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.
Transmission dynamics and control methodology of COVID-19: A modeling study

Hongjun Zhu\textsuperscript{a,c,1,*}, Yan Li\textsuperscript{b,1}, Xuelian Jin\textsuperscript{d}, Jiangping Huang\textsuperscript{a,c}, Xin Liu\textsuperscript{a,c}, Ying Qian\textsuperscript{a,c,4}, Jindong Tan\textsuperscript{b}

\textsuperscript{a}School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
\textsuperscript{b}Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville 37919, USA
\textsuperscript{c}Chongqing Engineering Research Center of Software Quality Assurance, Testing and Assessment, Chongqing 400065, China
\textsuperscript{d}College of Media and Arts, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

\textbf{A R T I C L E   I N F O}

Article history:
Received 6 August 2020
Accepted 14 August 2020
Available online 21 September 2020

Keywords:
COVID-19
Transmission dynamics
Non-linear optimization
SEIR-HC model
Prevention and control

\textbf{A B S T R A C T}

The coronavirus disease 2019 (COVID-19) has grown up to be a pandemic within a short span of time. To investigate transmission dynamics and then determine control methodology, we took epidemic in Wuhan as a study case. Unfortunately, to our best knowledge, the existing models are based on the common assumption that the total population follows a homogeneous spatial distribution, which is not the case for the prevalence occurred both in the community and in hospital due to the difference in the contact rate. To solve this problem, we propose a novel epidemic model called SEIR-HC, which is a model with two different social circles (i.e., individuals in hospital and community). Using the model alongside the exclusive optimization algorithm, the spread process of COVID-19 epidemic in Wuhan city is reproduced and then the propagation characteristics and unknown data are estimated. The basic reproduction number of COVID-19 is estimated to be 7.9, which is far higher than that of the severe acute respiratory syndrome (SARS). Furthermore, the control measures implemented in Wuhan are assessed and the control methodology of COVID-19 is discussed to provide guidance for limiting the epidemic spread.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

A total of 113,702 confirmed cases have been reported worldwide as of 10:00 CET on Mar. 10. The statistical data shows that the COVID-19 outbreak constitutes an enormous threat to human in the world. Many factors, including human connectivity, urbanization, as well as international travel, pose difficulties for prevention and control of COVID-19 [1]. Fortunately, mathematical models offer valuable tools for understanding epidemiological patterns and then for decision-making in global health.

However, modeling the transmission dynamics of epidemic is still a challenging task, since the usefulness of a mathematical model depends on the existence of its solution [2] A prime difficulty is to obtain the reliable data due to the fact that the available data is often patchy, delayed, and even wrong [3–6]. The second cause is that the classical epidemic models describe epidemic transmission in the absence of interventions, which scarcely occurs in the real world. Daily activities

\textsuperscript{*} Corresponding authors at: School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.
\textsuperscript{1} E-mail addresses: zhuhj@cqupt.edu.cn (H. Zhu), qianying@cqupt.edu.cn (Y. Qian).
\textsuperscript{4} Hongjun Zhu, Yan Li contributed equally to this work.

https://doi.org/10.1016/j.apm.2020.08.056
0307-904X© 2020 Elsevier Inc. All rights reserved.
and travel, which are tightly linked to the spread of infection, tend to make the case more complex. The last but not least reason is that there are too many incidences, which cannot be ignored anyhow, happened in hospital where the contact rate is entirely different from that in the community. In this case, the general epidemic models do not work. Additional problem encountered here is related to the estimation of the parameters of the epidemic model [2].

In order to probe propagation characteristics and transmission mechanism, several attempts have been done on modeling the transmission dynamics of COVID-19, using observations of the index case in Wuhan [7–10]. Unfortunately, the models fail to consider the difference of contact rate between in communities and in hospitals. This may limit the accuracy of the proposed models and hence the reliability of results. And hence, the results are very different from one another and even contradictory. Most of them indicate that the basic reproduction number of COVID-19 is lower than that of SARS. On the contrary, the number of the individuals infected by COVID-19 to date has been far higher than that infected by SARS (8422) in 2003 [11]. Moreover, there is no comparison of model output with real-world observations, though such a comparison is necessary to establish the performance of the models.

To resolve this problem, this paper presents a novel epidemic model, which based on the standard assumption: the population is divided into susceptible (S), exposed (E), infectious (I), and recovered (R) groups [12]. The model proposed here, referred to as SEIR-HC for simplicity, extends the susceptible–exposed–infectious–removed (SEIR) model to handle epidemic transmission in two different but coupled social circles. Furthermore, a two-step optimization is built for the parameter estimation.

With the limited data, transmission process of COVID-19 in Wuhan city reproduced by the proposed model. Then, the propagation characteristics and unreported data are estimated, and the end time of COVID-19 is also predicted. In the end, the control measures implemented in Wuhan are assessed and the control methodology of COVID-19 is discussed to provide guidance to cease the COVID-19 outbreak.

The remainder of the paper is structured as follows: Section 2 introduces the previous work; Section 3 defines the related terminology; Section 4 explains the SEIR-HC model in detail; Section 5 describes the two-step optimization for parameter estimation; Section 6 shows the analysis results, and finally, Section 7 states the conclusions.

