



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Estimating unconfirmed COVID-19 infection cases and multiple waves of pandemic progression with consideration of testing capacity and non-pharmaceutical interventions: A dynamic spreading model

Choujun Zhan<sup>a</sup>, Lujiao Shao<sup>b</sup>, Xinyu Zhang<sup>b</sup>, Ziliang Yin<sup>b</sup>, Ying Gao<sup>c</sup>, Chi K. Tse<sup>d</sup>, Dong Yang<sup>e</sup>, Di Wu<sup>f</sup>, Haijun Zhang<sup>b,\*</sup>

<sup>a</sup> School of Computing, South China Normal University, Guangzhou 510641, China

<sup>b</sup> Department of Computer Science, Harbin Institute of Technology, Shenzhen 518055, China

<sup>c</sup> School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China

<sup>d</sup> Department of Electrical Engineering, City University of Hong Kong, Hong Kong, China

<sup>e</sup> Department of Management of Complex Systems, Ernest and Julio Gallo Management Program, School of Engineering, University of California, Merced, CA 95343, USA

<sup>f</sup> Department of ICT and Natural Science, Norwegian University of Science and Technology, Norway

## ARTICLE INFO

### Article history:

Received 19 November 2021

Received in revised form 16 April 2022

Accepted 27 May 2022

Available online 6 June 2022

### Keywords:

COVID-19

Testing capacity

Extended SEIR models

Spreading dynamics

Infection estimation

## ABSTRACT

The novel coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has unique epidemiological characteristics that include presymptomatic and asymptomatic infections, resulting in a large proportion of infected cases being unconfirmed, including patients with clinical symptoms who have not been identified by screening. These unconfirmed infected individuals move and spread the virus freely, presenting difficult challenges to the control of the pandemic. To reveal the actual pandemic situation in a given region, a simple dynamic susceptible-unconfirmed-confirmed-removed (D-SUCR) model is developed taking into account the influence of unconfirmed cases, the testing capacity, the multiple waves of the pandemic, and the use of non-pharmaceutical interventions. Using this model, the total numbers of infected cases in 51 regions of the USA and 116 countries worldwide are estimated, and the results indicate that only about 40% of the true number of infections have been confirmed. In addition, it is found that if local authorities could enhance their testing capacities and implement a timely strict quarantine strategy after identifying the first infection case, the total number of infected cases could be reduced by more than 90%. Delay in implementing quarantine measures would drastically reduce their effectiveness.

© 2022 Elsevier Inc. All rights reserved.

## 1. Introduction

The novel coronavirus disease 2019 (COVID-19), as a new type of pandemic [43], has swept through almost all countries around the world, and had caused over 499 million infected cases and 6.1 million deaths by April 13, 2022. COVID-19 has

\* Corresponding author.

E-mail address: [hjzhang@hit.edu.cn](mailto:hjzhang@hit.edu.cn) (H. Zhang).

become one of the worst pandemics since the emergence of H1N1 influenza in 1918 [10]. As a result, a critical question arises as to what unique epidemiological and clinical features of COVID-19 endow it with the ability to affect the entire world. The reproduction number of COVID-19 ranges from 1.4 to 6.49 [41,31], which is higher than the average rates associated with other epidemics [3]. During the incubation period of SARS-CoV-2 virus presymptomatic patients (or exposed individuals) have a strong person-to-person transmission ability [41,11], while the transmission of other epidemics only occurs during the symptomatic period [16]. It means that this virus can spread silently through the population. In addition, the SARS-CoV-2 virus also has a much longer incubation period (five to 14 days) than normal influenza (one to four days) [25,29], which makes it more dangerous than other viruses. Additionally, there is a large proportion of asymptomatic patients [12], and these have been estimated by different methods to represent 20–70% of the actual number of COVID-19 infections [19,7]. Given the lack of external symptoms such as abnormal temperature in the early stages of infection, asymptomatic cases have a low rate for seeking medical assistance [35], resulting in a high rate of escaping symptom-based detection [23]. However, asymptomatic patients have a similar level of infectivity as symptomatic patients. Asymptomatic carriers of COVID-19 can spread the virus freely to their contacts [23], and thus can be an important source contributing to the rapid spread of COVID-19 [15,6]. Hence, the high transmissibility and long incubation time and the large number of asymptomatic cases have made it a novel type of disease that is very difficult to control.

A quantitative real-time reverse transcriptase-polymerase chain reaction (RT-PCR) assay is generally utilized to determine the presence of SARS-CoV-2 from respiratory secretions, and is used as a clinical diagnostic criterion [42]. However, due to a lack of medical resources, widespread RT-PCR testing is a challenging task for low-income countries or regions, resulting in a far smaller number of confirmed COVID-19 patients than the actual ground truth [32]. The majority of asymptomatic cases do not seek medical help and difficult to be detected, due to the lack of obvious clinical symptoms and the poor awareness of prevention among some people [13]. Hence, presymptomatic, asymptomatic, and undiagnosed COVID-19 patients form a large group of unconfirmed infected cases, who may travel from one area to another [21,40], leading to the spread of the virus to individuals encountered in the transportation and mobility networks, business venues, hotels, restaurants, and other venues [22,5]. The movement of unconfirmed cases in the population is a major contributor to the spread of COVID-19 [46], and may trigger community transmission [30] and create difficulties in terms of epidemic prevention. An estimation of the actual number of unconfirmed infections can therefore improve the understanding of the real pandemic situation and the trends of COVID-19 in a region, providing insight into the spread of the epidemic and allowing policy-makers to determine the transmission of SARS-CoV-2 and to develop appropriate prevention and control strategies in advance [35,36]. In conventional approaches, the number of unconfirmed cases is estimated using seroepidemiological data, the collection of which requires significant cost, time, and logistical effort [4]. Moreover, serological testing has limitations due to the fact that these tests vary in terms of their sensitivity and specificity. The results of testing may also be complicated by the presence of existing antibodies to other diseases, such as MERS-CoV, SARS-CoV or common cold coronaviruses [32]. A reasonable epidemic model is therefore a necessary tool to investigate the development process and to characterize the dynamic behavior of a disease.

Several studies have been carried out with the aim of estimating the ratio of unconfirmed or asymptomatic infections. The work of Nishiura et al. [35] provided a simple method for estimating the ratio of asymptomatic cases by using Japanese nationals evacuated from Wuhan, China, on Charter Flights. Also, Mizumoto et al. utilized information of COVID-19 cases on the Diamond Princess Cruise ship to develop a statistical modeling analysis for estimate the proportion of asymptomatic cases [34]. The use of epidemiological models, including the classical Susceptible-Exposed-Infected-Recovered (SEIR) model or augmented SEIR models, is another potential way of describing and estimating the spread of SARS-CoV-2 [50]. The SEIR model has motivated the development of many variants with promising enhancements, such as those with new epidemiological variables [18,33] or which consider the influence of human migration [46,49,14]. These studies have given rise to various transmission-control methods of modeling the dynamical spread of SARS-CoV-2 [47]. However, most of these studies ignore the testing capacity and the influence of presymptomatic and undocumented patients with symptoms. In addition, most traditional epidemic models do not consider the influence of non-pharmacological interventions (NPIs), leading to a modeling system with time-dependent parameters; for instance, the efficacy of the social distancing interventions imposed by governments is not constant but time-varying, resulting in a time-dependent transmission rate of COVID-19. To accurately reveal the real situation in a country or region, we need to consider the influence of unconfirmed COVID-19 patients and to incorporate data on testing capacity. Here, we propose a simple and easy-to-implement epidemiological model called Dynamic-Susceptible-Unconfirmed-Confirmed-Removed (D-SUCR, leveraging our prior work [45]), in which the testing capacity, the time-dependent influence of NPIs, the number of unconfirmed cases, population demographics, and multiple waves of spread are considered. The D-SUCR model allows us to easily evaluate the actual pandemic situation in all stages of multiple waves in a region or country, using an evolutionary computation-based system identification algorithm under rational epidemiological constraints. To our knowledge, most previously studied epidemic models can only describe the evolution of a first or separate wave [14,45,2,17,39], and with the emergence of mutant viruses, this is no longer adequate to control the pandemic. Our proposed D-SUCR model is able to dynamically provide accurate descriptions over multiple waves of the COVID-19 pandemic, which will provide more accurate and effective scientific guidance to policy decision-makers.

Most of the regions in the world have seen a second, third or even a fifth wave of the epidemic due to imported cases or the application of reopening policies [38,1]. The performance of the D-SUCR model is evaluated based on officially reported confirmed cases from 51 areas of the United States and 116 countries worldwide. Our experimental results prove the effectiveness and accuracy of the D-SUCR model in terms of simulation and estimation of the long-term trends of a pandemic

with multiple waves. In addition, the parameters used in the model, such as the reproduction number and dynamic transmission rate, can provide further insight into the characteristics of SARS-CoV-2 transmission and the efficacy of various NPIs.

Since countries differ in terms of the progression of the pandemic, the scaling-up of test capacities is one possible method of assessing the epidemiological risk and the pandemic situation, with the aim of allowing policy-makers to initiate and implement effective NPIs to prevent onward transmissions [36]. Strict quarantine is another effective NPIs for containing the spread of COVID-19. In this research, we investigate the efficacy of enhancing testing capacities and implementing strict quarantine measures. Our experimental results indicate that imposing a strict quarantine immediately after detection of the first COVID-19 patient would have effectively contained the spread of the pandemic, and that the total infections could have been reduced to 10% of the actual scenario. However, a slight delay in implementing the strict quarantine measure, such as imposing it 60 days after the emergence of the first COVID-19 patients, would have meant that the total number of infections would only have been reduced to about 80% of the actual scenario. Hence, if the virus has spread widely and there is a large group of unconfirmed cases, strictly quarantining only the confirmed cases can only slightly suppress the transmission, and involves the waste of a lot of resources in terms of isolating the confirmed cases. It therefore seems reasonable that the USA has not implemented strict quarantine measures recently, as this would have required large resources for isolation but would have had little efficacy in controlling the spread of SARS-CoV-2.

The main contributions of this work are:

- We present a novel model called D-SUCR, which considers the testing capacity and the influence of NPIs. Our D-SUCR model can estimate the number of total patients, even if there are multiple waves of COVID-19 in a region. Furthermore, we can derive the time-variant transmission rate for COVID-19, which can be used to evaluate the efficacy of NPIs.
- The proposed D-SUCR model is applied to 51 areas in the United States and 116 countries worldwide. The ratio of confirmed cases to the actual numbers of infected people is derived from Mar 1, 2020, to May 10, 2021, and the results show that at the beginning of the pandemic, less than 10% of COVID-19 patients were confirmed. Then, with an increase in the testing capacity from May 10, 2021, onwards, the proportion of unconfirmed cases decreased to approximately 40–60% of the total number of infected cases, namely, the actual number of infected individuals is likely to have been at least 2.5 times the official number.
- The D-SUCR model is also applied to investigate the influence of an enhanced testing capacity and strict quarantine measures. The results indicate that if a strict quarantine is imposed immediately after detection of the first COVID-19 patient and the testing capacity is greatly enhanced (for example by a factor of five), the number of infections could be reduced to only 5% of the actual scenario. However, if a strict quarantine measure is implemented a few days after the emergence of the first COVID-19 patient, the efficacy of containing the spread of COVID-19 is very low. Hence, if a region contains a small number of infected cases, the implementation of strict quarantine measures and increasing the testing capacity would be efficient means of containing the spread of the virus, while for a region with a large group of unconfirmed cases, implementing the same measures would only slightly reduce the number of infections but would consume a great deal of resources.

## 2. Dynamic spreading models for COVID-19

We describe the process used to extend the classical SEIR model, based on the clinical features of COVID-19, to a novel Susceptible-Exposed-Asymptomatic-Unreported-Confirmed-Recovered-Removed (SEAU CRD) model, which can then be simplified to give a basic Susceptible-Unconfirmed-Confirmed-Removed (SUCR) model with only five different categories of individuals. Finally, by incorporating the testing capacity and the NPIs in a region, the basic SUCR model is extended to give our novel D-SUCR model.

### 2.1. The SEIR model

SEIR, a classical compartmental model, has been widely used to simulate the spread of various epidemic diseases. The population of this model is usually classified into four distinct epidemic categories: susceptible ( $S$ ), exposed ( $E$ ), infected ( $I$ ), and recovered ( $R$ ) [26]. Specific explanations of these four classes are given below.

- Susceptible ( $S$ ): These are individuals who have not been infected yet and are vulnerable to disease.
- Exposed ( $E$ ): For many epidemics, there exists an incubation time during which an individual can become infected but shows no visible clinical symptoms. During this incubation period, an exposed individual has no infective properties and cannot spread the disease.
- Infectious ( $I$ ): After the incubation period, an infected individual develops obvious clinical symptoms. In this state, the infected individual has the ability to shed the virus and infect susceptible individuals.
- Recovered ( $R$ ): An infected individual has overcome the disease and is no longer infectious. The classical SEIR model assumes that a recovered individual has developed natural immunity to the disease, and hence has a low (or even no) probability of being infected by the same disease within a certain time period. In some cases, a certain proportion of the infectious individuals will not survive the disease. These individuals in state  $R$  therefore include deceased individuals, who have lost their susceptible and infective properties and can be classified into the 'removed' category.

The number of individuals at time  $t$  in each of the above four categories is defined as  $S(t)$ ,  $E(t)$ ,  $I(t)$ , and  $R(t)$ , respectively. Specific explanations of these four variables are given as follows:

- Susceptible  $S(t)$ : The amount of uninfected individuals at time  $t$ .
- Exposed  $E(t)$ : The amount of individuals who have been infected but are still in the incubation period at time  $t$ . In classical models, exposed individuals have no obvious clinical symptoms and only low infectivity.
- Infectious  $I(t)$ : The amount of infected individuals, who have high infectivity and obvious clinical symptoms.
- Recovered  $R(t)$ : The amount of recovered (or removed) individuals at time  $t$ .

