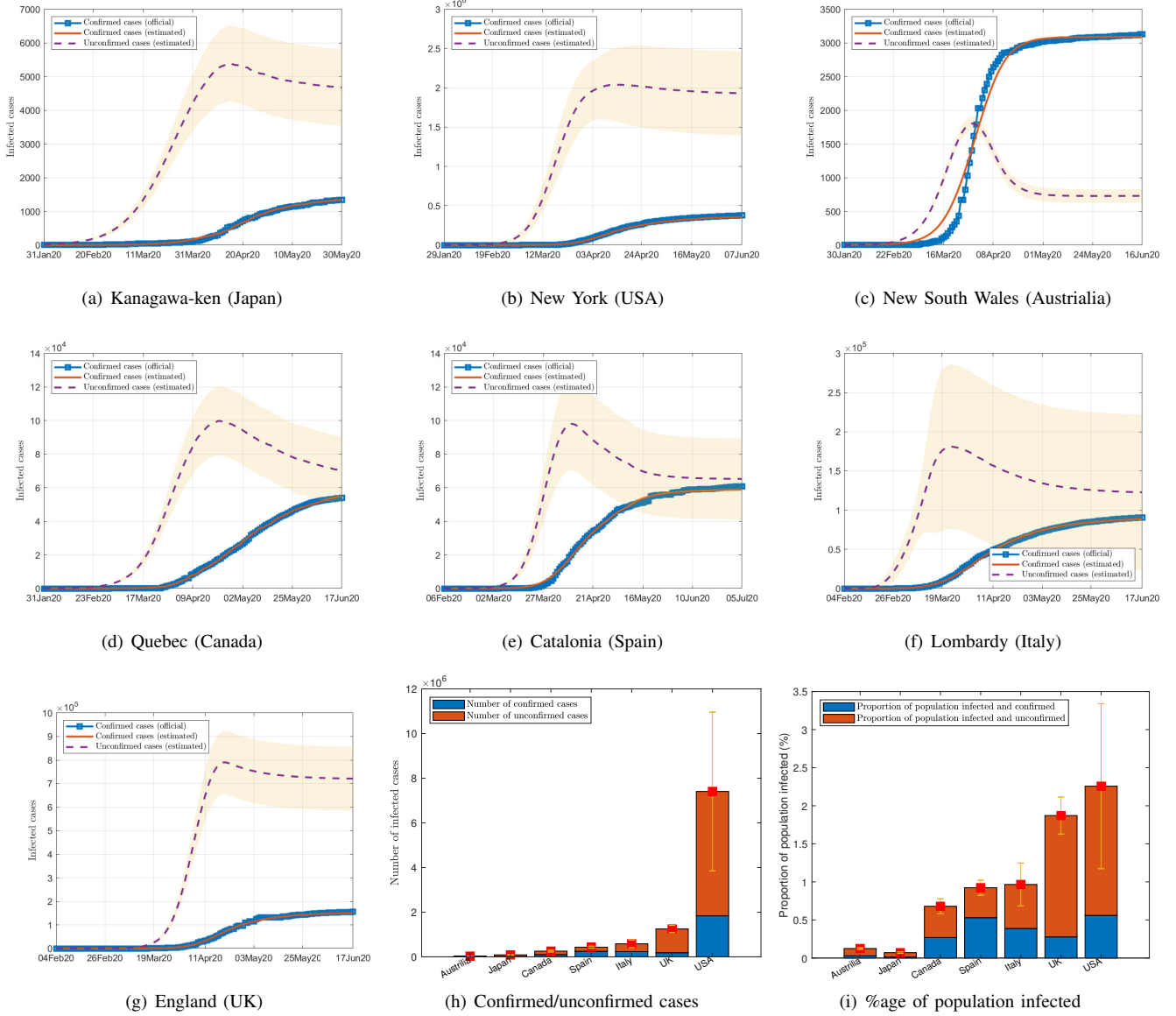
Fig. 5. (a)–(g) Official and estimated numbers of confirmed and unconfirmed (or unidentified) infected individuals in 7 selected regions. The squares are official number of cumulative infected cases, and the red solid curve is the mean estimated infected cases generated by the model. The dash curve is the estimated number of unconfirmed cases with 95% confidence interval. (h) Confirmed, unconfirmed and total cases; (i) estimated percentage of population infected until June 17, 2020.
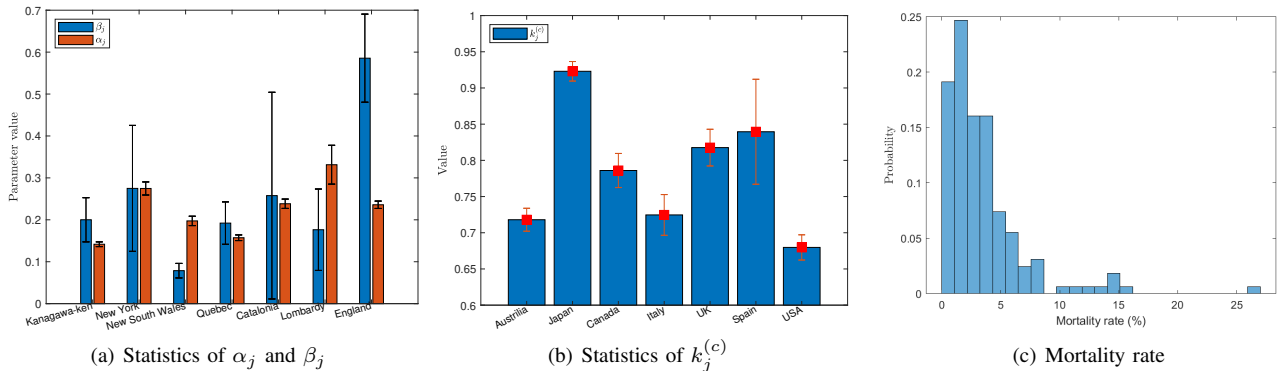