2. Previous work

Transmission dynamics of an epidemic disease have long been a topic of active research. In 1866, Farr [13] mathematically described the cattle-plague epidemic by means of curve fitting method. In 1889, P. D. En’ko presented a discrete-time model in which the population consists of infectious individuals and susceptible individuals [14]. In 1906, Hamer [15] proposed that the incidence depends on the product of the densities of susceptibles and infectives. The susceptible–infectious–removed (SIR) model was formulated by Martini in 1921 [16,17]. In 1927, Kermack and McKendrick [18] investigated the SIR model in a homogeneous population by using differential equations and discovered that the epidemic course would inevitably be terminated once the density of population becomes smaller than a critical threshold. Stone et al. [19] analyzed the rationale of the pulse vaccination strategy with the SIR model.

The SIR model is ideally suited to describe the process of virus spread. However, it is not always congruent with epidemic course. Some infections do not confer the lasting immunity. In this case, one should resort to the SIS model, in which the infectious becomes susceptible again, rather than recovered. The SIS model has many potential applications such as modeling the spread behavior of computer viruses [20]. On the other hand, if the epidemic lasts for a very long time, births and deaths heavily affect the population size. So, Hethcote [21] introduced births and deaths into the deterministic models of SIS and SIR.

On account of the fact that the SIS and SIR model only hold for the case without a latent period, which is not the case for many kinds of infectious diseases. To alleviate this problem, Cooke [22] proposed an epidemic model for the case that after a fixed time the infected susceptibles can get infectious. In this model, the population is divided into four classes of individuals: susceptibles, exposed individuals, infectives and recovered individuals. Such an epidemic model is known as the SEIR model. Later, Longini [23] presented a general formulation of the discrete-time epidemic model with permanent immunity. Lipsitch et al. [24] estimated the infectiousness of SARS and likelihood of an outbreak with modified SEIR model. Krylova and Earn [17] found that the dynamical structures of SEIR models have a less effect on the stage duration distribution, relative to that of the SIR models. Indeed, a family of SEIR models is being used to support epidemic control, elimination, and eradication efforts. See [14,25] for an extensive review of classic epidemic models.

During the disease spread, a high degree of chance enters into the conditions under which fresh infections take place. Therefore, statistical fluctuations should be taken into account for a more precise analysis [26]. For this reason, Bailey [27] introduced probability distribution into the SIS epidemic model. In the standard models, the incubation and infectious periods are typically assumed to be exponential distribution, which makes the model sensitive to stochastic fluctuations [12,28]. However, because of more robustness, Weibull distribution is used to investigate SARS by Marc Lipsitch et al [24].

For the stochastic case, even the simplest representations present difficulties in obtaining algebraic solutions [2]. To address this issue, Saunders [29] constructed an approximate maximum likelihood estimator for the chain SEIR model by using the Poisson approximation to the Binomial distribution.

Recently, Wu et al. [10] provided an estimate of the size of the COVID-19 epidemic based on the SEIR model. Using an assumption of Poisson-distributed daily time increments, Read et al. [9] fitted a deterministic SEIR model to the daily number of confirmed cases in Chinese cities and cases reported in other countries/regions. Zhao et al. [5] estimated the number of
unreported cases in mainland China based on the assumption that the initial growth phase followed an exponential growth pattern.

To sum up, a number of models have been proposed to formulate the transmission process of epidemic disease, which lay a good foundation for our work. Unfortunately, none of these studies have been done on the epidemic transmission under two different but coupled conditions. Such a mathematical model is desired to explore, understand, predict and anticipate COVID-19, including changes caused by intervention.

3. Terminology

In many cases, canonical representation will be the starting point for obtaining a clear or concise expression. For this reason, let us introduce the notation used throughout the paper before describing our work. Fig. 1 presents a pictorial display of disease natural history in order to make these terms easy to be understood.

(1) latent period: time from exposure to onset of infectiousness, during which the infectious agent develops in the vector, and at the end of which the vector can infect a susceptible individuals [22];
(2) Incubation period: time from exposure to first appearance of clinical symptoms of infection [25];
(3) Infectious period: period during which an infected person can transmit a pathogen to a susceptible [10];
(4) Length of stay: time from the day of admission in the hospital to the day of discharge, i.e., the number of days a patient stayed in a hospital for treatment [30];
(5) Serial interval: time from the onset of symptoms in an index case to the onset of symptoms in a subsequent case infected by the index patient [24];
(6) Susceptibles: individuals who have possibility of contacting with infectious individuals but still stay uninfected [18];
(7) Exposed individuals: individuals in the latent period, who are infected but not yet infectious [25];
(8) Infectives: individuals who are infectious in the sense that they are capable of transmitting the infection [25];
(9) Removed individuals: individuals who are removed by recovery and death, among them recovers obtain permanent infection-acquired immunity [18,25];
(10) Quarantined individuals: suspected or exposed individuals who are separated and controlled to see if they become sick [24];
(11) Isolated individuals: infections who are separated and controlled to avoid disease transmission [24];
(12) Basic reproduction number: expected number of secondary infectious cases generated by an average infectious case in an entirely susceptible population [24];
(13) Contact rate: average number of individuals with which one infective have an adequate contact in unit time, where an adequate contact is an interaction which results in infection [25,31];
(14) Incidence rate: rate of new infections [32], or, more precisely, total number of exposed individuals who move into infective class in unit time;
(15) Death rate: the death probability of a person per day on average [18].

4. SEIR-HC epidemic model

Based on the basic principle that if the model is tightly close to the real world, then the optimization algorithm will converge to the most reasonable solution, we build a new SEIR model with intervention mechanism which takes two different social circles into consideration.