The basic assumption underlying the classical SEIR model is that all individuals will cycle through these four classes based on the state transition probabilities (shown in Fig. 1(a)). More precisely, the mathematical representation of the SEIR model is shown as follows:

$$\begin{aligned}
 \dot{S} &= \mu(N - S(t)) - \frac{\beta S(t)I(t)}{N}, \\
 \dot{E} &= \frac{\beta S(t)I(t)}{N} - (\mu + \sigma)E(t), \\
 \dot{I} &= \sigma E(t) - (\mu + \gamma)I(t), \\
 \dot{R} &= \gamma I(t) - \mu R(t),
 \end{aligned}
 \tag{1}$$

where  $S(t)$ ,  $E(t)$ ,  $I(t)$  and  $R(t)$  are the system variables. The total population is  $N = S(t) + E(t) + I(t) + R(t)$ . In general, we assume that  $N$  is a constant value, although in reality,  $N$  is time-dependent, as the death and birth rates are always unequal. All parameters have physical meanings and can be described as follows (shown in Table 1):

- $\mu$  is the natural mortality rate without the specific epidemic disease.
- $\beta$  is the transmission rate from susceptible individuals to exposed individuals due to the presence of currently infected individuals, i.e., the susceptible-to-exposed transition rate.
- $\sigma$  is transition rate from exposed cases to infected cases.
- $\gamma$  represents the recovered (or removed) rate of infected individuals.

The discrete form of the SEIR model is as follows:

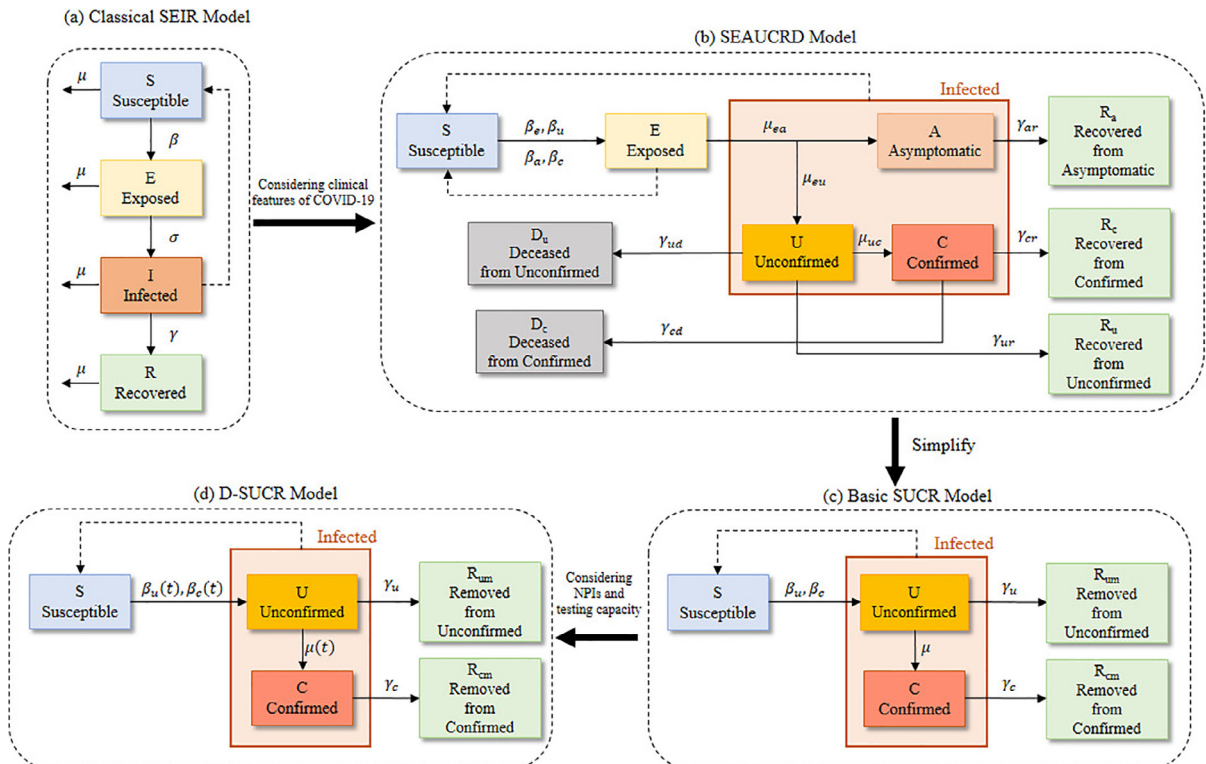


Fig. 1. Schemes used in the SEIR, SEAUCRD, Basic SUCR, and D-SUCR models, and the relationships between these four different models.

**Table 1**  
System parameters used in the SEIR model.

Variable	Description
$\mu$	Natural mortality without considering the epidemic disease
$\beta$	The contact and infection rate of transmission per contact from infected class
$\sigma$	Transition rate of exposed individuals to the infected class
$\gamma$	Recovery (or removed) rate of infected individuals

$$\begin{aligned}
 S(t + 1) &= S(t) + \mu(N - S(t)) - \frac{\beta S(t)I(t)}{N}, \\
 E(t + 1) &= E(t) + \frac{\beta S(t)I(t)}{N} - (\mu + \sigma)E(t), \\
 I(t + 1) &= I(t) + \sigma E(t) - (\mu + \gamma)I(t), \\
 R(t + 1) &= R(t) + \gamma I(t) - \mu R(t),
 \end{aligned}
 \tag{2}$$

where  $t$  is the time step. Although in some cases, the SEIR model has limitations in terms of representing the actual scenario, it still provides a basic tool for analyzing the spread of an epidemic.

### 2.2. The SEAUCRD model

Here, we first discuss the epidemiological characteristics of COVID-19, and then propose an SEAUCRD model based on these characteristics, which are very different from those of normal epidemic diseases. According to the known epidemiological characteristics of COVID-19, individuals in a population can be roughly classified into the following 10 classes:

- Susceptible ( $S$ ): A susceptible individual is vulnerable to SARS-CoV-2 but has not yet been infected.
- Exposed ( $E$ ): The mean incubation time of the original COVID-19 is about one week, and the 95th percentile of the distribution is about two weeks. In the incubation period, exposed (or presymptomatic) individuals have no typical clinical symptoms, such as fever and pneumonia, but have infective properties and can spread SARS-CoV-2. This is different from many epidemic diseases.
- Asymptomatic COVID-19 infections ( $A$ ): After the incubation period, some of the exposed individuals will pass to the group of asymptomatic COVID-19 patients, who still show no visible abnormalities in a lung computed tomography scan and have no apparent clinical symptoms, in the same way as exposed individuals. There is therefore a low probability of asymptomatic patients taking RT-PCR tests and being documented by the authorities. However, asymptomatic patients still have a certain infectivity rate. For simplicity, we can assume that the detection rate of asymptomatic patients is close to zero.
- Unreported infections with clinical symptoms ( $U$ ): After the incubation period, some COVID-19 infections begin to show clinical symptoms and can also shed the virus. Infected people with clinical symptoms have a high probability of taking PCR tests, but have not yet taken such tests. However, due to limitations on the testing capacity and other factors, only a certain proportion of these infections with clinical symptoms will be identified by screening before they recover or pass away.
- Confirmed cases with clinical symptoms ( $C$ ): A proportion of the infections with clinical symptoms will be confirmed through PCR tests in the laboratory as positive for SARS-CoV-2, and will therefore be confirmed as COVID-19 cases and reported to local authorities. These confirmed cases will be asked to self-quarantine at home, or will be centrally quarantined or hospitalized. These isolated infected individuals, however, still have a possibility of infecting other susceptible individuals, and the infection rate of confirmed cases is influenced by the level of quarantine strategy imposed.
- Recovered asymptomatic infections ( $R_a$ ): As asymptomatic cases have no visible symptoms, it is reasonable to assume that such people will not seek medical help for COVID-19. Hence, we assume that all asymptomatic patients recover and return to normal life.
- Recovered unreported infections ( $R_u$ ): A proportion of the unreported COVID-19 infections with symptoms will recover.
- Deceased unreported infections ( $D_u$ ): A proportion of the unreported COVID-19 infections with symptoms will not survive the disease.
- Recovered confirmed cases ( $R_c$ ): Most confirmed cases will recover after appropriate medical treatment.
- Deceased confirmed cases ( $D_c$ ): A few confirmed cases will not survive the disease and will pass away.

The infectious period, which is the time interval between an individual being in states  $I$  and  $R$  in the SEIR model, corresponds to the time interval in which COVID-19 infections show infectivity, and includes the following states: exposed ( $E$ ), asymptomatic ( $A$ ), unreported ( $U$ ), and confirmed ( $C$ ). We therefore replace the infected individuals  $I$  in the traditional SEIR model with asymptomatic ( $A$ ), unreported ( $U$ ) and confirmed ( $C$ ) individuals, respectively.

We can then define the numbers of individuals in the above 10 categories at time  $t$  by  $S(t), E(t), A(t), U(t), C(t), R_a(t), R_u(t), D_u(t), R_c(t)$  and  $D_c(t)$ , respectively. The definitions of these classes are given as follows:

- Susceptible  $S(t)$ : The amount of uninfected individuals at time  $t$ .
- Exposed  $E(t)$ : The amount of individuals who have been infected but are still in the incubation period at time  $t$ .
- Asymptomatic  $A(t)$ : The amount of asymptomatic infections at time  $t$ .
- Unreported infections with symptoms  $U(t)$ : The amount of infected people with obvious clinical symptoms but who have not been detected at time  $t$ .
- Confirmed cases with symptoms  $C(t)$ : The amount of infected people with clinical symptoms who have been documented at time  $t$ .
- Recovered asymptomatic infections  $R_a(t)$ : The amount of recovered individuals from the asymptomatic class at time  $t$ .
- Recovered unreported infections  $R_u(t)$ : The number of recovered individuals from the unreported class at time  $t$ .
- Recovered confirmed infections  $R_c(t)$ : The amount of recovered individuals from the confirmed class at time  $t$ .
- Deceased unreported infections  $D_u(t)$ : The amount of deceased individuals from the unreported class at time  $t$ .
- Deceased confirmed infections  $D_c(t)$ : The amount of deceased individuals from the confirmed class at time  $t$ .

An individual may cycle through these 10 classes based on the state transmission probabilities shown in Fig. 1(b). Then, a generalized SEIR model called the SEAUCRD model can then be proposed. It is formulated as follows:

$$\begin{aligned}
 \dot{S} &= -\frac{\beta_e E(t)S(t)}{N} - \frac{\beta_u U(t)S(t)}{N} - \frac{\beta_a A(t)S(t)}{N} - \frac{\beta_c C(t)S(t)}{N}, \\
 \dot{E} &= \frac{\beta_e E(t)S(t)}{N} + \frac{\beta_u U(t)S(t)}{N} + \frac{\beta_a A(t)S(t)}{N} + \frac{\beta_c C(t)S(t)}{N} \\
 &\quad - \mu_{ea}E(t) - \mu_{eu}E(t), \\
 \dot{A} &= \mu_{ea}E(t) - \gamma_{ar}A(t), \\
 \dot{U} &= \mu_{eu}E(t) - \mu_{uc}U(t) - \gamma_{ur}U(t) - \gamma_{ud}U(t), \\
 \dot{C} &= \mu_{uc}U(t) - \gamma_{cr}C(t) - \gamma_{cd}C(t), \\
 \dot{R}_a &= \gamma_{ar}A(t), \\
 \dot{R}_u &= \gamma_{ur}U(t), \\
 \dot{R}_c &= \gamma_{cr}C(t), \\
 \dot{D}_u &= \gamma_{ud}U(t), \\
 \dot{D}_c &= \gamma_{cd}C(t).
 \end{aligned} \tag{3}$$

For clarity, the descriptions of the variables involved are shown in Table 2. A discrete form of the SEAUCRD model can then be developed in the following form:

$$\begin{aligned}
 S(t+1) &= S(t) - \frac{\beta_e E(t)S(t)}{N} - \frac{\beta_u U(t)S(t)}{N} - \frac{\beta_a A(t)S(t)}{N} - \frac{\beta_c C(t)S(t)}{N}, \\
 E(t+1) &= E(t) + \frac{\beta_e E(t)S(t)}{N} + \frac{\beta_u U(t)S(t)}{N} + \frac{\beta_a A(t)S(t)}{N} + \frac{\beta_c C(t)S(t)}{N} \\
 &\quad - \mu_{ea}E(t) - \mu_{eu}E(t), \\
 A(t+1) &= A(t) + \mu_{ea}E(t) - \gamma_{ar}A(t), \\
 U(t+1) &= U(t) + \mu_{eu}E(t) - \mu_{uc}U(t) - \gamma_{ur}U(t) - \gamma_{ud}U(t), \\
 C(t+1) &= C(t) + \mu_{uc}U(t) - \gamma_{cr}C(t) - \gamma_{cd}C(t), \\
 R_a(t+1) &= R_a(t) + \gamma_{ar}A(t), \\
 R_u(t+1) &= R_u(t) + \gamma_{ur}U(t), \\
 R_c(t+1) &= R_c(t) + \gamma_{cr}C(t), \\
 D_u(t+1) &= D_u(t) + \gamma_{ud}U(t), \\
 D_c(t+1) &= D_c(t) + \gamma_{cd}C(t),
 \end{aligned} \tag{4}$$

**Table 2**  
System variables of the SEAUCRD model.

Variable	Description
$S$	Susceptible
$E$	Exposed (or pre-asymptomatic) infections
$A$	Asymptomatic infections without clinic symptoms
$U$	Unreported infections with clinical symptoms
$C$	Confirmed infections with clinical symptoms
$R_a$	Recovered asymptomatic infections
$R_u$	Recovered unreported infections
$R_c$	Recovered confirmed infections
$D_u$	Deceased unreported infections
$D_c$	Deceased confirmed infections

where  $N$  represents the total population in a region before the start of the pandemic;  $\beta_e$  is the transmission rate ( $\text{day}^{-1}$ ) from the susceptible class to the exposed class due to the current exposed class;  $\beta_a$  is the transmission rate ( $\text{day}^{-1}$ ) from the susceptible class to the exposed class due to the current asymptomatic class;  $\beta_u$  is the transmission rate ( $\text{day}^{-1}$ ) from the susceptible class to the exposed class due to the current unreported class with symptoms;  $\beta_c$  is the transmission rate ( $\text{day}^{-1}$ ) from the susceptible class to the exposed class due to the current confirmed cases;  $\mu_{ea}$  is the transmission rate ( $\text{day}^{-1}$ ) from the exposed class to the asymptomatic class;  $\mu_{eu}$  is the transmission rate ( $\text{day}^{-1}$ ) from the exposed class to the unreported infections class;  $\mu_{uc}$  is the rate of unreported infections documented by the authorities in a region at time  $t$ ;  $\gamma_{ar}$  represents the rate of recovery from asymptomatic infections;  $\gamma_{ur}$  represents the rate of recovery from unreported infections;  $\gamma_{cr}$  represents the rate of recovery for confirmed cases;  $\gamma_{ud}$  denotes the mortality rate for unreported infections; and  $\gamma_{cd}$  denotes the mortality rate for confirmed infections. Table 3 shows the summary of these parameters.