Fig. 6. (a) Mean value and 95% confidence interval of $\alpha_j$ and $\beta_j$ of selected regions; (b) mean value and 95% confidence interval of $k_j^{(c)}$ of 7 countries; (c) distribution of mortality rates of 163 countries.

TABLE IV
ESTIMATED NUMBER OF TOTAL AND UNCONFIRMED CASES BASED ON DATA COLLECTED UNTIL JUNE 17, 2020.

| Country | Mean total no. of infected cases ($I_T$) | 95% confidence interval (CI $I_T$) | Mean no. of unconfirmed (unidentified) cases ($I_u$) | Percentage of unconfirmed (unidentified) cases ($I_u/I_T$) | Ratio of unconfirmed to confirmed cases ($I_u : C$) |
|---|---|---|---|---|---|
| Australia | 33,810 | [31096, 36523] | 26,628 | 78.75% | 3.7078 : 1 |
| Canada | 35,3427 | [289660, 417194] | 251,793 | 71.24% | 2.4775 : 1 |
| Italy | 959,294 | [571075, 1347513] | 722,998 | 75.36% | 3.0597 : 1 |
| Japan | 88,642 | [75572, 101713] | 72,226 | 81.48% | 4.3997 : 1 |
| Spain | 433,276 | [387009, 479542] | 184,750 | 42.64% | 0.7433 : 1 |
| UK | 1,246,991 | [1085017, 1408964] | 1,062,089 | 85.17% | 5.7441 : 1 |
| USA | 7,406,623 | [3850704, 10962541] | 5,569,695 | 75.19% | 3.0320 : 1 |

## IV. RESULTS

### A. Estimated Number of Unconfirmed Cases

The pandemic progression profiles of 175 regions in the seven selected countries were examined. Data fitting of the model described by (10) was performed using historical daily data of confirmed and recovered cases up to June 17, 2020. For each country, we applied the parameter identification procedures repeatedly to identify 100 suitable candidate sets of parameters that satisfied the fitting criteria. For each set of parameters, we also performed a separate simulation run to generate propagation trajectories.

Fig. 5 shows the mean estimated cumulative number of infected cases generated by the sets of suitable candidate parameters and official number of infected cases for 7 selected regions, which were epicenters in the seven countries before June 17, 2020, including New South Wales (Australia), Quebec (Canada), Lombardy (Italy), Kanagawa-ken (Japan), Catalonia (Spain), and England (UK), and New York (USA).