4.1. Assumptions

To estimate the parameters, the model of a complex physical situation tends to involve a certain amount of simplification for real world applications. However, it is not straightforward to obtain a balance between simplicity and practicality. To
resolve this dilemma, generality of epidemic and particularity of COVID-19 are considered simultaneously in our work and then a set of assumptions are determined carefully as follows.

1. The population is homogeneous and uniformly mixed [22,31,33].
2. Recovered individuals are permanently immune and newborn infants have temporary passive immunity to the infection [25].
3. Infectiousness remains constant during an infectious period [25].
4. The natural disease-independent death rate is constant throughout the population [28].
5. The disease-caused death rate is a time-independent constant.
6. Latent period and infectious period follow Weibull distribution [24], which is a versatile distribution that has the ability to take on the characteristics of other types of distribution.
7. Contact rate is constant over the entire infectious period [23].
8. The first index case is infected on Dec 1, 2019 [10].
9. Travel behavior was not affected by disease before lockdown on Jan 23,2020 [10].

4.2. Symbols

Consider a time interval $[t, t+h]$, where $h$ represents the length between the time points at which measurements are taken, here $h = 1$ day. For convenience, a variable $X$ at a time interval $(t, t+1]$ is represented as $X_{(t)}$. Then, the variables and parameters are denoted as follows:

- $S_{(t)}$: number of susceptibles at time $t$;
- $E_{(t), t}$: number of exposed individuals at time $t$ who are first infected at time $t - \tau$;
- $E_{(t)}$: total number of exposed individuals at time $t$;
- $I_{(t), t}$: number of infectives at time $t$ whose infectiousness starts at time $t - \tau$;
- $I_{(t)}$: total number of infectives at time $t$;
- $R_{(t)}$: number of removed individuals at time $t$;
- $Z_{(t)}$: number of infectives which is equivalent to force of infection of wildlife in the Huanan Seafood Wholesale Market;
- $N_{(t)}$: size of population at time $t$, that is, total number of susceptibles, exposed individuals, infectives and removed individuals at time $t$;
- $C_{\text{in}}(t)$: number of inbound travellers every day in Wuhan at time $t$;
- $C_{\text{out}}(t)$: number of outbound travellers every day in Wuhan at time $t$;
- $\alpha_c$: contact rate in the community;
- $\alpha_h$: contact rate in hospitals;
- $\beta_{(t)}$: incidence rate of the exposed individuals who are infected $\tau$ days ago, which follows the Weibull distribution;
- $k_\beta$: shape parameter of the Weibull probability density function (PDF) of incidence rate $\beta$;
- $\lambda_\beta$: scale parameter of the Weibull PDF of the incidence rate $\beta$;
- $\gamma_{(t)}$: removal rate of infectives by disease-caused death or recovery who have been infectious for $\tau$ days, which follows the Weibull distribution;
- $k_\gamma$: shape parameter of the Weibull PDF of the removal rate $\gamma$;
- $\lambda_\gamma$: scale parameter of the Weibull PDF of the removal rate $\gamma$;
- $\xi$: proportion of hospitalized infectives to total number of infectives;
- $\zeta$: proportion of quarantined susceptibles to total number of susceptibles;
- $P_L$: maximum of latent period;
- $P_I$: maximum of infectious period;
- $\nu$: disease-independent death rate.

For notational convenience, index c is used to denote community and h hospital in the following expressions. The variables without the indices c and h mean they are applicable in both cases.

4.3. Formulating SEIR-HC epidemic model

From classical SEIR model, the population is roughly classified as four classes: susceptible, exposed, infectious, and recovered individuals. Among them, exposed and infectious individuals fall into a series of groups according to disease progression in our work so that the Weibull distribution, which armed with high generalization capability [34], can match accurately number of individuals and duration time.

In order to accommodate the quarantine and isolation measures, and take the infectivity difference between hospital and community into consideration, the standard SEIR structure should be modified as shown in Fig. 2. Note that the size of population is varied with control measures and is also affected by the inbound and outbound travellers.

For the individuals respectively in hospital and community, disease transmissions are different in infection pattern but share the common nature of the virus. For this reason, two populations are considered simultaneously and analyzed separately for the sake of accuracy. With the reservations mentioned in Section 4.1, we used the SEIR-HC model to simulate the
epidemic process in Wuhan. From the rules of node dissemination, the dynamic transfer equations of the SEIR-HC model are stated as follows.

For the individuals in the community,

\[ S_{c(t+1)} = S_{c(t)} - \frac{\alpha_c S_{c(t)} \left(1 - \xi \right) I_{c(t)} + Z_{c(t)}}{N_{c(t)}} + C_{in} - C_{out} \frac{S_{c(t)}}{N_{c(t)}} - \nu S_{c(t)}, \]

\[ E_{c(t+1)} = E_{c(t)} + \frac{\alpha_c S_{c(t)} \left(1 - \xi \right) I_{c(t)} + Z_{c(t)}}{N_{c(t)}} - \sum_{\tau=0}^{\infty} \beta(\tau) E_{c(0,t-\tau)} - C_{out} \frac{E_{c(t)}}{N_{c(t)}} - \nu E_{c(t)}, \]

\[ I_{c(t+1)} = I_{c(t)} + \sum_{\tau=0}^{\infty} \beta(\tau) E_{c(0,t-\tau)} - \sum_{\tau=0}^{\infty} \gamma(\tau) I_{c(0,t-\tau)} - C_{out} \frac{I_{c(t)}}{N_{c(t)}} - \nu I_{c(t)}, \]

\[ R_{c(t+1)} = R_{c(t)} + \sum_{\tau=0}^{\infty} \gamma(\tau) I_{c(0,t-\tau)} - C_{out} \frac{R_{c(t)}}{N_{c(t)}}, \]

\[ N_{c(t)} = S_{c(t)} + E_{c(t)} + I_{c(t)} + R_{c(t)}. \]