For COVID-19, there are tremendous numbers of asymptomatic and unreported cases, and these are considered in the SEAUCRD model. Note that the incidence rates  $\frac{\beta_e E(t)S(t)}{N}$ ,  $\frac{\beta_a A(t)S(t)}{N}$ ,  $\frac{\beta_u U(t)S(t)}{N}$ , and  $\frac{\beta_c C(t)S(t)}{N}$  are used to describe the transmission of disease [27]. These incidence rates play a significant role in the epidemic, and can be applied to describe the evolution of an infectious disease. The incidence rates and parameters used in the SEAUCRD model can be influenced by numerous factors, such as public interventions, age, gender, genetic profile, and health status, and these constant parameters therefore represent the mean rates over a certain period. Three main channels are considered in the SEAUCRD model (shown in Fig. 1(b)). The first one goes to  $S \rightarrow E \rightarrow A \rightarrow R_a$ , while the second is  $S \rightarrow E \rightarrow U \rightarrow R_u$  or  $D_u$ , and the third is  $S \rightarrow E \rightarrow U \rightarrow C \rightarrow R_c$  or  $D_c$ . These reflect the influence of the asymptomatic, unreported and confirmed cases, respectively. In fact, the classical SEIR model is a simplified version of the proposed SEUCRD model, which does not consider asymptomatic ( $A(t) = 0$ ) or unreported cases ( $U(t) = 0$ ).

### 2.3. Basic SUCR model by simplifying SEAUCRD

The historical pandemic data released by local authorities mainly consist of confirmed infections  $C$ , recovered confirmed infections  $R_c$ , and the death toll (or deceased confirmed infections)  $D_c$ . However, there are also asymptomatic infections ( $A$ ), unreported infections with symptoms  $U$ , recovered asymptomatic infections  $R_a$ , recovered unreported infections  $R_u$ , and deceased unreported infections  $D_u$ . As a consequence, we can simplify the 10 classes in the SEAUCRD model into five different classes: susceptible individuals  $S$ , active unconfirmed cases  $U$ , active confirmed cases  $C$ , officially recorded removed cases  $R_{cm}$ , and unrecorded removed cases  $R_{um}$ . Descriptions of these five classes are given below:

- Susceptible  $S$ : a susceptible individual is vulnerable but has not yet been infected.
- Unconfirmed infections  $U$ : Active, exposed, asymptomatic, and unreported infections with symptoms all have infective properties and can spread SARS-CoV-2, but have not been confirmed by local authorities. We can therefore combine the classes  $E, A$ , and  $U$  in the SEAUCRD model into a single large class called unconfirmed infections  $U$ , i.e.,  $(E + A + U) \rightarrow U$ .
- Confirmed infections with clinical symptoms  $C$ : A proportion of the unconfirmed infections will be confirmed in the laboratory, and hence will be recorded as COVID-19 patients and reported to local authorities.
- Removed unconfirmed infections  $R_{um}$ : These are individuals who have been removed from the unconfirmed class and have lost their infective or susceptible properties, i.e.,  $(R_a + R_u + D_u) \rightarrow R_{um}$ .
- Removed confirmed infectious  $R_{cm}$ : These are individuals who have been removed from the confirmed class and have lost their infective or susceptible properties, i.e.,  $(R_c + D_c) \rightarrow R_{cm}$ .

**Table 3**  
System parameters of the seaucrd model.

Variable	Description
$N$	The number of population in a region before the start of the COVID-19 pandemic
$\beta_e$	The contact and infection rate of transmission per contact from exposed class
$\beta_a$	The contact and infection rate of transmission per contact from asymptomatic class
$\beta_u$	The contact and infection rate of transmission per contact from unreported class
$\beta_c$	The contact and infection rate of transmission per contact from Confirmed class
$\mu_{ea}$	Transition rate ( $\text{day}^{-1}$ ) from the exposed class to the asymptomatic class
$\mu_{eu}$	Transition rate ( $\text{day}^{-1}$ ) from the exposed class to the unreported class
$\mu_{uc}$	Transition rate ( $\text{day}^{-1}$ ) from the unreported class to the confirmed class
$\gamma_{ar}$	Transition rate of the asymptomatic class to the recovered class
$\gamma_{ur}$	Transition rate of the unreported infectious class to the recovered class
$\gamma_{cr}$	Transition rate of the confirmed infectious class to the recovered class
$\gamma_{ud}$	Transition rate of the unreported infectious class to the deceased class
$\gamma_{cd}$	Transition rate of the confirmed infectious class to the deceased class



Hence, in this scenario, we replaced the exposed ( $E$ ), asymptomatic ( $A$ ) and unreported ( $U$ ) individuals of the SEAUCRD model with the unconfirmed individuals ( $U$ ) of the SUCR model. The confirmed individuals ( $C$ ) indicate the COVID-19 patients who have been detected and quarantined.

At time  $t$ , the amount of individuals in the above five classes is denoted by  $S(t)$ ,  $U(t)$ ,  $C(t)$ ,  $R_{cm}(t)$ , and  $R_{um}(t)$ , respectively:

- Susceptible  $S(t)$ : The number of uninfected individuals at time  $t$ ;
- Unconfirmed infectious  $U(t)$ : The number of unconfirmed active COVID-19 infections, including exposed, asymptomatic, and unreported infected cases with symptoms;
- Confirmed infectious  $C(t)$ : The number of confirmed COVID-19 patients;
- Removed unconfirmed infectious  $R_{um}(t)$ : The number of individuals removed from the unconfirmed class, i.e.,  $R_{um} = R_a(t) + R_u(t) + D_u(t)$ ;
- Removed confirmed infections  $R_{cm}$ : The number of individuals removed from the confirmed class, i.e.,  $R_{cm} = R_c(t) + D_c(t)$ .

An individual can cycle through these five classes based on the state transition probabilities, as shown in Fig. 1(c). We can then simplify the SEAUCRD model and develop our basic SUCR model. In this model, susceptible individuals can be infected through contact with closely infected individuals, thus becoming unconfirmed cases who have been infected but not confirmed. Unconfirmed cases are usually not quarantined, and can transmit the virus freely before being screened out. Unconfirmed patients transit to the confirmed state at a rate proportional to both the number of unconfirmed cases and the testing capacity. In general, confirmed cases are either hospitalized or quarantined, and finally transit into the removed state, meaning they have either recovered or passed away and cannot infect susceptible individuals. The basic SUCR model can then be summarized as:

$$\begin{aligned}
 \dot{S}(t) &= -\frac{\beta_c C(t)S(t)}{N} - \frac{\beta_u U(t)S(t)}{N}, \\
 \dot{U}(t) &= \frac{\beta_c C(t)S(t)}{N} + \frac{\beta_u U(t)S(t)}{N} - \mu U(t) - \gamma_{ur} U(t), \\
 \dot{C}(t) &= \mu U(t) - \gamma_{cr} C(t), \\
 \dot{R}_{cm}(t) &= \gamma_{cr} C(t), \\
 \dot{R}_{um}(t) &= \gamma_{ur} U(t),
 \end{aligned} \tag{5}$$

At time  $t$ , the amount of infected cases is  $I(t) = C(t) + U(t)$ , while the amount of removed cases is  $R(t) = R_{cm}(t) + R_{um}(t)$ . The definitions of these variables are summarized in Table 4. The discrete form of the simplified SUCR model is formulated as follows:

$$\begin{aligned}
 S(t + 1) &= S(t) - \frac{\beta_c C(t)S(t)}{N} - \frac{\beta_u U(t)S(t)}{N}, \\
 U(t + 1) &= U(t) + \frac{\beta_c C(t)S(t)}{N} + \frac{\beta_u U(t)S(t)}{N} - \mu U(t) - \gamma_{ur} U(t), \\
 C(t + 1) &= C(t) + \mu U(t) - \gamma_{cr} C(t), \\
 R_{cm}(t + 1) &= R_{cm}(t) + \gamma_{cr} C(t), \\
 R_{um}(t + 1) &= R_{um}(t) + \gamma_{ur} U(t),
 \end{aligned} \tag{6}$$

where  $N$  is the population in an area before the start of the pandemic;  $\beta_u$  is the infection rate from susceptible to unconfirmed cases due to the current unconfirmed cases;  $\beta_c$  is the infection rate from susceptible to unconfirmed cases due to the current confirmed cases;  $\mu$  is the fraction of unconfirmed infections that are documented by the authority in an area;  $\gamma_{ur}$  represents the removal rate of unconfirmed individuals; and  $\gamma_{cr}$  is the removal rate of confirmed cases. For clarity, a summary of these variables is presented in Table 5, and a transmission diagram for the basic SUCR model is shown in Fig. 1(c).

It should be noted that different countries and regions may implement different quarantine strategies for confirmed cases. In regions with a more relaxed strategy, confirmed cases still have the possibility of spreading the virus to susceptible individuals [24]. The basic SUCR model does not take into account the influence of the testing capacity and public interventions implemented by the local authorities, and therefore has some limitations in terms of representing the actual situation. Obviously, the basic SUCR model cannot capture COVID-19 pandemics with multiple waves.

**Table 4**  
System variables of the SUCR and D-SUCR models.

Variable	Description
$S$	Susceptible individuals
$U$	Unconfirmed infections
$C$	Confirmed infections
$R_{um}$	Recovered unconfirmed infections
$R_{cm}$	Recovered confirmed infections

**Table 5**  
System parameters of the SUCR and D-SUCR models.

Variable	Description
$N$	The number of individuals in a region before the start of the pandemic
$\beta_u$	The contact and infection rate of transmission per contact from unconfirmed class
$\beta_c$	The contact and infection rate of transmission per contact from confirmed class
$\gamma_{ur}$	Transition rate of unconfirmed infectious class to the removed class
$\gamma_{cr}$	Transition rate of confirmed infectious class to the removed class
$N_s$	The amount of patients with COVID-19 like symptoms but not COVID-19 infections
$k_T$	A constant ratio utilized to calibrate the testing performance
$k_c$	The average number of close contacts of a confirmed COVID-19 infections
$\phi$	The fluctuation rate referring to the efficiency of public intervention.

2.4. The D-SUCR model

Due to the limitations on testing capacities, detection and reporting may not be done in a timely manner in some regions, or possibly in an entire country. In a real-world scenario, the more tests that are administered, the higher the probability of screening out unconfirmed cases. An increase in the testing capacity can therefore increase the percentage of detection of unconfirmed cases. Hence, the rate at which unconfirmed cases become confirmed cases should not be a constant ratio  $\mu$ , as in the SEACRD model (Eq. (3)) and the basic SUCR model (Eq. (5)), but should be proportional to the testing capacity and should be expressed as a time-dependent rate  $\mu(t)$ . In addition, a large group of patients with symptoms that are similar to COVID-19 but are actually not COVID-19 cases are also COVID-19 candidates who have a probability of being tested. For simplicity, let  $N_s$  represent the amount of patients with COVID-19-like symptoms, and suppose that the close contacts of a confirmed case can be always tested by the local authorities. We can assume that  $k_c$  is the average number of close contacts of a confirmed case; then, the number of close contacts is roughly  $k_c C(t)$ . The total number of individuals who are COVID-19 candidates is then roughly equal to  $U(t) + k_c C(t) + N_s$ . We can assume that unconfirmed infections and COVID-19 candidates are evenly distributed. The rate at which unconfirmed infections are diagnosed can then be formulated as:

$$\mu(t) = \frac{k_T N_T(t)}{U(t) + k_c C(t) + N_s}, \tag{7}$$

where  $k_T$  is a constant ratio for calibration. The number of unconfirmed cases detected then becomes  $\frac{k_T N_T(t) U(t)}{U(t) + k_c C(t) + N_s}$ . Eq. 7 indicates that the higher the amount of COVID-19 candidates, the lower the probability of detecting a COVID-19 infection in a single test.