Based on the data up to June 17, 2020, our model estimated that about 78.75% (Australia), 71.24% (Canada), 75.36% (Italy), 81.48% (Japan), 42.64% (Spain), 85.17% (UK) and 75.19% (USA) of COVID-19 patients were not identified or confirmed. In other words, the ratio of unconfirmed cases to confirmed cases are 3.7078 (Australia), 2.4775 (Canada), 3.0597 (Italy), 4.3997 (Japan), 0.7433 (Spain) and 5.7441 (UK) and 3.0320 (USA). In applying the parameter identification procedure, we identify 100 suitable candidate sets of parameters that satisfied the fitting criteria, and for each set of parameters, we performed a separate simulation run to generate propagation trajectories and estimated the total number of COVID-19 patients with the 95% confidence interval, as detailed in Table IV for each country. The number of confirmed and unconfirmed cases, and the total number of cases are presented in Fig. 5(h). Moreover, the percentage of population infected in the seven countries is graphically presented in Fig. 5(i). Results showed that the USA had the highest number of actual infected cases (7,406,623 cases) which was four-fold as many as the confirmed number, accounting for 2.2567% of its population. In terms of active intervention, all seven countries have an estimated $k_j^{(c)}$ of less than 1, based on data collected until June 17, 2020. This suggested that reasonable efforts had been made by these countries in controlling the spread of COVID-19 before June 17, 2020. To illustrate the small variation of parameters found using our parameter estimation method, Fig. 6(a) shows the statistics of $\alpha_j$, $\beta_j$ for selected regions, and 6(b) shows the statistics of $k_j^{(c)}$ for the seven countries under study.

The mean estimated ratio of unconfirmed to confirmed cases is 3.3092 for the seven selected countries under study. As of July 26, 2020, the worldwide confirmed number of cases has reached 16 million, and using this same ratio, it can be estimated that 69 million ($16 \times 4.309$) people would have actually been infected. Furthermore, for comparison purposes, we estimated the mortality rates for 163 countries and the distribution is shown in Fig. 6(c), with the mean mortality rate being 3.49% (CI, [2.84, 4.15]%). The estimated death toll would thus be 1.8543 million (c.f. the official death toll of 654,327).

### B. Enhancing Testing Capacity

To assess the effect of enhancing testing capacity, we re-ran the simulation with a 10-fold increase in the testing capacity for each country since February 1, 2020, while keeping other parameters unchanged. Fig. 7 shows a drastic reduction in the number of unidentified cases as well as the total number of cases. Specifically, our model estimated 64.62% (Australia), 24.82% (Canada), 35.07% (Italy), 44.27% (Japan), 35.39% (Spain), 53.66% (UK) and 39.90% (USA) of COVID-19 patients being unidentified or unconfirmed. This translates to much reduced ratios of unconfirmed to confirmed cases as well as the total number of infected cases, as given in Table V. The number of confirmed, unconfirmed, and total cases in the seven countries, with a 10-fold increase in testing capacity, are shown in Fig. 7(h), and the corresponding percentages of population infected are shown in Fig. 7(i). The total number of infected cases would reduce by 4.65% (Australia), 58.63% (Canada), 37.51% (Italy), 52.81% (Japan), 10.88% (Spain), 52.15% (UK) and 57.03% (USA), and the overall percentage of population infected would reduce to less than 1%.