Given contact rate \( \alpha \), \( \alpha S_{c(t)}/N_{c(t)} \) is the average number of susceptibles with which per infective contacts in unit time, and thus \( \alpha S_{c(t)}/N_{c(t)} \) is the total number of susceptibles with which \( I_{c(t)} \) infectives contacts in unit time (i.e., the total number of new infections in unit time). Notice that the population of Wuhan city was suddenly reduced from 14 million to 9 million before lockdown on the morning of Jan 23, 2020. On the day, the preventive and control measures of category A infectious diseases were implemented to fight against COVID-19. Therefore, we assume that \( N_{c(t)}, S_{c(t)}, E_{c(t)}, I_{c(t)}, R_{c(t)} \) decrease proportionately by 5/14 and then the number of susceptibles became \( 5/14 N_{c(t)} \) for restrictions on outdoor activities on the same day. We further assume that the disease-independent death rate \( \nu \) is 1.50959 \times 10^{-3}, which is the same as that reported by the Wuhan government in December 2019 [35]. According to the data presented by Wu et al [10], \( C_{in}, C_{out} \) are set to 490,856, 505,646 before Jan 10, 814,046, 720,859 from Jan 11 to Jan 22, and 0, 0 after Jan 22, respectively.

For the individuals in hospital,

\[ S_{h(t+1)} = \xi I_{h(t)} + h_{h(t)}, \]

\[ E_{h(t+1)} = E_{h(t)} + \frac{\alpha_h S_{h(t)} \xi I_{h(t)}}{S_{h(t)} + \xi I_{h(t)}} - \sum_{\tau=0}^{\infty} \beta(\tau) E_{h(0,t-\tau)} - \nu E_{h(t)}, \]

\[ I_{h(t+1)} = I_{h(t)} + \sum_{\tau=0}^{\infty} \beta(\tau) E_{h(0,t-\tau)} - \sum_{\tau=0}^{\infty} \gamma(\tau) I_{h(0,t-\tau)} - \nu I_{h(t)}, \]
The number of new cases every day from Dec 1, 2019, to Jan 8, 2020 [4].

\[ R_{0(t+1)} = R_{0(t)} + \sum_{\tau=0}^{n} I_{t}(\tau) I_{0(t-\tau)}, \tag{9} \]

where

\[ E_{t}(\tau) = \sum_{\tau=0}^{p} E_{t}(\tau), \tag{10} \]

\[ I_{t}(\tau) = \sum_{\tau=0}^{l} I_{t}(\tau), \tag{11} \]

\[ \beta_{t}(\tau) = \int_{\tau}^{\tau+1} \frac{x}{\lambda} e^{-x/(\lambda x)} \left( \frac{x}{\lambda} \right)^{k_r-1} dx, \quad \tau \in [0, 100). \tag{12} \]

\[ \gamma_{t}(\tau) = \int_{\tau}^{\tau+1} \frac{x}{\lambda} e^{-x/(\lambda x)} \left( \frac{x}{\lambda} \right)^{k_r-1} dx, \quad \tau \in [0, 100). \tag{13} \]

\[ Z_{t}(\tau) = \begin{cases} z & \text{before Jan 1, 2020} \\ 0 & \text{others} \end{cases} \tag{14} \]

Depending on the data provided by Li et al (in Fig. 3), there are no more than 5 new cases every day from Dec 8 to Dec 28. At the same time, the basic reproduction number of the COVID-19 in Wuhan must be more than 1 or else the outbreak is impossible [24,36]. In addition, a susceptible can be infected within 15 s of standing next to an infective. For these reasons, it is highly probable that the parameter \( z \), which reflects the force of infection of the Huanan Seafood Market, is small and hence let \( z = 1 \).

As defined in Eq. (6), the number of susceptibles in hospital is equal to the number of infectives stayed in hospital. In fact, \( N_{i(t)} + I_{h(t)} \) in (6) is substituted by the number of the hospitalized patients in Wuhan city, which is reported by Wuhan Municipal Health Commission (WMHC) [37] and Health Commission of Hubei Province (HCHP) [38]. In addition, the execution of (12) and (13) is a time-consuming process. So, they are reformulated by

\[ \beta_{t}(\tau) = e^{\left( \frac{-x}{\tau} \right)^{k_r}} - e^{\left( \frac{-x}{\tau} \right)^{k_r}}, \tag{15} \]

\[ \gamma_{t}(\tau) = e^{\left( \frac{-x}{\tau} \right)^{k_r}} - e^{\left( \frac{-x}{\tau} \right)^{k_r}}. \tag{16} \]