Preventive (or containment) measures, such as promoting self-protection, maintaining social distance, wearing face masks in public areas, tracing close contacts, quarantining infected cases, or even locking down cities, were introduced and implemented at certain times to curb the spread of COVID-19 during emergency periods. However, local authorities cannot impose preventive measures all the time, as there would be negative impacts on the economy and human well-being. Obviously, tightening or relaxing containment measures can influence the transmission rate of the disease, and the transmission rates  $\beta_u$  and  $\beta_c$  in the basic SUCR model should be time-variant, according to the strictness of the containment measures. It is therefore reasonable to introduce dynamic transmission rates  $\beta_u(t)$  and  $\beta_c(t)$ , which can reflect the time-dependent efficiency of NPIs and improve the model. Here, we assume that the transmission rate gradually increases or decreases in the form of:

$$\begin{aligned} \beta_u(t+1) &= (1 + \phi(t))\beta_u(t), \\ \beta_c(t+1) &= (1 + \phi(t))\beta_c(t), \end{aligned} \tag{8}$$

where  $|\phi(t)| \leq \phi_T$  is the fluctuation rate. Here  $\phi_T > 0$  stands for a threshold. Obviously, we have  $1 - \phi_T \leq 1 + \phi(t) \leq 1 + \phi_T$ , meaning that the transmission in one-time step cannot be less than  $1 - \phi_T$  or larger than  $1 + \phi_T$  times the transmission rate of the previous step. Based on this, we propose a dynamic epidemiological model called D-SUCR in which we leverage the testing capacity to reveal the actual pandemic situation and the efficacy of NPIs, as shown in Fig. 1(d). More precisely, our D-SUCR model is formulated as:

$$\begin{aligned} S(t+1) &= S(t) - \frac{\beta_c(t)C(t)S(t)}{N} - \frac{\beta_u(t)U(t)S(t)}{N}, \\ U(t+1) &= S(t) + \frac{\beta_c(t)C(t)S(t)}{N} + \frac{\beta_u(t)U(t)S(t)}{N} - \frac{k_T N_T(t)U(t)}{U(t) + k_c C(t) + N_s} - \gamma_u U(t), \\ C(t+1) &= C(t) + \frac{k_T N_T(t)U(t)}{U(t) + k_c C(t) + N_s} - \gamma_c C(t), \\ R_{cm}(t+1) &= R_{cm}(t) + \gamma_c C(t), \\ R_{um}(t+1) &= R_{um}(t) + \gamma_u U(t), \\ \beta_u(t+1) &= (1 + \phi(t))\beta_u(t), \\ \beta_c(t+1) &= (1 + \phi(t))\beta_c(t), \end{aligned} \tag{9}$$

where the parameters  $N_s, k_c, k_T$ , and  $\phi(t)$  are defined as follows:

- $N_s$  is the amount of patients that have some COVID-19-like symptoms but are not actual COVID-19 cases;
- $k_c$  is the average number of close contacts for a confirmed case;
- $k_T$  is a ratio that is used to calibrate the testing performance (as not every COVID-19 test is performed accurately);
- $\phi(t)$  is the fluctuation rate which indicates the efficiency of NPIs.

For clarity, Table 5 summarizes these variables. The D-SUCR model consists of only one main channel,  $S \rightarrow U \rightarrow C \rightarrow R$  (shown in Fig. 1(d)). In summary, our proposed model has five variables:

$$X(t) = \{S(t), U(t), C(t), R_{cm}(t), R_{um}(t)\}, \tag{10}$$

and seven parameters

$$\theta = \{\beta_u, \beta_c, \gamma_u, \gamma_c, k_T, k_c, N_s, \phi\}. \tag{11}$$

In this model,  $\beta_c$  is the transmission rate of confirmed cases. In some countries or regions, confirmed cases and individuals exposed to the SARS-CoV-2 may be quarantined in a hospital, hotel, or mobile hospital, resulting in  $\beta_c \approx 0$ ; however, some local authorities have suggested that confirmed cases should self-quarantine at home for 14 days. The main limitation of this approach is that family members may be exposed to COVID-19, resulting in new infections [21] and  $\beta_c > 0$ .  $\beta_u$  represents the infection rate of the group of unconfirmed cases, including presymptomatic, asymptomatic and COVID-19 infections with symptoms. Studies have shown that the rapid spread of the virus is mainly attributable to new undiagnosed COVID-19 infections [28]. To our knowledge, apart from our previous work [45], almost no prior researchers have considered incorporating the testing capacity into their epidemiological models. Our proposed D-SUCR model incorporates historical data on testing capacities in order to model the trends in COVID-19. Moreover, few epidemiological models have considered the influence of unconfirmed cases or have tried to estimate the actual number of total infections from historical data. In contrast, the D-SUCR model incorporates officially released pandemic data, including information on testing capacities and the influence of NPIs, to reveal the actual pandemic situation and to estimate the actual amount of infected cases.

### 3. Evolutionary computation method for parameter estimation

---

**Algorithm 1:** Sub-algorithm for deriving time series  $\hat{y}_i$

---

**Input:** The initial set and parameter of the model:

$$\Theta = \{S_0, U_0, C_0, R_{cm,0}, R_{um,0}, \beta_{u,0}, \beta_{c,0}, \gamma_u, \gamma_c, k_T, N_s, \phi\}, \tag{12}$$

where  $\phi = \{\phi_0, \phi_1, \phi_2, \dots, \phi_{K-1}\}$ ;

**Output:** A time series  $\{\hat{C}(t_0), \hat{C}(t_1), \dots, \hat{C}(t_K)\}$  and  $\{\hat{R}_{cm}(t_0), \hat{R}_{cm}(t_1), \dots, \hat{R}_{cm}(t_K)\}$ ;

*Initialisation:*

1: Set  $\beta_u^{(0)} = \beta_{u,0}$ , and  $\beta_c^{(0)} = \beta_{c,0}$

*LOOP Process*

2: **for**  $i = 1$  to  $K - 1$  **do**

3: Derive  $\hat{X}(t_i)$

$$\hat{X}(t_i + 1) = \hat{X}(t_i) + f(\hat{X}(t_i) | \Theta),$$

and  $\beta_u^{(i)}, \beta_c^{(i)}$

$$\begin{aligned} \beta_u^{(i+1)} &= (1 + \phi_i)\beta_c^{(i)}, \\ \beta_c^{(i+1)} &= (1 + \phi_i)\beta_u^{(i)}. \end{aligned}$$

4: **end for**

5: Note that  $\hat{X}_i = \{\hat{S}_i, \hat{U}_i, \hat{C}(t_i), \hat{R}_{um}(t_i), \hat{R}_{cm}(t_i)\}$ , hence  $\hat{C}(t_i)$  and  $\hat{R}_{cm}(t_i)$  can be extracted.

6: **return:** The estimated confirmed cases  $\hat{C} = \{\hat{C}(t_0), \hat{C}(t_1), \dots, \hat{C}(t_K)\}$  and removed cases

$$\hat{R}_{cm} = \{\hat{R}_{cm}(t_0), \hat{R}_{cm}(t_1), \dots, \hat{R}_{cm}(t_K)\}.$$


---

**Algorithm 2:** Algorithm for estimating the optimal parameter set  $\Theta^*$ **Input:** The initial parameter set  $\Theta_0$  that needs to be optimized:**ensure:** Optimal parameter set  $\Theta^*$ ;*Initialisation:*1: Initialize the temperature  $T$  and randomly adopt a starting point

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

where  $k_{rand} \in [0, 1]$  is a random real number. Initialize temperature  $T$  and set

$$\Theta_{current} = \Theta_0.$$

Compute the value of the objective function in Eq. (15) with the parameter set  $\Theta_{current}$  with the estimated time series from Algorithm 1

$$cost_{current} = \sum_{i=1}^N w_{C,i} \left( C(t_i) - \hat{C}(t_i|\Theta) \right)^2 + w_{R,i} \left( R_{cm}(t_i) - \hat{R}_{cm}(t_i|\Theta) \right)^2.$$

*LOOP Process*2: **for**  $i_{iter} = 0$  to  $i_{max}$  **do**

3:

$$i_{iter} = i_{iter} + 1$$

$$temp_{iter} = 0$$

$$\Theta_{previous} = \Theta_{current}$$

$$cost_{previous} = cost_{current}$$

4: **while**  $temp_{iter} \leq n_{rep}$  **do**

5:

$$temp_{iter} = temp_{iter} + 1$$

6: **if**  $\theta_j \in \Phi$  **then**7:   Adopt a new set of parameters  $(\theta_j)$  from the neighborhood

$$\theta_j \leftarrow \theta_j$$

8: **else**9:   Keep  $\theta_j = \theta_{j,0}$ 10: **endif**11: Derive the value of the objective function in Eq. (15) and compute  $\delta$ 

$$\delta = cost_{current} - cost_{previous}$$

12: **if**  $\delta < 0$  **then**

13:   Keep this new parameter set

14: **else**15:   Keep this new parameter set with probability  $\exp(-\delta/T)$ 16: **end if**17: **end while**

18:

$$T = a * T, (0 < a < 1)$$

19: **end for** 20: **return**  $\Theta^*$

One of the essential stages in the development and evaluation of a newly developed epidemiological model is the estimation of unknown system parameters. Parameter estimation (or system identification) from historical data is a procedure for tuning unknown model parameters to fit the historical epidemic data. This process is necessary to evaluate the ability of a model to capture real situations accurately, in a reasonable and verifiable manner. The results inferred from the tuned model can then be compared with historical records to either disprove or affirm the basic assumptions of the model.

Here, we define  $X_0 = \{S_0, U_0, C_0, R_{cm,0}, R_{um,0}\}$  as the initial numbers of susceptible, unconfirmed, confirmed, removed confirmed, and removed unconfirmed individuals at time  $t_0$ , respectively. The D-SUCR model relies on a set of unknown parameters, i.e.,  $\theta = \{\beta_u, \beta_c, \gamma_u, \gamma_c, k_T, k_c, N_s, \phi\}$ . We assume that at the initial time  $t_0, \beta_u(t_0) = \beta_{u,0}, \beta_c(t_0) = \beta_{c,0}$  and  $\phi_0 = 1$ . This set of unknown parameters determines the transmission of a disease. Let  $X(t)$  represent the extended state vector, i.e.,  $X(t) = \{S(t), U(t), C(t), R_{cm}(t), R_{um}(t)\}$ , as shown in Eq. (10). Then, the D-SUCR model can be formulated as:

$$X(t_i + 1) = X(t_i) + f(X(t_i)|\theta), \tag{13}$$

where  $f(x)$  represents the right side of the D-SUCR model as shown in Eq. (9); and  $\theta$  is the set of unknown parameters. From Eq. (9), we can derive an algorithm for calculating the estimated pandemic trajectories  $\hat{X}(t_i)$ , as summarized in Algorithm 1. The unknown set is given by

$$\Theta = \{X_0, \theta\} = \{S_0, U_0, C_0, R_{um,0}, R_{cm,0}, \beta_{u,0}, \beta_{c,0}, \gamma_u, \gamma_c, k_T, k_c, N_s, \phi(t)\}, \tag{14}$$

which essentially has 12 unknown parameters to characterize the trajectory generated by the D-SUCR model.

The D-SUCR model is a nonlinear dynamic model in which the parameters are hard to estimate by an explicit method in a closed form. In this study, we solve this problem by using a nonlinear optimization approach in which a least-squares error function is minimized. The purpose of parameter estimation is to search for suitable parameters so that the estimated spreading trajectories closely match the historical records. The problem of parameter estimation can therefore be considered as a constrained nonlinear optimization problem, i.e.,

$$P_0 : \min_{\Theta} \sum_{i=1}^N w_{C,i} (C(t_i) - \hat{C}(t_i|\Theta))^2 + w_{R,i} (R_{cm}(t_i) - \hat{R}_{cm}(t_i|\Theta))^2$$

$$\text{s.t.} \begin{cases} \text{(i)} & X(t_i + 1) = X(t_i) + f(X(t_i)|\Theta) \\ \text{(ii)} & X(t_0) = X_0 \\ \text{(ii)} & \Theta_U \geq \Theta \geq \Theta_L \end{cases}, \tag{15}$$

where  $\hat{C}(t_i|\theta)$  and  $\hat{R}_{cm}(t_i|\theta)$  are the estimated amount of confirmed and removed cases, respectively, with initial condition  $X_0$  and parameter set  $\theta$ ; and  $w_{C,i}$  and  $w_{R,i}$  represent the weighted coefficients. The upper and lower bounds on the unknown parameter set are  $\Theta_U$  and  $\Theta_L$ , respectively. An evolutionary computation algorithm is adopted to search for the optimal parameters and initial states by solving Eq. (15).

The optimization problem described in Eq. (15) is subject to the constraints specified by the lower and upper bound vectors  $\Theta_L$  and  $\Theta_U$ , respectively, as summarized in Table 6. Here, we adopt an evolutionary computation algorithm minimize the problem, as this type of approach is suitable for dealing with nonlinear constrained optimization problems. We used official COVID-19 records from 116 countries and 51 regions of the USA, and applied a simulated annealing (SA) algorithm (in which the main idea is similar to the approach used in our previous work [48]) to calibrate the unknown parameter set  $\theta$  match the real scenario. The pseudocode for the optimization algorithm is given in Algorithm 2. The pandemic data were collected from

**Table 6**  
Search space of the system parameters of the D-SUCR model.

Variable	lower bound	upper bound
$\hat{S}_0$	$0.8(C(t_K) + R_{cm}(t_K) + 1)$	$\min\{0.7N_p, 100(C(K) + R_{cm})\}$
$\hat{U}_0$	$0.01C(t_0) + 0.1$	$100C(t_0) + 1$
$\hat{C}_0$	$0.8C(t_0) + 0.1$	$10C(t_0) + 1$
$\hat{R}_{cm,0}$	$0.5R_{cm}(t_0) + 0.1$	$3R_{cm}(t_0) + 1$
$\hat{R}_{um,0}$	$0.5R_{cm}(t_0) + 0.1$	$3R_{cm}(t_0) + 1$
$\beta_{u,0}$	$0.0001/\hat{N}$	$0.4/\hat{N}$
$\beta_{c,0}$	$0.0001/\hat{N}$	$0.4/\hat{N}$
$\gamma_u$	0.001	0.2
$\gamma_c$	0.01	0.2
$k_T$	0.01	1000
$k_c$	0.1	$N_p / \max\{C(t_0), C(t_1), \dots, C(t_K)\}$
$N_s$	$0.01(C(K) + R_{cm}(K))/K$	$0.1N_p$
$\phi_i$	-0.1	0.1
$\hat{N}$	$\hat{N} = \hat{S}_0 + \hat{U}_0 + \hat{C}_0 + \hat{R}_{cm,0} + \hat{R}_{um,0}$	

the website of Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>). The settings of the SA algorithm are as follows: the termination tolerance of function value is  $10^{-10}$ , the initial value of the temperature is 100, and the maximum number of iterations is 150,000. In some cases, the optimization algorithm is not able to converge, and we therefore set the maximum time for the algorithm to half an hour, meaning that the program would stop if the running time was longer than this. The model is fitted using a nonlinear optimization approach, by calculating the normalized least-squares error which represents the difference between the estimated trajectory generated by the model and the historical trajectory recorded by the local authority, as shown in Eq. (15). Here, the initial values of the unknown parameters are set randomly, with a uniform distribution. This procedure has been carried out at least 2,000 times with random initial conditions to avoid always falling into the same local minima. When the parameters had been determined by the optimization algorithm, the model could be used to characterize the tendency of COVID-19 outbreak and to investigate the pandemic situation in a region or country.