## V. DISCUSSIONS

Our results have highlighted an alarming number of unidentified COVID-19 patients who have been allowed to move freely and interact with the susceptible population. The current
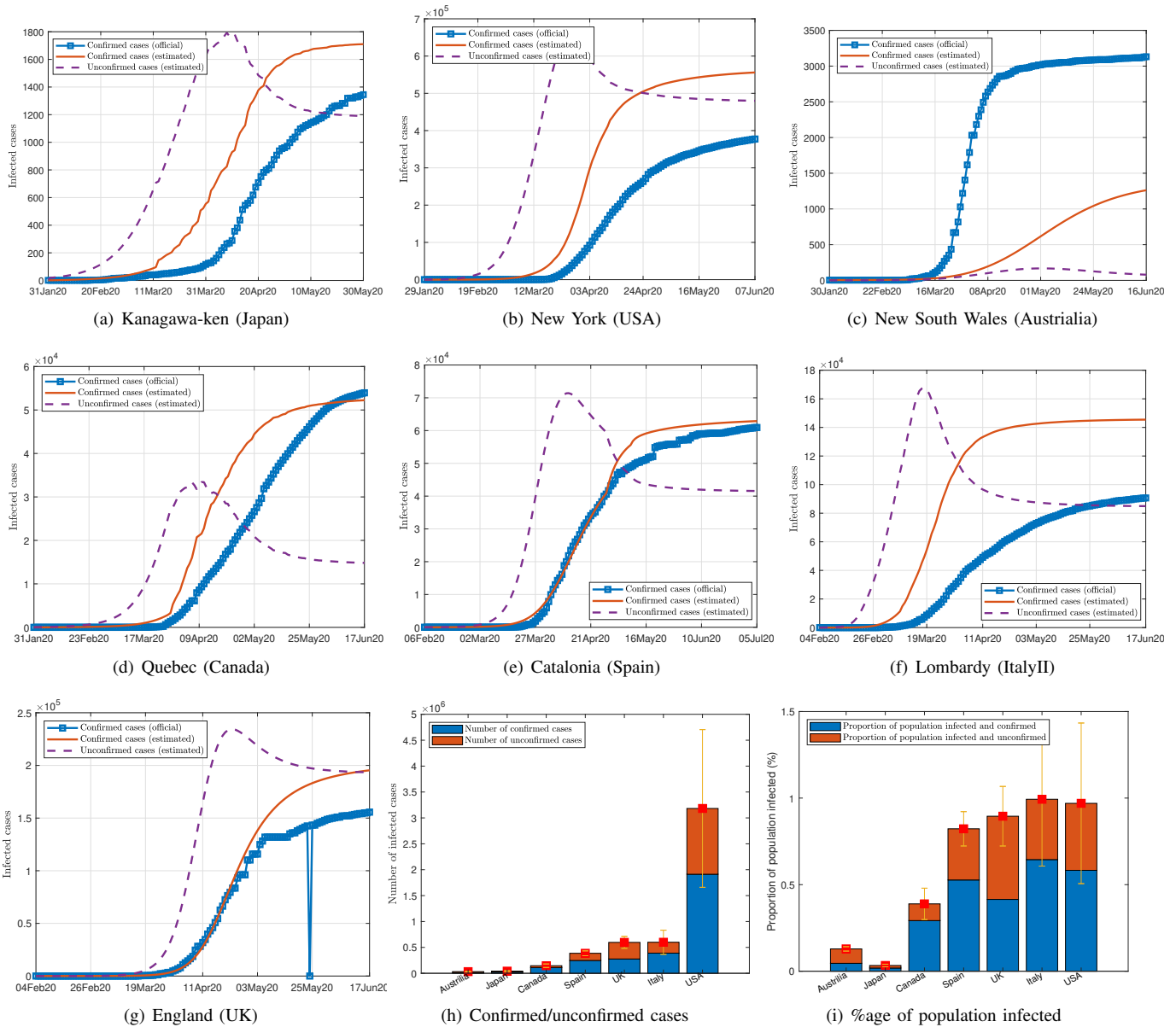
Fig. 7. Effect of 10-fold increase in testing capacity from February 1, 2020. (a)–(g) Official and estimated numbers of confirmed and unconfirmed (or unidentified) infected individuals in 7 selected regions. The squares are official number of cumulative infected cases, and the red solid curve is the mean estimated infected cases generated by the model. The dash curve is the estimated number of unconfirmed cases. (h) Confirmed, unconfirmed and total cases; (i) estimated percentage of population infected until June 17, 2020.

official or confirmed figures of infected cases represent only 23% of the actual figures (1/(3.309+1)). Authorities could have easily underestimated the scope of the outbreak, and the public could have been ill-informed of the risk of getting infected as official information had omitted the existence of 77% of the transmission chains. Moreover, the countries under study are all developed countries with high testing capacity and advanced medical technology. We can thus reasonably expect that other countries would have an even higher ratio of unconfirmed to confirmed cases. As of July 26, 2020, the official number of infected cases was 16 million, and we may reasonably project that more than 69 million people would have actually been infected.

Our study also showed that the level of active intervention applied in the seven developed countries, as reflected by

parameter $k_j^{(c)}$, were generally satisfactory up to mid June, 2020. The situation, however, could quickly deteriorate as the transmissibility of the disease remains high due to the high ratio of unidentified patients who are actively infecting others in the community.