It is readily seen from Eq. (10) that the exposed individuals can be classified as belonging to one of \( P_{E} + 1 \) groups. And, the sizes of these groups are changed every day. The update formula is

\[ E_{t}(\tau) = \begin{cases} \alpha S_{t(t-1)} \left( 1 - E_{t(t-1)} + Z_{t(t-1)} \right) / N_{t(t-1)} & \text{if } \tau = 0 \\ E_{t(t-1)} - \beta_{t(\tau-1)} E_{t(\tau-\tau)} - C_{out} E_{t(\tau-\tau)} N_{t(t-1)} - \nu E_{t(\tau-\tau-1)} & \text{others} \end{cases} \tag{17} \]

Similarly, there are
\( E_h(t, t) = \begin{cases} \frac{\alpha_3 S_h(t) I(t)}{N_{SEIR}}, & \text{if } \tau = 0 \\ E_h(t-1, t-1) - \beta(t-1) E_h(0, t-\tau) - \nu E_h(t-1, t-1), & \text{otherwise.} \end{cases} \) (18)

\( I_c(t, t) = \begin{cases} \sum_{t=0}^n \beta(t) \epsilon_c(0, t-1) \\ \frac{\epsilon_c(0, t-1) - \gamma(t-1) \epsilon_c(0, t-1)}{\epsilon_c(0, t-1)} - \frac{\epsilon_c(t-1, t-1)}{\epsilon_c(t-1, t-1)} + \nu \epsilon_c(t-1, t-1), & \text{if } \tau = 0 \\ \text{otherwise.} \end{cases} \) (19)

\( I_h(t, t) = \begin{cases} \sum_{t=0}^n \beta(t) \epsilon_h(t, t-1) \\ \frac{\epsilon_h(t, t-1) - \gamma(t-1) \epsilon_h(t, t-1)}{\epsilon_h(t, t-1)} - \frac{\epsilon_h(t-1, t-1)}{\epsilon_h(t-1, t-1)} + \nu \epsilon_h(t-1, t-1), & \text{if } \tau = 0 \\ \text{otherwise.} \end{cases} \) (20)

5. Solution of parameters

In this section, we propose a solution based on constraint optimization to estimate the parameters of the SEIR-HC model mentioned in Section 4.

5.1. Objective function

The epidemic model is devised to estimate the unobserved variables and to predict the transmission process. The output data of the model is desired to be close to the real. Unfortunately, it is difficult for the lack of the credible observations. To alleviate this problem, reasonable data is considered as an alternative. The data collection is completed by integrating multiple data sources:

1. Number of new cases every day from Dec 1, 2019, to Jan 8, 2020, provided by Li et al (in Fig. 3) [4],
2. Number of new cases every day from Feb 20 to Mar 10, 2020, reported by WMHC [37] and HCHP [38],
3. Number of the infectives among the nationals who returned to America [39], Japan [40], South Korea [41] and Singapore [42] from Jan 29 to Jan 31, 2020 (as tabulated in Table 1),
4. Number of the infectious hospital staff members is 1102 by Feb 11, 2020, reported by The State Council Information Office of China [43].
5. Numbers of discharged and dead patients reported by WMHC [37] and HCHP [38], sum of which is theoretically equal to the number of the individuals removed by both recovery and death. Note that, among them, several numbers were corrected for the contradiction between accumulated cases and new cases.

Based on the above data, the objective function consists of the five error sums of squares as follows:

\[ F = ESS(1) + ESS(2) + ESS(3) + ESS(4) + ESS(5) \] (21)

where

\[ ESS(1) = \sum_{t=1}^{39} (I_c(0, t) + I_h(0, t) - \hat{I}_{(0, t)})^2. \] (22)

\[ ESS(2) = \sum_{t=82}^{101} (I_c(0, t) + I_h(0, t) - \hat{I}_{(0, t)})^2. \] (23)

\[ ESS(3) = (I_c(61) - \hat{I}_{(61)})^2. \] (24)

\[ ESS(4) = \left[ I_h(73) + R_h(73) - (\hat{I}_{(73)} + \hat{R}_{(73)}) \right]^2. \] (25)

\[ ESS(5) = \sum_{t=82}^{101} (R_c(0, t) + R_h(t) - \hat{R}_{(t)})^2. \] (26)
Consequently,

\[ 5.3. \]

Similarly, the national in Wuhan city can be regarded as the samples of population in the community. The ratio of the cases confirmed during the time from Jan 29 to Jan 31 among nationals is approximately equal to that on Jan 30. In this sense, the number of infectives is likely to be about 36880 in Wuhan city on Jan 30. Hence, \( \hat{I}_{(61)} \) is set to 36,880.

5.2. Initialization of parameters

The parameter estimation is, in general, a complex task partly due to so many unknown parameters: \( \xi, \zeta, \alpha_c, \alpha_h, k_\beta, \lambda_\beta, k_\gamma, \lambda_\gamma \). Correct convergence is hard to reach unless a reliable initial guess is provided. Therefore, an optimum initial estimation based on knowledge is crucial in this case. For this reason, all the parameters are dichotomized into two classes. The parameters \( \xi, \zeta, \alpha_c, \alpha_h \) belong to class one and the others class two. Among the first-class parameters, \( \xi, \zeta \) are determined by intervention, \( \alpha_c, \alpha_h \) reflect the combined effect of virus and control. The second-class parameters are related to the characteristic of the virus.