#### 4. Experimental results

We collected COVID-19 pandemic data for 190 countries based on national public health agencies around the world. The USA has released the most detailed data about testing capacities, including the daily testing capacity of 51 different regions, including the 50 states and Washington DC. Most other nations in the world have not released detailed testing capacity information for each region, but instead have released the total testing capacity for the whole country. Although 190 countries have released pandemic information, only 116 have released information on testing capacities for the whole country. We therefore applied our model to these 51 regions of the United States and 116 other countries worldwide.

##### 4.1. Estimated pandemic situation in 51 regions of the USA

We first generated the numbers of confirmed and unconfirmed cases for the 51 regions in the USA from the proposed D-SUCR model with the optimal parameter sets. Parameter estimation for the D-SUCR model, as described by Eq. (15), was performed using the official number of the confirmed and removed cases and the testing capacity, up to May 10, 2021. For each region, the parameter estimation procedures were applied repeatably to derive more than 100 suitable candidate parameters sets that satisfied the fitting criteria for more than 2,000 identification procedures, for ensuring the reliability of the analysis results. Our experimental results indicate that this model can accurately estimate the daily records of unconfirmed, confirmed, and removed cases in a region with multiple waves. The estimated values closely matched the actual situation. Here, the coefficient of determination ( $R^2$ ) was adopted to evaluate the performance of the proposed D-SUCR model:

$$R^2 = 1 - \frac{\sum_{i=1}^K (y_i - y_{e,i})^2}{\sum_{i=1}^K (y_i - \bar{y}_i)^2}, \tag{16}$$

where  $y_i$  is the actual value and  $y_{e,i}$  is the estimated value. For data up to May 2021, most of the values of  $R^2$  for the 51 regions were all larger than 0.97 (as shown in Fig. 2). This indicates that the estimated values fit well with the actual scenario.

Due to space limitation, we used four typical regions as illustrative examples: California, New York, Washington, and the USA as a whole. Fig. 3 shows the numbers of officially confirmed cases and the estimated total infections for these three states and the whole USA, as of mid-May 2021. The historical pandemic and testing capacity data for the model simulation were collected from the CDC in the USA, from March 2020 to mid-May 2021. Examples of the officially released data and model estimates for the pandemic situation in three states and the US as a whole are displayed in Figs. 3, which show the official numbers of infected cases, mean estimated cumulative numbers of infected patients with 95% confidence interval (CI) generated by the D-SUCR model with suitable candidate parameter sets, mean estimated unconfirmed cases with 95 CI, and estimated total confirmed cases with 95% CI for California, New York, Washington, and the whole USA, respectively (more results are given in the **Supplementary Material**, which can be downloaded from [https://dl2link.com/SUCR\\_supplementary\\_material.pdf](https://dl2link.com/SUCR_supplementary_material.pdf)). We can observe that the estimated confirmed cases from the D-SUCR model fit well with the historical confirmed cases under different NPIs. In addition, the percentages of the population infected in the three regions and the USA as a whole are also shown graphically (see the right axis), and these results indicate that about 16% of the USA population had been infected as of mid-May 2021. The discovery rate at time  $t$ ,  $\gamma_d(t)$ , is defined as the ratio between the value of officially confirmed cases and the estimated value of total cases, i.e.,

$$\gamma_d(t) = \frac{C(t) + R_{cm}(t)}{C(t) + R_{cm}(t) + \hat{U}(t) + \hat{R}_{um}(t)}, \tag{17}$$

where  $C(t)$  and  $R_{cm}(t)$  are the numbers of officially confirmed and removed cases, respectively, and  $\hat{U}(t)$  and  $\hat{R}_{um}(t)$  are the estimated unconfirmed and removed unconfirmed cases, respectively. The discovery rates  $\gamma_d(t)$  for California, New York, Washington, and the whole USA from Mar 2020 to May 2021 are shown in Figs. 3, respectively. These experimental results

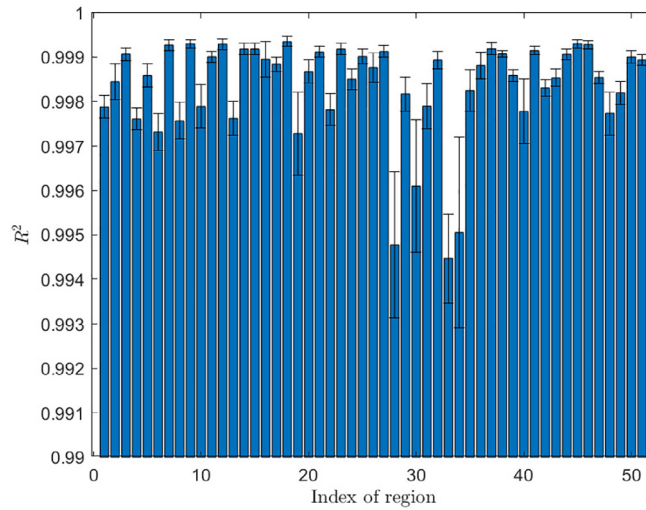


Fig. 2. Mean and 95% CI for the coefficient of determination ( $R^2$ ) for 51 regions in the USA.

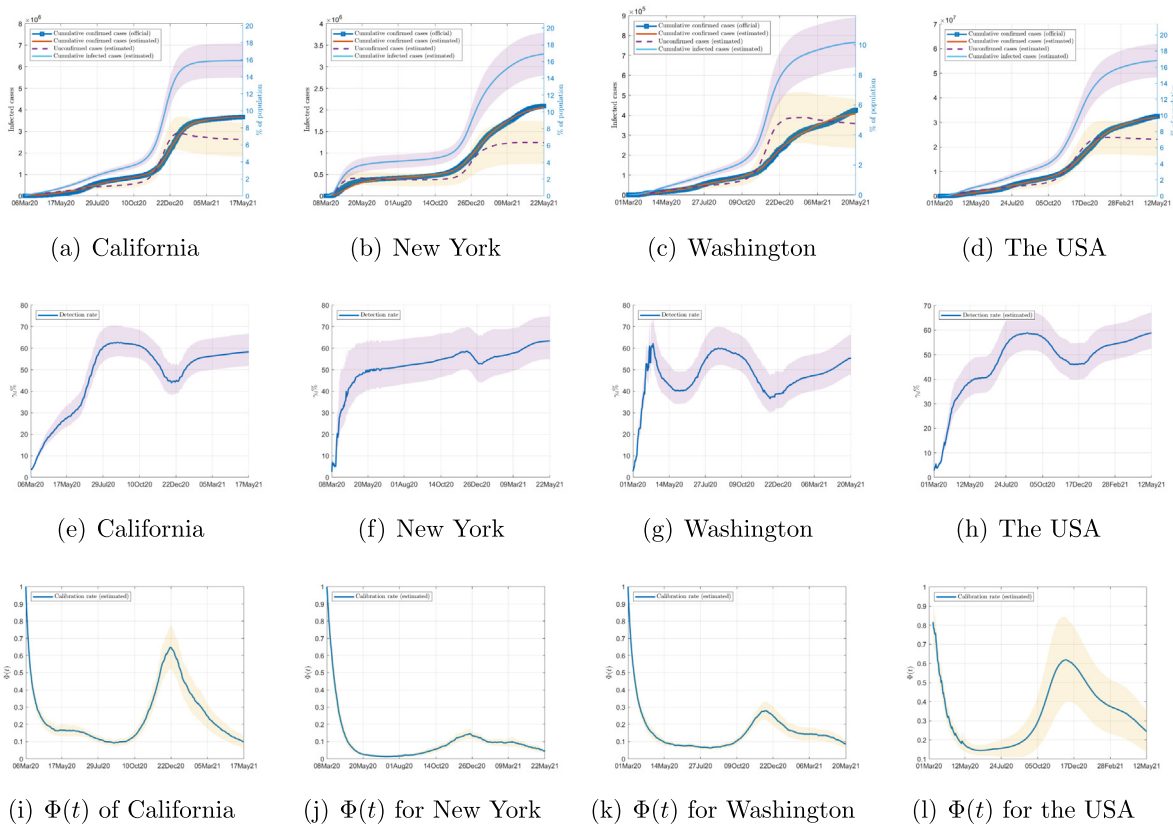


Fig. 3. Results for three states as illustrative examples and the whole USA. Lines and shaded areas represent the median and 5<sup>th</sup> to 95<sup>th</sup> percentiles, respectively, from 1,000 simulations: (a–d) Official cumulative confirmed, estimated cumulative confirmed, estimated unconfirmed and estimated total cases in California, New York, Washington, and the USA as a whole; (e–h) ratios between officially confirmed cases and the estimated total cases (%); (i–l) dynamic enhancement rates  $\Phi(t)$ .

clearly indicate that at the beginning of the pandemic (from Mar 2020 to April 2020), only about 10% of the total infections had been identified by screening. Across all simulations, our D-SUCR model suggests that there were a significant proportion of unconfirmed patients at the beginning of the pandemic outbreak. The ratio of unconfirmed to confirmed cases was more than 10. Taking New York as an example, our D-SUCR model indicates that the number of infected individuals surged from a

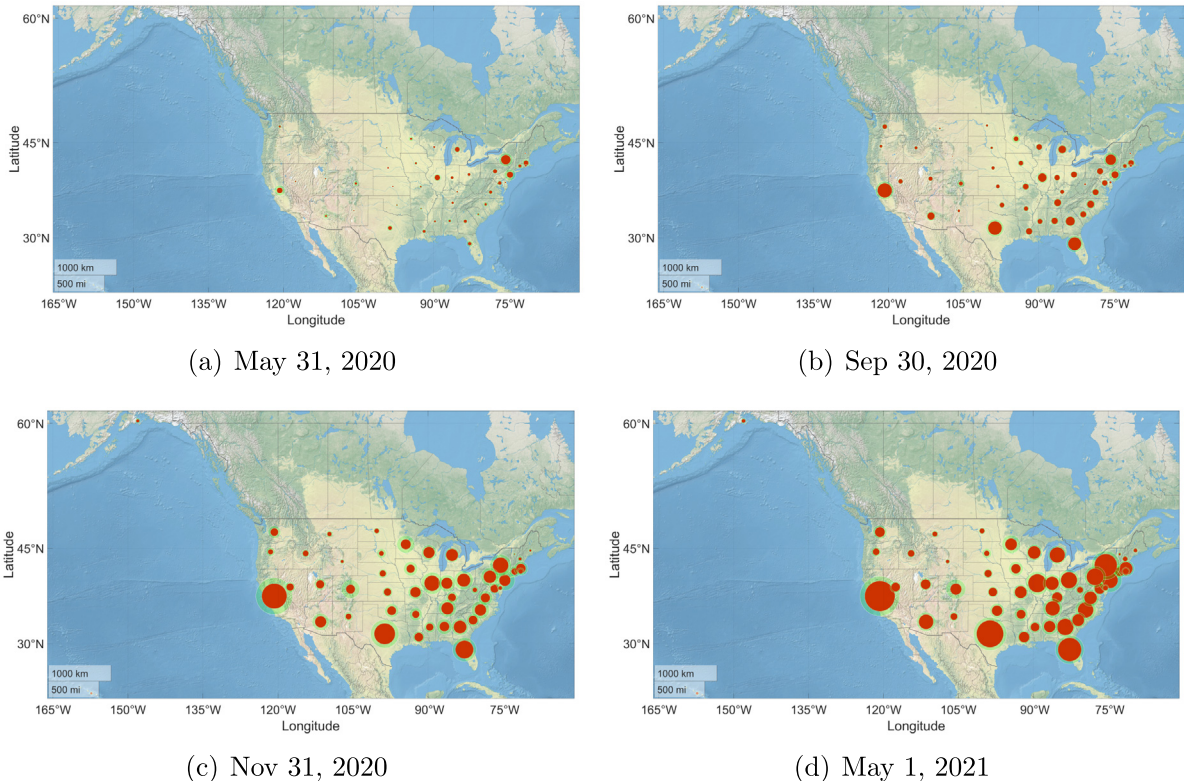
few infected individuals to 3.76% (5<sup>th</sup> to 95<sup>th</sup> percentiles: 2.63–4.89%) of the population infected (5<sup>th</sup> to 95<sup>th</sup> percentiles: 514,037–955,452) by May 31, 2020. The proportion of unconfirmed cases was very high at the beginning of the outbreak, with a mean of about 10 (5<sup>th</sup> to 95<sup>th</sup> percentiles: 8.51–11.24) times the number of confirmed cases as of mid-April 2020. An examination of the temporal dynamics shows that the discovery rate  $\gamma_d(t)$  increased dramatically with an increase in the testing capacity (see Fig. 3f). The number of confirmed, estimated unconfirmed, and total confirmed cases in California and Washington DC followed a similar trajectory, and similar results can be found for the other 48 regions in the USA. These findings suggest that at the beginning of the COVID-19 pandemic, there was a large proportion of unconfirmed cases and only a small group of COVID-19 patients were confirmed (less than 10% in the USA as a whole).