Our study has pinpointed the importance of increasing the testing capacity, and our results showed that a drastic 10-fold increase in testing capacity since February 2020 could have brought the total number of infected cases down by about 50%, which means over 300,000 lives could have been saved. Besides, with a drastically reduced number of unidentified cases, authorities would get hold of more accurate and realistic information regarding the progression of the pandemic, and hence would be able to formulate and implement effective health responses. In the absence of effective vaccination, mass testing is thus *the* available practical option for all governments

TABLE V
ESTIMATED NUMBER OF TOTAL AND UNCONFIRMED CASES WITH 10-FOLD INCREASE IN TESTING CAPACITY BASED ON DATA COLLECTED UNTIL JUNE 15, 2020.

| Country | Mean total no. of infected cases ($I_T$) | 95% confidence interval (CI $I_T$) | Mean no. of unconfirmed (unidentified) cases ($I_u$) | Percentage of unconfirmed (unidentified) cases ($I_u/I_T$) | Ratio of unconfirmed to confirmed cases ($I_u : C$) |
|---|---|---|---|---|---|
| Australia | 32,237 | [29080, 35394] | 20,833 | 64.62% | 1.8267 : 1 |
| Canada | 146,219 | [112178, 180260] | 36,293 | 24.82% | 0.3302 : 1 |
| Italy | 599,470 | [366550, 832389] | 210,275 | 35.07% | 0.5402 : 1 |
| Japan | 41,828 | [32412, 51244] | 18,519 | 44.27% | 0.7945 : 1 |
| Spain | 386,121 | [339570, 432672] | 138,745 | 35.93% | 0.5608 : 1 |
| UK | 596,700 | [482097, 711302] | 320,234 | 53.66% | 1.1583 : 1 |
| USA | 3,182,331 | [1659784, 4704878] | 1,269,953 | 39.90% | 0.6640 : 1 |

to consider adopting in order to contain the spread of COVID-19 as soon as possible.

## VI. CONCLUSION

The main challenge in the modeling and prediction of COVID-19 pandemic progression has been the large number of factors that need to be considered in order to achieve the needed accuracy. Mobility of people, information latency (missing and delayed reporting), country or region specific transmission characteristics, and varying levels of intervention are important factors that affect the rapidity and extent of the spread of the virus. In this work, we propose a new model that takes these factors into consideration. The model considers multiple regions with different sets of parameters (characteristics) in one system, incorporates intercity travel data, distinguishes unconfirmed cases from the official data, allows active intervention to be included as a control parameter, as well as takes into account the effect of testing capacity. The model inevitably has a large parameter set, and requires advanced computational techniques for parameter estimation. In this work, we adopted an evolutionary algorithm to perform parameter extraction, which then permits prediction of future progression profiles. The model and the computational method have been applied to study the spread of COVID-19 pandemic in 175 states or regions in Australia, Canada, Italy, Japan, Spain, the UK and USA. Our results showed that the actual number of infected individuals was 3.309 times (average) as many as the official confirmed figures, with the USA having the highest number of unidentified patients. The actual total number of infected cases could have approached 70 million. Our study has also focused on the impact of increasing the testing capacity, and we showed that the scope of the COVID-19 outbreak would have been halved, had governments provided 10 times more tests since February 2020. As the pandemic continues to progress, a tracking system based on the model and results of this study may be developed for facilitating healthcare professionals to assess the scope of the pandemic and to formulate effective and timely measures to control its progression.

## REFERENCES

[1] R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, and C. Rothe, "Virological assessment of hospitalized patients with COVID-2019," *Nature*, vol. 581, no. 7809, pp. 465–469, 2020.

[2] M. F. Boni, P. Lemey, X. Jiang, T. T.-Y. Lam, B. Perry, T. Castoe, A. Rambaut, and D. L. Robertson, "Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic," *Nature Microbio.*, doi: 10.1038/s41564-020-0771-4, Jul. 2020.

[3] 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.*, vol. 27, no. 2, doi: 10.1093/jtm/taaa021, Mar. 2020.

[4] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application," *Annals Intern. Med.*, vol. 172, no. 9, pp. 577–582, 2020.

[5] K. K.-W. To, O. T.-Y. Tsang, W.-S. Leung, A. R. Tam, T.-C. Wu, D. C. Lung, C. C.-Y. Yip, J.-P. Cai, J. M.-C. Chan, and T. S.-H. Chik, "Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study," *The Lancet Infect. Dis.*, vol. 20, no. 5, pp. 565–574, May 2020.