To estimate the parameters of the first class, we first investigated the implementation of control measures in Wuhan city. From the report of HCHP [38], we found that an absolute increment of the number of confirmed infectives was more than 10 thousand on Feb 12, 2020 compared to about one thousand the day before that day. This implies that many patients are likely to fail to be hospitalized before Feb 12. Taking into account the strict restriction on outdoor activities imposed before, \( \xi \) and \( \zeta \) are set at 0.8 and 0.2.

The permanent resident population of Wuhan city is about 14 million [35]. Moreover, Wuhan is well known for being the transport hub of China. The contact rate \( \alpha_c \) was set 3 for the infectiousness in the community. On the other side, ideally, infections hardly spread in hospital. But, this is not the case for the inadequacy of medical supplies. And, according to the report of HCHP, a lot of medical supplies had been distributed after Feb 12. From these points under consideration, the initial contact rate \( \alpha_h \) is set equal to 0.05. At the same time, the contact rate \( \alpha_h \) is assumed to decrease by 99% after Feb 12.

With regard to class two, the parameters are relevant to the Weibull PDF of the latent period and the infectious period. According to the medical records of 138 patients at Zhongnan Hospital of Wuhan University, the median hospital stay is 10 days [44]. Li et al [4] declared that the mean incubation period is 5.2 days. Assuming that the latent period and the infectious period is approximately equal to the incubation period and the length of stay, the mean values of the latent period and the infectious period are preliminarily estimated to be 5 and 10 days, respectively. The profiles of the Weibull PDF with various shape parameters are shown in Fig. 4. As a result, the initial guesses of the parameters \( k_\beta, \lambda_\beta, k_\gamma, \lambda_\gamma \) are 2.4, 5.64, 2.4 and 11.28, respectively.

Similarly, we determined the low bound and upper bound of the parameters. It is worth noting that to cut down the domain may not be a good strategy for global optimization since that may block the way to the global optimal point.

5.3. Two-step iterative optimization

Global solutions are usually difficult to locate, whereas the situation may be improved when constraints are add [45]. Consequently, an inequality constraint is defined to allow algorithms to make good choices for search directions. Based on

\[
\text{Fig. 4. Weibull PDF with various shape parameters while (a) mean value is 5, and (b) mean value is 10.}
\]
the knowledge that isolation and quarantine are useful control measures [36], the numbers of exposed and infectious individuals in the presence of control efforts are consequently no more than those in the absence of interventions.

In theory, if only the mean values of the latent period and the infectious period are unchanged, the result is almost invariable in number. In this sense, for a given mean value, the second-class parameters only affect the shape of the SEIR-HC model.

As a result, we explored a two-step optimization by adopting the sequential quadratic programming (SQP) method. In the first step, the first-class parameters are estimated by the SQP method. Here, the results are taken as the initial guesses of the first class in the second step. In the second step, the second-class parameters are determined by the same way. The above process is iterated many times. The flow chart is presented in Fig. 5.

In brief, the complete process of the two-step parameter optimization can be divided into four steps:

1. Guess all the parameters and their low bound and upper bound as described in Section 5.2.
2. Estimate the first-class parameters using the SQP method with the objective function presented in Section 5.1 and inequality constraints stated in this section.
3. Estimate the second-class parameters in the same manner with the first-class parameters updated in step (2).
4. Repeat the computing process from (2) to (3) until the bias is small than a given threshold value or cycle index reaches 5.
6. Result and discussion

6.1. Estimated value of SEIR-HC parameters

After the two-step iterative optimization, all the parameters of the SEIR-HC model are determined for COVID-19 in Wuhan. The low bound, the upper bound, the initial value and the estimated results are summarized in Table 2. All results fall within the range between the upper and lower limits. Some are very close to the initial value, but some far from the guess.

6.2. Epidemiology characteristic of COVID-19

With the parameters estimated, it is easy to derive that the mean and variance of the latent period are 4.19 and 1.53, and those of the infectious period are 12.53 and 11.40. The Weibull PDFs of the latent period and the infectious period are shown in Fig. 6. It can be observed that a large proportion of exposed individuals become infectious in a short time and most cases are mild. The cumulative probability of the latent period for 10 days is up to 99.99%, by which the 14-day period of active monitoring is well supported. The difference between the latent period and the incubation period is equal to 1.01 days and the difference between the infectious period and the length of stay is 2.53 days.

It is worth noting that \(1 - \xi(= 20.1\%)\) infectives, which is equivalent to, on the average, 2.52 days per infective, still stayed in the community before Feb 12. Assuming that the infectiousness is constant during the entire infectious period, the basic reproduction number is up to 7.90 where everyone is susceptible. The basic reproduction number estimated here is compared with others in Table 3. Furthermore, even though an infective is hospitalized at the beginning of clinical symptoms, he still can make 3.48 individuals infected. As a result, the outbreak is inevitable in the absence of interventions for the difference between the latent period and the incubation period. Of course, this is not necessary true for districts other than Wuhan city, because basic reproduction number varies from population density and social enthusiasm besides characteristics of pathogenic bacteria [46].