For the USA as a whole, the number of unconfirmed cases was also 10 times higher than the official number at the beginning of the outbreak. The USA government increased the testing capacity over the subsequent two months, resulting in a dramatic increase in the discovery rate. As of mid-May 2020, about 40% of the infections had been screened out. After that, the discovery rates of the 51 regions are likely to saturate at about 40–60%. This result indicates that about half of infected individuals were not confirmed, and that the actual number of infections is likely to be about twice the official number in the USA (shown in Figs. 3d and 3h). Based on the data up to mid-May 2021, the proposed D-SUCR model estimates that approximately 59.23% (California), 63.52% (New York), 54.37% (Washington), and 41.63% (the USA as a whole) of the infected cases were confirmed. The ratios of unconfirmed to actual infected cases in the four regions were 0.4077 (California), 0.3648 (New York), 0.4563 (Washington), and 0.5814 (the whole USA), respectively. Our results show that the United States had a much higher number of actual infected individuals than the official number (in fact about twice the official number), accounting for about 17% of its population. The numbers of officially confirmed cases and estimated total infections in the USA on four specific dates are shown in Figs. 4a (May 31, 2020), 4b (Sep 30, 2020), 4c (Dec 31, 2020), and 4d (May 1, 2021).

Here, at time  $t$ , we define the transmission-calibration factor as

$$\Phi(t) = \prod_{i=1}^t \phi(i). \tag{18}$$

The transmission rates with respect to unconfirmed and confirmed cases are  $\beta_u(t) = \beta_{u,0}\Phi(t)$  and  $\beta_c = \beta_{c,0}\Phi(t)$ , respectively. The smaller the value of  $\Phi(t)$ , the lower the infection rates  $\beta_u(t)$  and  $\beta_c(t)$  become. The calibration rate  $\Phi(t)$  represents the efficacy of public health strategies in terms of containing the transmission of the virus. The transmission-calibration fac-



**Fig. 4.** Numbers of officially confirmed and estimated total cases per 100 people in different regions in the USA on four specific dates: (a) May 31, 2020; (b) Sep 31, 2020; (c) Dec 31, 2020; (d) May 1, 2021. The sizes of the bubbles represent the numbers of confirmed cases.



tors for the three selected states and the whole USA are shown in Figs. 31. Taking the whole USA as an example (shown in Fig. 31), we can see that the transmission-calibration factor decreased dramatically from Mar 2020 to mid-Jul 2020, during a period in which most local authorities implemented lockdown policies. The lowest value was about 0.18, which means the NPIs worked well, resulting in an infection rate of only 0.18 times the value at the beginning of the pandemic. Then, as reopening policies were implemented by states, the transmission-calibration factor  $\Phi(t)$  gradually increased from its low values, from mid-Jul 2020 to mid-Mar 2021. By mid-Dec 2020, the transmission factor had reached its local peak (about 0.61), representing an increase of nearly 330% (from 0.18 to 0.61), compared to the value for mid-Jul 2020. The transmission-calibration factor then gradually decreased over the first half of 2021.

#### 4.2. Estimated pandemic situation in 116 countries

We then applied our D-SUCR model to an analysis of the pandemic situation in additional 116 countries. Based on data up to May 2021, we can see that most of the values of  $R^2$  for these 116 countries are larger than 0.90 (as shown in Fig. 5). Obviously, the estimated trajectory fits well with the actual situation. Fig. 6 shows the number of officially confirmed and the estimated total infections for four countries (the Philippines, Japan, Italy, and Russia) up to May 24, 2021, under the reference pandemic scenario. The historical pandemic and testing capacity data for simulations in the model were drawn from March 2020 to May 2021, and the demographic data were drawn from the Census Bureau for each region. Examples of officially released and estimated pandemic data for the four example countries are displayed in Fig. 6a–6d, which show the official number of infected individuals, mean estimated cumulative number of infected individuals with 95% CI generated by the proposed D-SUCR model, mean estimated unconfirmed cases with 95% CI, and estimated total confirmed cases with 95% CI, for the Philippines, Japan, Italy, and Russia (results for other countries are given in the **Supplementary Material**, which can be downloaded from [https://dl2link.com/SUCR\\_supplementary\\_material.pdf](https://dl2link.com/SUCR_supplementary_material.pdf)). The discovery rates  $\gamma_d(t)$  for these four countries from Mar 2020 to May 2021 are shown in Figs. 6, respectively. Our experimental results clearly indicate that at the beginning of the pandemic (between Mar and April 2020), less than 10% of the total infections were confirmed. Taking Russia as an example, our D-SUCR model shows that the outbreak surged from a few infected individuals to 7.96% (5<sup>th</sup>–95<sup>th</sup> percentiles: 6.69–9.72%) of the population infected (5<sup>th</sup>–95<sup>th</sup> percentiles: 1,011,153–1,795,077) by May 31, 2020. The proportion of unconfirmed cases was very high at the beginning of the outbreak, with a mean of more than 15 times the number of confirmed individuals on April 6, 2020. An examination of the temporal dynamics shows that the discovery rate  $\gamma_d(t)$  increased dramatically as the testing capacity increased (Fig. 6h). The numbers of confirmed, estimated unconfirmed, and total confirmed cases in the other three countries followed a similar trajectory. In particular, substantial proportions of the population were unconfirmed by Mar 6, 2020. Similar results can be found for the other countries worldwide. These findings suggest that at the beginning of the COVID-19 pandemic, there was a significant proportion of unconfirmed individuals and only a small number of infected individuals were confirmed (less than 10%) in almost all of countries. The authorities in most countries then increased the testing capacity over the next two months, resulting in a rapid increase in the discovery rate. Then, about 40% of infections had been picked up by testing in mid-May 2020. Based on data up to May 2021, the proposed D-SUCR model estimated that about 22.44% (Philippines), 43.09% (Japan), 55.83% (Italy), and 40.66% (Russia) of the total infections had been confirmed; in other words, the ratios of unconfirmed to confirmed cases were 3.4563 (Philippines), 1.3207 (Japan), 0.7912 (Italy), and 1.4594 (Russia). The numbers of officially confirmed cases and estimated total infections

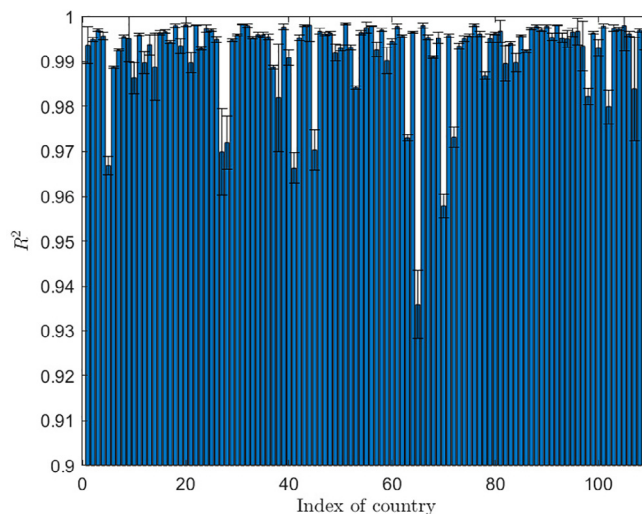
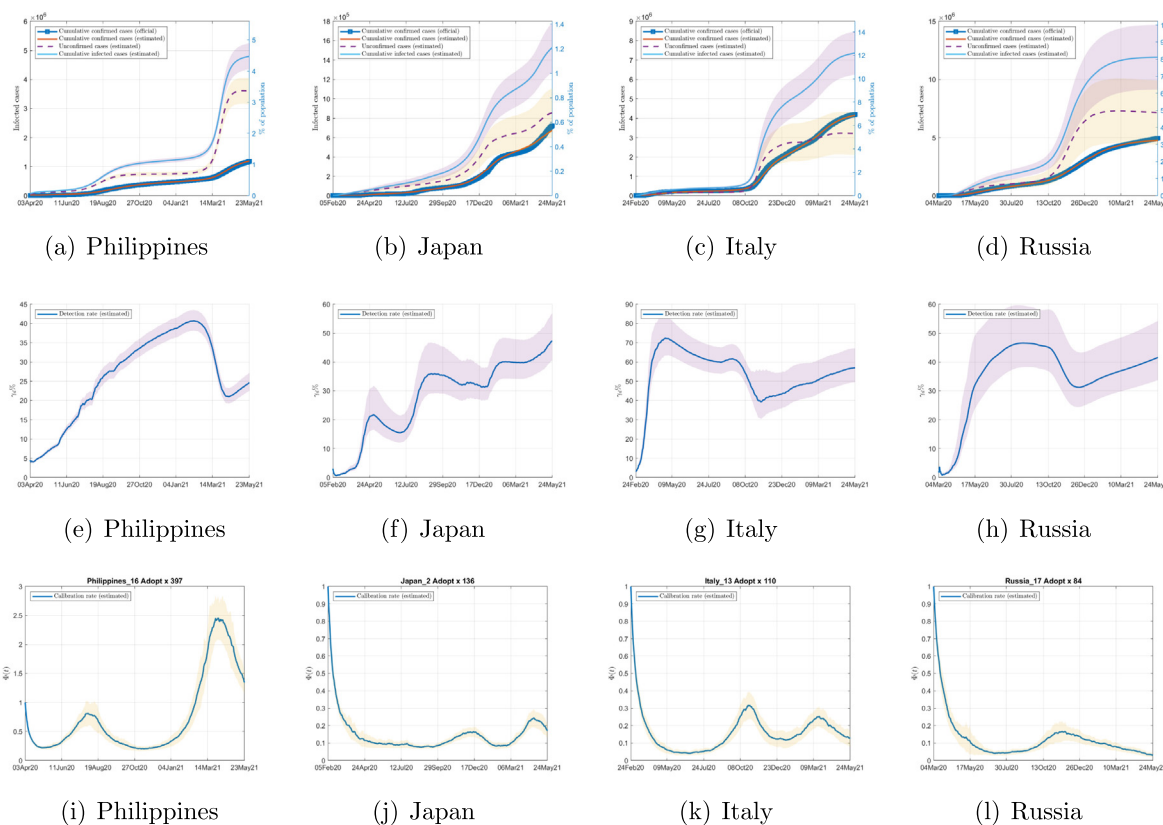


Fig. 5. Mean and 95% CI for the coefficient of determination ( $R^2$ ) for 116 countries worldwide.

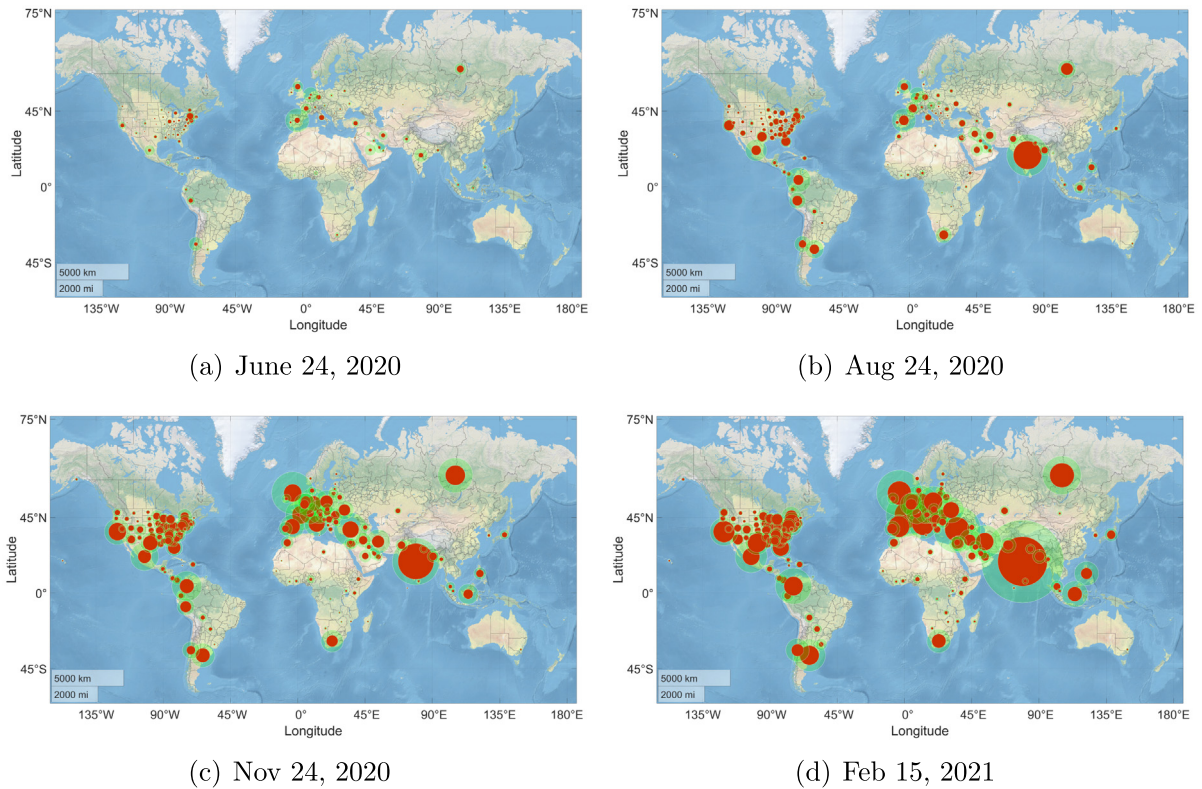


**Fig. 6.** Results for four countries (the Philippines, Japan, Italy, and Russia) as illustrative examples. Lines and shaded areas stand for the median and 5<sup>th</sup>–95<sup>th</sup> percentiles, respectively: (a–d) Official cumulative confirmed cases, estimated cumulative confirmed, estimated unconfirmed and estimated total infections for the Philippines, Japan, Italy and Russia; (e–h) ratios between the official confirmed cases and the estimated total cases (%); (i–l) dynamic enhancing rates  $\Phi(t)$ .

for 116 countries on four example dates are shown in Figs. 7a (June 24, 2020), 7b (Aug 24, 2020), 7c (Nov 24, 2020), and 7(d) (Feb 15, 2021).