[6] M. Gandhi, D. S. Yokoe, and D. V. Havlir, "Asymptomatic transmission, the achilles' heel of current strategies to control COVID-19," *New England J. Med.*, vol. 382, pp. 2158–2160, May 2020.

[7] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang, "Presumed asymptomatic carrier transmission of COVID-19," *J. American Med. Assoc.*, vol. 323, no. 14, pp. 1406–1407, 2020.

[8] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[9] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *New England J. Med.*, vol. 362, no. 13, pp. 1199–1207, 2020.

[10] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun *et al.*, "Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts," *The Lancet Global Health*, vol. 8, no. 4, pp. e488–496, Apr. 2020.

[11] S. Mallapaty, "Antibody tests suggest that coronavirus infections vastly exceed official counts," *Nature*, doi: 10.1038/d41586-020-01095-0, Apr. 2020.

[12] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, "Population flow drives spatio-temporal distribution of COVID-19 in China," *Nature*, vol. 582, pp. 389–394, Apr. 2020.

[13] 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," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.

[14] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang *et al.*, "An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China," *Science*, vol. 368, no. 6491, pp. 638–642, 2020.

[15] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. Du Plessis, N. R. Faria, R. Li, W. P. Hanage *et al.*, "The effect

of human mobility and control measures on the COVID-19 epidemic in China," *Science*, vol. 368, no. 6490, pp. 493–497, 2020.

[16] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. Piontti, K. Mu, L. Rossi, and K. Sun, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, 2020.

[17] C. Zhan, C. K. Tse, Y. Fu, Z. Lai, and H. Zhang, "Modelling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data," *PLoS ONE*, vol. 15, no. 10, pp. e0 241 171–1–17, Oct. 2020.

[18] Worldmeter.com, "Coronavirus cases by country, territory, or conveyance." 2020. [Online]. Available: https://www.worldometers.info/coronavirus/#countries

[19] J. Peto, "Covid-19 mass testing facilities could end the epidemic rapidly," *British Med. J.*, vol. 368, p. m1163, 2020.

[20] M. Salathé, C. L. Althaus, R. Neher, S. Stringhini, E. Hodcroft, J. Fellay, M. Zwahlen, G. Senti, M. Battegay, A. Wilder-Smith *et al.*, "COVID-19 epidemic in switzerland: on the importance of testing, contact tracing and isolation." *Swiss Med. Weekly*, vol. 150, no. 11–12, p. w20225, 2020.

[21] J. H. Tanne, E. Hayasaki, M. Zastrow, P. Pulla, P. Smith, and A. G. Rada, "Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide," *British Med. J.*, vol. 368, p. m1090, 2020.

[22] C. Zhan, C. K. Tse, Z. Lai, X. Chen, and M. Mo, "General model for COVID-19 spreading with consideration of intercity migration, insufficient testing and active intervention: modeling study of pandemic progression in Japan and USA," *JMIR Public Health Surveill.*, vol. 6, no. 3, pp. e18 880–1–10, Jul. 2020.

[23] "Japan government e-stat," 2020. [Online]. Available: https://www.e-stat.go.jp/dbview?sid=0003280293

[24] "Australian bureau of statistics," 2020. [Online]. Available: http://stat.data.abs.gov.au/

[25] "Statistics canada," 2020. [Online]. Available: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710002201

[26] "Instituto nacional de estadistica," 2020. [Online]. Available: http://dati.istat.it/Index.aspx?QueryId=19748&lang=en

[27] "Bureau of transportation statistics," 2020. [Online]. Available: https://www.transtats.bts.gov/

[28] D. Cucinotta and M. Vanelli, "Who declares COVID-19 a pandemic," *Acta Bio-medica: Atenei Parmensis*, vol. 91, no. 1, pp. 157–160, 2020.

[29] World Health Organization, "Coronavirus disease outbreak situation." 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[30] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press, 2013.

[31] M. Small, P. Shi, and C. K. Tse, "Plausible models for propagation of the SARS virus," *IEICE Trans. Fund. Electron. Comm. Comp. Sci.*, vol. 87, no. 9, pp. 2379–2386, Sep. 2004.

[32] Center for Disease Control, "Priorities for testing patients with suspected COVID-19 infection," 2020. [Online]. Available: https://lhatrustfunds.com/priorities-testing-patients/

[33] World Health Organization, "COVID-19 strategy update 2020." 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf

[34] S. He, Y. Peng, and K. Sun, "SEIR modeling of the COVID-19 and its dynamics," *Nonl. Dynamics*, vol. 101, pp. 1667–1680, Aug. 2020.