6.3. Transmission process of COVID-19

With the SEIR-HC model proposed here, the transmission process of COVID-19 epidemic in Wuhan city is reproduced as shown in Fig. 7. In our baseline scenario, we estimate that the outbreak would be over before Apr 15, 2020, and the total...
Table 3
Comparison of basic reproduction numbers.

| Literature  | Assumption | Data | Epidemic model | Method                  | R estimated (average) |
|-------------|------------|------|----------------|-------------------------|-----------------------|
| Li [4]      | -          | Confirmed cases in Wuhan (Dec 10 to Jan 4) | Unspecified            | Derived from SI        | 2.2                   |
| Tuite [6]   | -          | Exported cases on Jan 12 and Jan 18       | Renewal equation       | Derived from SI        | 2.3                   |
| Lai [7]     | -          | Confirmed cases in China (Jan 10 to Jan 24) | Exponential growth     | Unspecified            | 2.24-3.58             |
| Zhao [5]    | SI(COVID-19) = SI(SARS) | Confirmed cases in China (Dec 1 to Jan 24) | Exponential growth     | Derived from SI        | 2.56                  |
| Wu [10]     | SI(COVID-19) = SI(SARS) | Exported cases (Dec 25 to Jan 19) | SEIR                   | MLE                     | 2.68                  |
| Tang [8]    | -          | Confirmed cases in China (before Jan 22)  | SEIR                   | MLE                     | 6.47                  |
| Ours        | -          | Confirmed cases in Wuhan (Dec 1 to Jan 8, Feb 20 to Mar 10), exported cases on Jan 30, and others. | SEIR-HC                 | SQP                     | 7.90                  |

Note that SI denotes serial interval and MLE means maximum likelihood estimation.

Fig. 7. Transmission process of 2019-nCov epidemic in Wuhan city. In the picture, the capital letters E, I, R indicate the exposed, infectious and removed individuals, respectively. And, the lowercase letters c and h mean community and hospital. Note that t indicates the number of days delayed from Dec 1, 2019.

of infectives no longer increases by Mar 15, 2020. At the same time, the total number of removed individuals would reach 111383 finally (as shown in Fig. 7(a)). Among them, the number of hospital staff members would be up to 2950, which is likely to be slightly more than the reported cases due to asymptomatic infection. The number of the infected individuals in Wuhan on Jan 25, 2020, is estimated to be 46029, which is less than the count 75815 estimated by Wu et al [10]. The number of infectives on Jan 30 is 35962 (2.52% difference to 36880), which is much higher than 20767 (the number of the
infected individuals estimated by Nishiura et al [3]). Number of infectious hospital staff members on Feb 11 is equal to 1241 (11.87% difference to 1022).

It is readily seen from Fig. 7(b)-(d) that there is a sudden decrease of the number of the exposed individuals on Jan 23 and on Feb 12, which implies control measures launched have a conspicuous effect on the infection rate.

Though mathematical models of epidemic transmission often scarce contrast model output with real-world observations [1], the comparison is necessary to demonstrate the performance of the models and the validity of the results. For this reason, Fig. 8 provides a pictorial display of comparison of the estimated data and the reported ones. The two curves show the same trend after Feb 12. But, before that day, the number of infectious individuals estimated here is far more than that reported by WMHC and HCHP. A probable reason for this is the underreporting of incidences before Feb 12 [3–6], which is the main reason for the underestimation of basic reproduction number.

From Fig. 8(a), it seems that there is a delay of the estimated number of the removed individuals relative to the reported. The underlying cause of the delay may be the fact that the discharge time is later than the end time of infectious period for functional recovery. In this sense, it appears to be appropriate that the outbreak in Wuhan terminates later than the expected time. However, the warmer weather is helpful in preventing the virus from reproducing. Given these points, the outbreak is likely to be ended as we expected before if the control measures are kept on as usual. Additionally, it also can be observed from Fig. 8(a) that, in the worst case, up to 33620 infectives failed to be hospitalized on Jan 30.

6.4. Control measures

To assess the control measure, a series of experiments are carried out using the SEIR-HC model. The control measures are simulated by the first-class parameters and the effect is captured by the number of infectives.

6.4.1. Effect of control level

Since the proportion of quarantined susceptibles to the total number of susceptibles $1 - \xi$ implies control level, the function of quarantine is tested by adopting $\xi$. From the SEIR-HC model, we can see that $\xi$ impacts primarily on the community infection. So, only $1c$-t and $(1c+Rc)$-t graphs are shown here for space limitations. Fig. 9(a) shows $I_c$ varying when $\xi$ increases from 0.032 to 0.932 and Fig. 9(b) shows $I_c + R_c$. Fig. 9(c) and (d) display the results with a delay of 0–9 days. It was clear that the number of infectives dramatically increases with the proportion $\xi$ and delay time. As a result, control measures played a key role in prevention of the spread. It also can be found from Fig. 9 that the resulting divergence begins after Jan 28. Therefore, Jan 23, on which the measures imposed, is the right time to stop the outbreak.

6.4.2. Effect of the time of hospital admission

Given the proportion of hospitalized infectives to total number of infectives $\xi$, $1 - \xi$ is equivalent to the proportion of the average time during which infectives stay in the community to the average infectious period, which reflects how quickly the infectives are allowed to hospitalize. In this section, the effect of the time of hospital admission is test by changing parameter $\xi$. Fig. 10(a) shows that the number of infectives in the community varies with $\xi$ and Fig. 10(b) is the corresponding accumulative value. Fig. 10(c) and (d) depict the case in hospital. The delay of hospital admission makes an increase in the number of infectives both in the community and in hospital. However, the increase of the number of infectives slows down when $\xi$ is small enough due to the depletion of susceptibles in the population. It can be observed
Fig. 9. Simulated function of control measures: (a) number of infectives at various control levels, (b) accumulative total of infectives at various control levels, (c) number of infectives when control is delayed, (d) accumulative total of infectives varying with delayed time. Note that $t$ indicates the number of days delayed from Dec 1, 2019, while delay=$x$ means $x$ days delayed from Jan 23, 2020.

from Fig. 10(b) that, as a result of a large number of inbound travellers every day, the accumulative total of the infectives can be even more than permanent resident population. In fact, this is almost impossible because the implicit assumption that travel behavior was not affected by epidemic is not valid.