The transmission-calibration factors for the four example countries are shown in Fig. 6l. Taking Russia as an example (shown in Fig. 6l), we see that the transmission-calibration factor decreased dramatically from Mar to Jul 2020, during a period in which efficient NPIs were implemented. The lowest value was about 0.08, meaning that the public health strategies worked well and the transmission rate was only 0.08 times the value at the beginning of the pandemic. Then, as reopening strategies were implemented in the country, the transmission-calibration factor  $\Phi(t)$  gradually increased from its low point in Jul 30, 2020, until Dec 24, 2020, when the transmission factor reached its local peak (about 0.21), growing by nearly 260% (from 0.08 to 0.21) in comparison with the value on Jul 30, 2020. The transmission-calibration factor then slowly decreased until Jun 2021.

We found similar results for the other countries in the world. At the beginning of the pandemic, the discovery rate was less than 10% by Apr 2020. The authorities in most countries then enhanced the testing capacity, resulting in an increase in the discovery rate. As of mid-Feb, 2021, about 40% of the infections had been picked up by screening, and the discovery rates are then likely to be stable at about 40%. This indicates that about 60% of infections were not confirmed, and that the actual number of patients was about 2.5 times the official number (as shown in Figs. 8a and 8b). The mean and 95% CIs for the transmission-calibration factors for the 116 countries are shown in Fig. 8c. It can be seen that the transmission-calibration factor decreased dramatically between the beginning of Feb 2020 and mid-Jun, 2020, during a period in which most of the local authorities in the world implemented NPIs. The lowest value was about 0.22, indicating that these NPIs worked well and the infection rate was only 0.22 times its value at the beginning of the pandemic. Then, as reopening was successively implemented in most countries, the transmission-calibration factor  $\Phi(t)$  gradually increased from its lowest value in mid-Jun 2020 to Apr 2021. In mid-Apr 2021, the transmission factor reached its local peak (about 0.86), representing an increase of about 390% (from 0.18 to 0.61) compared with its value in mid-Jun, 2020.

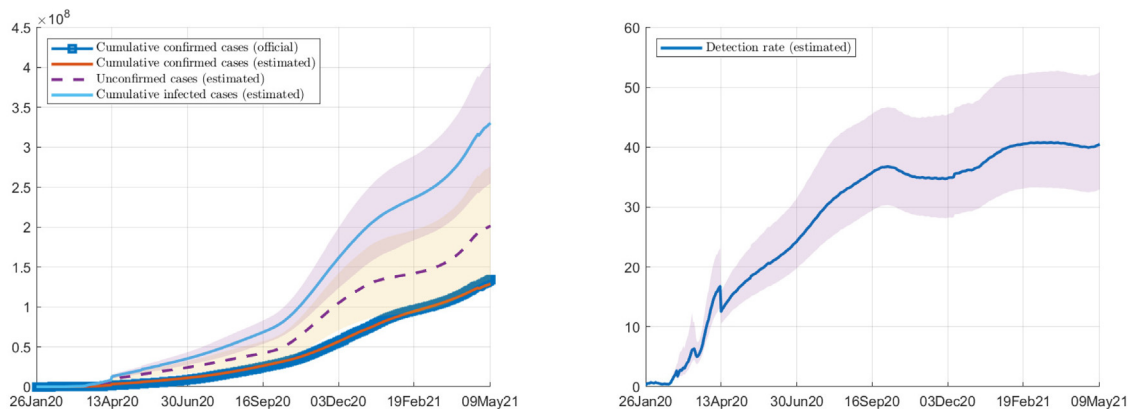


**Fig. 7.** Numbers of officially confirmed cases (blue bubbles) and estimated numbers of total infectious (transparent red bubbles) for different states in the USA on four example dates: (a) June 24, 2020; (b) Aug 24, 2020; (c) Nov 24, 2020; (d) Feb 15, 2021. The sizes of the bubbles represent the numbers of cases.

### 4.3. Simulated pandemic situation with strict control and massive tests

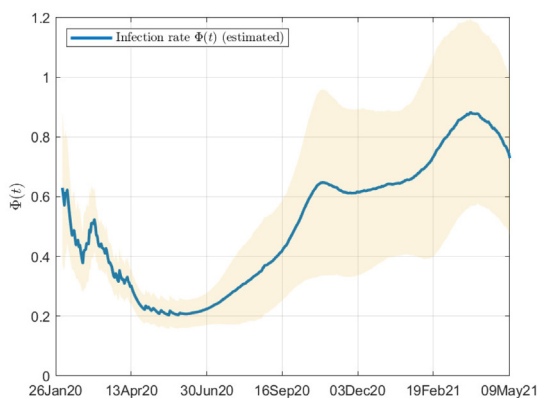
In this study, we investigate the efficacy of two public health strategies, namely, quarantining of confirmed cases, and an increase in the testing capacity. Here, we take the USA as an illustrative example, and assume that all of the states in the USA followed similar, strict quarantine strategies in which all confirmed cases were strictly quarantined and had no chance of spreading the virus, i.e.,  $\beta_c = 0$ . We assume that the authorities implemented strict quarantine measures  $T_c$  days after the detection of the first COVID-19 patient. Here, we adopt  $T_c = 1, 15, 30,$  and  $60$  days as illustrative examples, meaning that local governments implemented strict quarantine measures 1, 15, 30, and 60 days after the first COVID-19 cases was detected. The second public health strategy relies on massive testing. Despite the apparent benefits, the testing capacity remains largely suboptimal in many areas, leading to a significant number of infections, and particularly asymptomatic or mildly symptomatic infections, going unconfirmed. Here, we assume that from the beginning of Mar 2020 to mid-May 2021, the testing capacity in each region in the United States was one or five times that of the historical testing capacity, namely,  $\hat{N}_T(t) = N_T(t)$  and  $\hat{N}_T(t) = 5N_T(t)$ , where  $\hat{N}_T(t)$  is the hypothetical testing capacity. We consider eight scenarios: (i)  $T_c = 1$  day and  $\hat{N}_T(t) = N_T(t)$ ; (ii)  $T_c = 15$  days and  $\hat{N}_T(t) = N_T(t)$ ; (iii)  $T_c = 30$  days and  $\hat{N}_T(t) = N_T(t)$ ; (iv)  $T_c = 60$  days and  $\hat{N}_T(t) = N_T(t)$ ; (v)  $T_c = 1$  day and  $\hat{N}_T(t) = 5N_T(t)$ ; (vi)  $T_c = 15$  days and  $\hat{N}_T(t) = 5N_T(t)$ ; (vii)  $T_c = 30$  days and  $\hat{N}_T(t) = 5N_T(t)$ ; (viii)  $T_c = 60$  days and  $\hat{N}_T(t) = 5N_T(t)$ .

The estimated numbers of total infections, estimated confirmed cases and the 95% CI for each of these eight scenarios are summarized in Fig. 9. If the USA had implemented strict quarantine measures instantaneously ( $T_c = 1$  day) and had applied large-scale testing, the total number of infections would have decreased dramatically. In this scenario, the total number of patients would have been less than 5% of the real numbers (shown in Fig. 9a). If strict quarantine measures had been imposed immediately, the total number of infections would have been less than 10% of the real numbers (as shown in Fig. 9e). Depending on the scenario, massive testing and strict quarantine measures could have led to a dramatic reduction of between 90% to 95% in the number of overall cases (Figs. 9a and 9e). However, if the USA did not impose strict quarantine measure quickly, but implemented these measures 15 days (or more) after the first COVID-19 patient was diagnosed, the number of infections would be substantially higher than in scenarios 1 and 5 (as shown in Figs. 9b and 9f). When  $T_c = 30$  and 60 days, our experimental results indicate that the estimated total infections are much higher than those in scenarios 1 and 5. The reductions in the number of overall cases are between 10% and 60%. Fig. 9f suggests that increasing the testing



(a) Estimated total confirmed cases

(b) Ratio of official confirmed over estimated total infected (%)



(c) Estimated  $\Phi(t)$

**Fig. 8.** Statistical results for 116 countries. Lines and shaded areas stand for the median and 5<sup>th</sup>-95<sup>th</sup> percentiles, respectively: (a) Official cumulative confirmed, estimated cumulative confirmed, estimated total infections in 116 countries (%); (b) ratios between estimated total infections and officially confirmed cases; (c) dynamic enhancement rates  $\Phi(t)$ .

capacity by a factor of five with a lagged strict quarantine measure leads to a reduction of only about 25% in the number of overall cases. The amount of infected individuals can be reduced through mass testing, but implementing a strict, immediate quarantine leads to reductions of about 90% in the number of overall cases.

Fig. 10 shows the ratio of the estimated total number of infected individuals with the implementation of a strict quarantine and increased testing capacity to the actual number of infected people under real conditions. Here, we assume the testing capacity is one, two, five, and 10 times the actual testing capacity, while the period after which strict quarantine measures are implemented ranges from one to 60 days after the identification of the first COVID-19 patient. Obviously, a timely quarantine measure can significantly reduce the number of infections (to about 10% of the actual scenario), while a slightly lagging quarantine measure will not have significant efficacy (only reducing the infections to about 80% of the actual scenario, as shown in Fig. 10). These experimental results reveal that strict quarantining of confirmed cases may be effective in terms of containing the outbreak of the COVID-19 pandemic. They also suggest that in a scenario with a large group of infected individuals, only strictly quarantine confirmed cases are unable to stop the spread of the virus. As there is a large group of unconfirmed individuals, isolating only the confirmed cases would mean that unconfirmed individuals would be overlooked, and these form the main source of spreading of the virus. Hence, strict quarantine is not always an efficient measure for containing the virus, and can be efficient only when the group of infections is small. In summary, a combination of immediate, strict quarantine with massive testing is useful method of controlling the spread of disease.

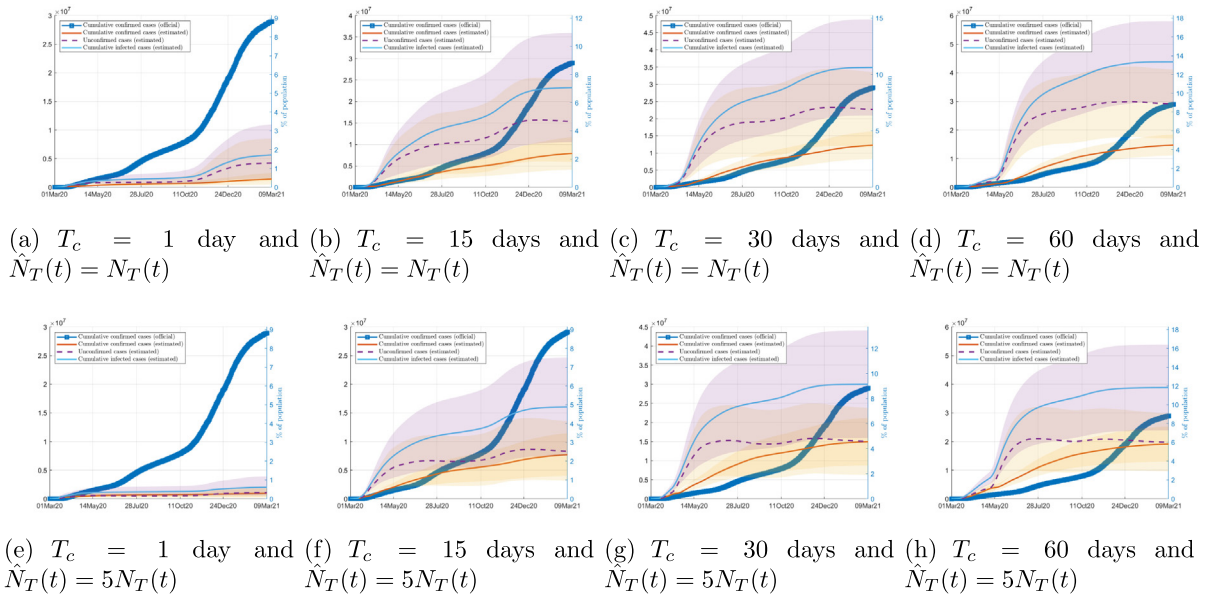


Fig. 9. The estimated total infections, confirmed, and unconfirmed cases under strict controlling measurement in the USA.

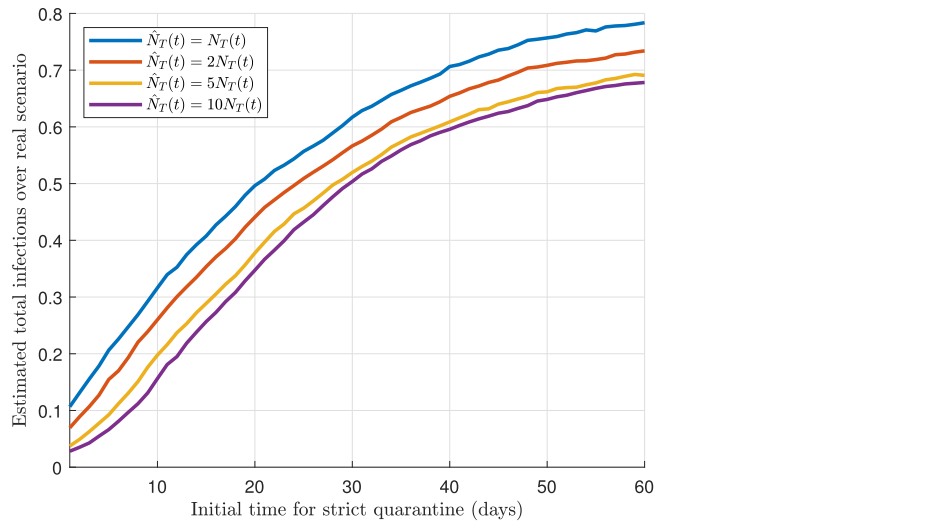


Fig. 10. Estimated numbers of total infections, confirmed, and unconfirmed cases under strict control measures in the USA.