[35] H. Wang, Z. Wang, Y. Dong, R. Chang, C. Xu, X. Yu, S. Zhang, L. Tsamlag, M. Shang, J. Huang, and Y. Wang, "Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China," *Cell Discovery*, vol. 24, pp. 1–6, Feb. 2020.

[36] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, "Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures," *Proc. Nat. Acad. Sci.*, vol. 117, pp. 10 484–10 491, May 2020.

[37] L. Basnarkov, "Seair epidemic spreading model of covid-19," *Chaos Sol. Fract.*, vol. 142, pp. 110 394–1–5, Jan. 2021.

[38] C. Cartis, N. I. M. Gould, and P. L. Toint, "On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods," *SIAM J. Numeric. Anal.*, vol. 53, no. 2, pp. 836–851, 2015.

[39] W. Shi, W.-N. Chen, Y. Lin, T. Gu, S. Kwong, and J. Zhang, "An adaptive estimation of distribution algorithm for multipolicy insurance investment planning," *IEEE Trans. Evol. Comput.*, vol. 23, no. 1, pp. 1–14, Jan. 2017.

[40] Q. Yang, W.-N. Chen, J. Da Deng, Y. Li, T. Gu, and J. Zhang, "A level-based learning swarm optimizer for large-scale optimization," *IEEE Trans. Evol. Comput.*, vol. 22, no. 4, pp. 578–594, Apr. 2017.

**Choujun Zhan** (Member, IEEE) received the BS degree in automatic control engineering from Sun Yat-Sen University, Guangzhou, China, in 2007 and the PhD degree in electronic engineering from City University of Hong Kong in 2012.

He is presently a Professor with the Nanfang College of Sun Yat-Sen University, Guangzhou, China, and also with the School of Computer Science, South China Normal University, Guangzhou, China. His research interests include complex networks, epidemiology, and collective human behavior.

**Chi K. Tse** (Fellow, IEEE) received the BEng (Hons) degree with first class honors and the PhD degree, both in electrical engineering, from the University of Melbourne, Australia, in 1987 and 1991, respectively. He is presently Associate Vice-President (Research) and Chair Professor of Electrical Engineering at City University of Hong Kong, Hong Kong. His interests include power electronics, complex networks and disease spreading models.

Dr. Tse serves and has served as Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II (2016-2019), *IEEE Circuits and Systems Magazine* (2012-2015) and *IEEE Circuits and Systems Society Newsletter* (since 2007), Associate Editor for three IEEE Journal/Transactions, Editor for *International Journal of Circuit Theory and Applications*, and is on the editorial board of the IEEE PROCEEDINGS. He was the recipient of a number of research and industry awards, including Prize Paper Awards by IEEE TRANSACTIONS ON POWER ELECTRONICS in 2001, 2015 and 2017, RISP Journal of Signal Processing Best Paper Award in 2014, Best paper Award by International Journal of Circuit Theory and Applications in 2003, two Gold Medals at the International Inventions Exhibition in Geneva in 2009 and 2013, a Grand Prize and Gold Medal at the Silicon Valley International Invention Festival in 2019, and a number of recognitions by the academic and research communities, including honorary professorship by several Chinese and Australian universities, Chang Jiang Scholar Chair Professorship, IEEE Distinguished Lectureship, Distinguished Research Fellowship by the University of Calgary, Gledden Fellowship and International Distinguished Professorship-at-Large by the University of Western Australia.

**Ying Gao** received the BS and MS degrees in information science and engineering from Central South University, Changsha, China, in 1993 and 1997, respectively, and the PhD degree in computer science and engineering from South China University of Technology, Guangzhou, China, in 2003.

She is currently a Professor with the School of Computer Science and Engineering at South China University of Technology, Guangzhou, China. Her interests include software architecture design, software engineering and big data analytics,

**Tianyong Hao** (Senior Member, IEEE) received the PhD degree in computer science from City University of Hong Kong, Hong Kong, in 2010. He worked as a postdoc research scientist with Columbia University, USA, in 2013.

He is presently a Professor with the School of Computer Science at South China Normal University, Guangzhou, China. His interests include natural language processing, text mining, and medical informatics.