6.4.3. Effect of contact rate $\alpha_c$

Contact rate is primarily determined by the nature of pathogen. In fact, it can also be changed in some degree by intervention strategies [24]. Assuming that $\alpha_c$ is varied from 3.3 to 5.1, the number of infectives in the community varying with $\alpha_c$ is demonstrated in Fig. 11. A clear result is that the number of infectives in the community increases exponentially with $\alpha_c$. Note that, following the same idea mentioned in Section 6.4.1, only the number of infectives in the community is shown here.

6.4.4. Effect of contact rate $\alpha_h$

The contact rate $\alpha_h$ partly reflects the prevention level in hospital. From Fig. 12, the number of infectives in hospital increase sharply with $\alpha_h$.

6.4.5. Control methodology of COVID-19

According to the experimental results, it is easy to find out about the control methodology. The basic principle is taking measures as early as possible to lower $\zeta$, $\alpha_c$, $\alpha_h$ and to enhance $\xi$. The control measures of COVID-19 list as follows.

(1) Keeping in home quarantine and reducing travel to lower $\zeta$.
Fig. 10. Effect of the time of hospital admission: (a) number of infectives in the community at various proportion \( \xi \), (b) accumulative total of infectives in the community at various proportion \( \xi \), (c) number of infectives in hospital at various proportion \( \xi \), (d) accumulative total of infectives in hospital at various proportion \( \xi \). Note that \( t \) indicates the number of days delayed from Dec 1, 2019.

Fig. 11. Effect of contact rate \( \alpha_c \): (a) number of infectives in the community at various levels of contact rate \( \alpha_c \), (b) accumulative total of infectives in the community at various levels of contact rate \( \alpha_c \). Note that \( t \) indicates the number of days delayed from Dec 1, 2019.
(2) Tracing, testing and quarantining the suspected case, immediately isolating symptomatic individuals and speeding hospital admission to enhance $\xi$;

(3) Strengthening personal protection to lessen $\alpha_c$ and $\alpha_h$.

7. Conclusion

This work provides the SEIR-HC model, a novel SEIR model with two different social circles. If let $\xi = 0$, the SEIR-HC model is easy to degrade into the standard SEIR model. For exploring the transmission dynamics of COVID-19, a two-step optimization method exclusively designed for parameter estimation of SEIR-HC model. With the model, the spread process of COVID-19 is reproduced clearly even without enough observation data. The latent period, infectious period and basic reproduction number of COVID-19 are estimated to be 4.19, 12.53 and 7.90, respectively. Obviously, COVID-19 is highly transmissible. Furthermore, the outbreak in Wuhan is anticipated to be over before Apr 15, 2020. The total number of removed individuals would reach 111383 finally. Among them, the number of hospital staff members would be up to 2950.

According to the SEIR-HC model, the principle of prevention and control of COVID-19 is taking measures as early as possible to lower $\zeta$, $\alpha_c$, $\alpha_h$ and to enhance $\xi$. A set of measures such as quarantine have significant impacts on lessening the spread. By the way, an international effort is required to prevent virus transmission since COVID-19 has spread all over the world.

As a whole, the conclusions are well interpretable and reasonable. As evidenced by the success in estimation and prediction, the SEIR-HC model is useful. Although the results are based on the data from Wuhan and hence they are not necessary to be reliable for other cities, the SEIR-HC model is valid everywhere, which allows us to capitalize on new data streams and lead to an ever-greater ability to generate robust insight and collectively shape successful local and global public health policy.

Acknowledgments

This work was supported by National Social Science Foundation of China (16BXW005), Chongqing Basic Science and Frontier Research Project, China (No. cstc2017jcyjAX0007, cstc2017jcyjAX0386), Project Foundation of Chongqing Municipal Education Committee, China (Grant no. 17SKG050, 19SKG046, KJQN201900620), and Sichuan Provincial Key Laboratory of Robot Technology Used for Special Environment (Southwest University of Science and Technology), China (Grant No. 17kftk02).

References

[1] H. Heesterbeek, et al., Modeling infectious disease dynamics in the complex landscape of global health, Science 347 (6227) (2015) 1216–1226.
[2] W. Coffman, V.A. Newill, Generalization of epidemic theory: an application to the transmission of ideas, Nature 204 (4955) (1964) 225–228.
[3] H. Nishiura, et al., The rate of underascertainment of novel coronavirus (2019-nCoV) infection: estimation using Japanese passengers data on evacuation flights, J. Clin. Med. 9 (2) (2020) 419–421.
[4] Q. Li, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, N. Engl. J. Med. 382 (3) (2020) 1199–1207.
[5] S. Zhao, et al., Estimating the unreported number of Novel Coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak, J. Clin. Med. 9 (2) (2020) 338–343.
[6] A.R. Tuite, D.N. Fisman, Reporting, epidemic growth, and reproduction numbers for the 2019 Novel Coronavirus (2019-nCoV) epidemic, Ann. Intern. Med. 257 (3) (2020) 1–7.