### 5. Discussion and conclusion

Recent evidence indicates that COVID-19 is the most severe pandemic event since the influenza epidemic of 1918 [9,8]. Information about daily testing capacities and the pandemic situation can enable trend analysis to be carried out, to determine the trajectory of infections [44]. Inadequate testing capacities can hinder contact tracing and the implementation of effective NPIs to contain the early spread and dissemination of the virus [37]. Due to the strong transmissibility of COVID-19, conducting massive testing over a short period has been suggested by many researchers [36,28]. However, massive testing is difficult to achieve, due to challenging problems such as manufacturing test kits, disseminating them to the population, and the correct collection, processing, and examination of the tests. Our analysis strongly suggests that testing capacity plays an essential role in determining the pandemic situation. It is very important to estimate the actual situation of this pandemic and the real size of the infected population.

This study provides an epidemiological model that takes into consideration unconfirmed infections, testing capacity, and NPIs to analyze the transmission dynamics of the virus and evaluate potential control strategies. Epidemiological models are

widely adopted to predict the pandemic trends, and to evaluate the efficacy of different NPIs for containing the spread. This study takes advantage of both epidemiological models and machine learning methods, and combines the explain ability of epidemiological models with the data-fitting ability of an evolutionary computation algorithm. Moreover, the parameters used in our D-SUCR model can provide some insights into prevention and containment measures for COVID-19. Without strong and effective NPIs, including mask-wearing, massive testing, quarantine, travel bans, and city lockdowns, the COVID-19 pandemic would have become a catastrophe for the world. Moreover, this study shows that an epidemiological model can be used to estimate the actual pandemic situation. The experimental results show that the proposed D-SUCR model also has the ability to describe the multiple waves of COVID-19 pandemic. Note that our D-SUCR model is applicable only when each individual in a region has an equal chance of undergoing testing, and that caution is advised when using the D-SUCR model in a country or region with a dramatically uneven distribution of tests.

Traditional methods of infection control rely on symptom-based case detection and subsequent testing, and these public health measures have worked well when combating other epidemics. For example, while symptomatic detection of infected individuals is an effective test in some diseases such as SARS CoV-1, due to the presence of a large number of asymptomatic and pre-symptomatic patients in SARS CoV-2 and their ability to transmit as well, using this strategy would create a large group of unconfirmed cases for SARS CoV-2. Testing only symptomatic individuals would mean overlooking these infected individuals, who have a high level of transmissibility. Thus, massive or population-based testing may be needed irrespective of symptoms [20]. Our results indicate that testing only symptomatic patients may overlook more than 50% of the COVID-19 patients who play an important role in the transmission of the virus. In addition, we find that using strict quarantines to avoid only the spread of the disease from confirmed cases to reduce the pandemic peak may not work, and may consume a great deal of resources in terms of imposing quarantines. Limited testing capacity and a lack of draconian and immediate strategies for tracing and quarantining infected cases are two of the principal limitations in terms of containing the spread of the virus. These results clearly indicate that close contact tracing and active case detection are key factors to contain the spread of the virus. A combination of active case detection, isolation of COVID-19 cases, community quarantine, quarantine of all close contacts, contact tracing, social distancing, and even locking down an area may eradicate SARS-CoV-2. Close contact tracing is a key factor to identify and isolating asymptomatic and mild cases in order to contain outbreaks in the following periods. A strict quarantine strategy must be implemented immediately after the detection of the first few infections, and massive testing should be used to reduce the number of unconfirmed cases to allow for outbreak control. Our proposed methodological template for an epidemiological model can be generalized to model the spread of other epidemics in any territory. Knowing how to contain the spread of the virus can help us to develop effective immunological defenses and enhance health care. We hope that our model will be useful for the control and prevention of this public health pandemic worldwide.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported by the Natural Science Foundation of Guangdong Province, China (2020A 1515 010 761) and by the Key Areas R&D Program of Science and Technology Program of Guangzhou (202103010005). This work was also supported in part by the National Natural Science Foundation of China under Grant No. 61972112 and No. 61832004, the Guangdong Basic and Applied Basic Research Foundation under Grant No. 2021B1515020088, the Shenzhen Science and Technology Program under Grant No. JCYJ20210324131203009, and the HITSZ-J&A Joint Laboratory of Digital Design and Intelligent Fabrication under Grant No. HITSZ-J&A-2021A01.

### References

- [1] F. Aiano, A.A. Mensah, K. McOwat, C. Obi, A. Vusirikala, A.A. Powell, J. Flood, J. Bosowski, L. Letley, S. Jones, et al, COVID-19 outbreaks following full reopening of primary and secondary schools in England: cross-sectional national surveillance, November 2020, *The Lancet Regional Health-Europe 6* (2021) 100120.
- [2] L. Basnarkov, SEAIR Epidemic spreading model of COVID-19, *Chaos Solitons Fractals 142* (2021) 110394.
- [3] M. Biggstaff, S. Cauchemez, C. Reed, M. Gambhir, L. Finelli, Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature, *BMC Infect. Diseases 14* (1) (2014) 1–20.
- [4] F. Carrat, E. Vergu, N.M. Ferguson, M. Lemaître, S. Cauchemez, S. Leach, A.-J. Valleron, Time lines of infection and disease in human influenza: a review of volunteer challenge studies, *Am. J. Epidemiol.* 167 (7) (2008) 775–785.
- [5] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.P.Y. Piontti, K. Mu, L. Rossi, K. Sun, et al, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, *Science 368* (6489) (2020) 395–400.
- [6] M. Day, Covid-19: four fifths of cases are asymptomatic, China figures indicate, *Br. Med. J.* 369 (2020) m1375.
- [7] M. Day, Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village, *Brit. Med. J.* 368 (2020) m1165.
- [8] W. Ding, M. Abdel-Basset, H. Hawash, RCTE: A reliable and consistent temporal-ensembling framework for semi-supervised segmentation of COVID-19 lesions, *Inf. Sci.* 578 (2021) 559–573.
- [9] A. Dixit, A. Mani, R. Bansal, CoV2-Detect-Net: Design of COVID-19 prediction model based on hybrid DE-PSO with SVM using Chest X-ray images, *Inf. Sci.* 571 (2021) 676–692.
- [10] J. Feehan, V. Apostolopoulos, Is COVID-19 the worst pandemic?, *Maturitas 149* (2021) 56–58.

- [11] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science* 368 (6491) (2020) eabb6936.
- [12] M. Gandhi, D.S. Yokoe, D.V. Havlir, Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19, *N. Engl. J. Med.* 382 (22) (2020) 2158–2160.
- [13] Z. Gao, Y. Xu, C. Sun, X. Wang, Y. Guo, S. Qiu, K. Ma, A systematic review of asymptomatic infections with COVID-19, *J. Microbiol. Immunol. Infect.* 54 (1) (2021) 12–16.
- [14] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures, *Proc. Natl. Acad. Sci.* 117 (19) (2020) 10484–10491.
- [15] W.J. Guan, Z.Y. Ni, Y. Hu, W.H. Liang, C.Q. Ou, J.X. He, L. Liu, H. Shan, C.L. Lei, D.S. Hui, et al, Clinical characteristics of coronavirus disease 2019 in China, *New Engl. J. Med.* 382 (18) (2020) 1708–1720.
- [16] Z. Guang, X. Shu-Yun, L. Qin, O. Jian-Ming, Infectivity of severe acute respiratory syndrome during its incubation period, *Biomed. Environ. Sci.* 22 (6) (2009) 502–510.
- [17] X. Hao, S. Cheng, D. Wu, T. Wu, X. Lin, C. Wang, Reconstruction of the full transmission dynamics of COVID-19 in Wuhan, *Nature* 584 (7821) (2020) 420–424.
- [18] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, *Nonlinear Dyn.* 101 (3) (2020) 1667–1680.
- [19] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, et al, Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China, *Sci. China Life Sci.* 63 (5) (2020) 706–711.
- [20] H.V. Huff, A. Singh, Asymptomatic transmission during the coronavirus disease 2019 pandemic and implications for public health strategies, *Clin. Infect. Dis.* 71 (10) (2020) 2752–2756.
- [21] J.S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, N.A. Christakis, Population flow drives spatio-temporal distribution of COVID-19 in China, *Nature* 582 (7812) (2020) 389–394.
- [22] 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* 368 (6490) (2020) 493–497.
- [23] A. Kronbichler, D. Kresse, S. Yoon, K.H. Lee, M. Effenberg, J.I. Shin, Asymptomatic patients as a source of COVID-19 infections: A systematic review and meta-analysis, *Int. J. Infect. Dis.* 98 (2020) 180–186.
- [24] S. Lai, N.W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J.R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, Effect of non-pharmaceutical interventions to contain COVID-19 in China, *Nature* 585 (7825) (2020) 410–413.
- [25] J. Lessler, N.G. Reich, R. Brookmeyer, T.M. Perl, K.E. Nelson, D.A. Cummings, Incubation periods of acute respiratory viral infections: a systematic review, *Lancet Infect. Dis.* 9 (5) (2009) 291–300.
- [26] M.Y. Li, J.S. Muldowney, Global stability for the SEIR model in epidemiology, *Math. Biosci.* 125 (2) (1995) 155–164.
- [27] 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, *N. Engl. J. Med.* 382 (13) (2020) 1199–1207.
- [28] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2), *Science* 368 (6490) (2020) 489–493.
- [29] N.M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A.R. Akhmetzhanov, S.-M. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data, *J. Clin. Med.* 9 (2) (2020) 538.
- [30] J. Liu, X. Liao, S. Qian, J. Yuan, F. Wang, Y. Liu, Z. Wang, F.-S. Wang, L. Liu, Z. Zhang, Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020, *Emerg. Infect. Dis.* 26(6) (2020) 1320–1323.
- [31] Y. Liu, A.A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus, *J. Travel Med.* (2020) 1–4.
- [32] Q.X. Long, X.J. Tang, Q.L. Shi, Q. Li, H.J. Deng, J. Yuan, J.L. Hu, W. Xu, Y. Zhang, F.J. Lv, et al, Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections, *Nat. Med.* 26 (8) (2020) 1200–1204.
- [33] L. López, X. Rodo, A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics, *Results Phys.* 21 (2021) 103746.
- [34] K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020, *Eurosurveillance* 25 (10) (2020) 2000180.
- [35] H. Nishiura, T. Kobayashi, T. Miyama, A. Suzuki, S.-M. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A.R. Akhmetzhanov, et al, Estimation of the asymptomatic ratio of novel coronavirus infections, COVID-19, *Int. J. Infect. Dis.* 94 (2020) 154–155.
- [36] J. Peto, Covid-19 mass testing facilities could end the epidemic rapidly, *Br. Med. J.* 368 (2020) m1163.
- [37] 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 medical weekly* 150 (11–12) (2020), w20225.
- [38] C. Stein-Zamir, N. Abramson, H. Shoob, E. Libal, M. Bitan, T. Cardash, R. Cayam, I. Miskin, A large COVID-19 outbreak in a high school 10 days after schools' reopening, Israel, May 2020, *Eurosurveillance* 25 (29) (2020) 2001352.
- [39] V. Sypsa, S. Roussos, D. Paraskevis, T. Lytras, S. Tsioufas, A. Hatzakis, Effects of social distancing measures during the first epidemic wave of severe acute respiratory syndrome infection, Greece, *Emerg. Infect. Dis.* 27 (2) (2021) 452.
- [40] 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* 368 (6491) (2020) 638–642.
- [41] 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, T.S.H. Chik, et al, Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study, *Lancet. Infect. Dis.* 20 (5) (2020) 565–574.
- [42] Y. Wang, H. Kang, X. Liu, Z. Tong, Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak, *J. Med. Virol.* 92 (6) (2020) 538–539.
- [43] R. Wölfel, V.M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M.A. Müller, D. Niemeyer, T.C. Jones, P. Vollmar, C. Rothe, et al, Virological assessment of hospitalized patients with COVID-2019, *Nature* 581 (7809) (2020) 465–469.
- [44] World Health Organization, COVID-19 strategy update 2020. <https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf>, 2020.
- [45] C. Zhan, J. Chen, H. Zhang, An investigation of testing capacity for evaluating and modeling the spread of coronavirus disease, *Inf. Sci.* 561 (2021) 211–229.
- [46] C. Zhan, C.K. Tse, Z. Lai, X. Chen, M. Mo, General Model for COVID-19 Spreading with Consideration of Intercity Migration, Insufficient Testing and Active Intervention: Application to Study of Pandemic Progression in Japan and USA, *JMIR Public Health Surveillance* 6 (3) (2020) e18880.
- [47] C. Zhan, C.K. Tse, Z. Lai, T. Hao, J. Su, Prediction of COVID-19 spreading profiles in South Korea, Italy and Iran by data-driven coding, *PLoS one* (2020), <https://doi.org/10.1371/journal.pone.0234763>.
- [48] C. Zhan, W. Situ, L.F. Yeung, P.W.M. Tsang, G. Yang, A parameter estimation method for biological systems modelled by ODE/DDE models using spline approximation and differential evolution algorithm, *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 11 (6) (2014) 1066–1076.
- [49] C. Zhan, Y. Zheng, Z. Lai, T. Hao, B. Li, Identifying epidemic spreading dynamics of COVID-19 by pseudoco-evolutionary simulated annealing optimizers, *Neural Comput. Appl.* 33 (10) (2021) 1–14.
- [50] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, et al, A novel coronavirus from patients with pneumonia in China, 2019, *New Engl. J. Med.* 382 (8) (2020) 727–733.