Nonlinear Dyn manuscript No.
(will be inserted by the editor)

Forecast analysis of the epidemics trend of COVID-19 in the United States by a generalized fractional-order SEIR model

Conghui Xu · Yongguang Yu* · YangQuan Chen* · Zhenzhen Lu

Received: date / Accepted: date

Abstract In this paper, a generalized fractional-order SEIR model is proposed, denoted by SEIQRP model, which has a basic guiding significance for the prediction of the possible outbreak of infectious diseases like COVID-19 and other insect diseases in the future. Firstly, some qualitative properties of the model are analyzed. The basic reproduction number R_0 is derived. When $R_0 < 1$, the disease-free equilibrium point is unique and locally asymptotically stable. When $R_0 > 1$, the endemic equilibrium point is also unique. Furthermore, some conditions are established to ensure the local asymptotic stability of disease-free and endemic equilibrium points. The trend of COVID-19 spread in the United States is predicted. Considering the influence of the individual behavior and government mitigation measurement, a modified SEIQRP model is proposed, defined as SEIQRPD model. According to the real data of the United States, it is found that our improved model has a better prediction ability for the epidemic trend in the next two weeks. Hence, the epidemic trend of the United States in the next two weeks is investigated, and the peak of isolated cases are predicted. The modified SEIQRP model successfully capture the development process of COVID-19, which provides an important reference for understanding the trend of the outbreak.

Keywords COVID-19 · Fractional order · Generalized SEIR model · Epidemic · Peak prediction

This work is supported by the National Nature Science Foundation of China under Grant 61772063, Beijing Natural Science Foundation under Grant Z180005.

Conghui Xu. Yongguang Yu (✉). Zhenzhen Lu.
Department of Mathematics, School of Science, Beijing Jiaotong University, Beijing 100044, China.
E-mail: ygyu@bjtu.edu.cn
YangQuan Chen (✉)
Mechatronics, Embedded Systems and Automation Lab, University of California, Merced, CA 95343, USA
E-mail: ychen53@ucmerced.edu

1 Introduction

The outbreak of the coronavirus disease (COVID-19) in 2019 occurred in Wuhan, China, at the end of 2019. This is a severe respiratory syndrome caused by the novel coronavirus of zoonotic origin [1]. The Chinese government has implemented many measures, including the establishment of specialized hospitals and restrictions on travel, to reduce the spread. By April 20, 2020, the outbreak in China has been basically controlled. However, the outbreak is still rampant all over the world. At present, the United States, Italy, Spain and other countries are still in the rising stage of the outbreak. It has posed a great threat to the public health and safety of the world.

At present, many countries have adopted mitigation measures to restrict travel and public gatherings, which have a serious impact on the economy. Therefore, it is very important to predict the development trend of this epidemic and estimate the peak of the isolated cases. The epidemic model is a basic tool to research the dynamic behaviours of disease and predict the spreading trend of disease. Establishing a reasonable epidemic model can effectively characterize the development process of the disease. Therefore, a generalized SEIR epidemic model is proposed in this paper, and is defined as SEIQRP model. Some qualitative properties of this model are first analyzed, including the existence and uniqueness of the disease-free and the endemic equilibrium points. Then, conditions are also established to ensure the local asymptotic stability of both disease-free and endemic equilibrium points.

So far, many scholars have researched COVID-19 from different perspectives [1–5, 23]. In [6], the epidemics trend of COVID-19 in China was predicted under public health interventions. In [7], the basic reproduction number of COVID-19 in China was estimated and the data-driven analysis was performed in the early phase of the outbreak. In order to pre-

dict the cumulative number of confirmed cases and combine the actual measures taken by the government on the outbreak of COVID-19, we further put forward an improved SEIQR-P model, denoted by SEIQRPD model. At present, we can obtain the epidemic data of COVID-19 outbreak in the United States before April 20, 2020. We use these data and the improved model to predict the epidemic trend of the United States in the next two weeks, and estimate the peak of isolated cases. Firstly, the data before April 5 was selected to identify the model parameters, and the prediction ability of the improved model with the epidemiological data from April 6 to April 20 is verified. Then, the cumulative number of confirmed cases and isolated cases are predicted in the next two weeks. The peak of isolated cases is thus predicted.

In recent years, with the continuous development of fractional calculus theory, fractional order systems modeling approaches have been applied in various engineering and non-engineering fields [19–21]. Compared with the short memory of the integer derivative, the fractional derivative has the information of the whole time interval or long memory. It is more accurate to describe the biological behavior of population by using fractional differential equation.

The rest of this manuscript is structured as below. In Section 2, the fractional SEIQRP model and the modified SEIQRP model is proposed. Some qualitative properties of the SEIQRP model are discussed in Section 3. In Section 4, the prediction ability of the modified SEIQRP model is verified by using real data. The disease development in the United States in the next two weeks after April 21, 2020 is predicted. Finally, conclusion and some future works are discussed in Section 5.

2 Preliminaries and Model Derivation

2.1 Preliminaries

In this section, some useful lemmas and definitions will be given to analyze some results of this paper.

Definition 1 [8] The Caputo fractional order derivative is given below

$$ {}_0^C D_t^\alpha f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-\xi)^{n-\alpha-1} f^{(n)}(\xi) d\xi, $$

where $n-1 \le \alpha \le n$. In order to simplify the symbolic representation, the Caputo fractional differential operator ${}_0^C D_t^\alpha$ in this paper are represented by D^α.

Definition 2 [8, 22] The Mittag-Leffler function is given below

$$ E_\alpha(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(k\alpha+1)}, $$

where $n-1 \le \alpha \le n$.

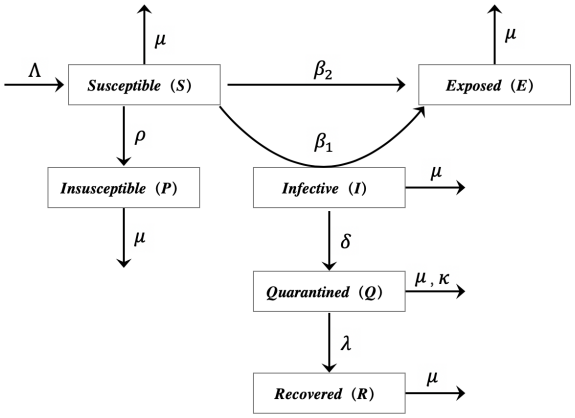

Fig. 1: Transmission diagram for the model involving six population classes.

Consider the following n-dimensional fractional order differential equation system

$$ D^\alpha X(t) = AX(t), \;\; x(0) = x_0, \tag{1} $$

where $\alpha \in (0,1)$, $X(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T$ is an n dimensional state vector and A is an $n \times n$ constant matrix.

Lemma 1 *[9] For the corresponding linear time-invariant system (1), the following results are true:*
(i) The zero solution is asymptotically stable, if and only if all eigenvalues s_j $(j = 1,2,\ldots,n)$ of A satisfy $|\arg(s_j)| > \frac{\alpha\pi}{2}$.
(ii) The zero solution is stable, if and only if all eigenvalues s_j of A satisfy $|\arg(s_j)| \ge \frac{\alpha\pi}{2}$ and eigenvalues with $|\arg(s_j)| = \frac{\alpha\pi}{2}$ have the same algebraic multiplicity and geometric multiplicity.

2.2 SEIQRP model

The outbreak of COVID-19 has had a great impact on the economic growth of any country and daily life of any human. In order to control and prevent the possible outbreak of infectious diseases like the COVID-19 or other insect diseases in the future, it is very important to establish an appropriate model. The transmission diagram of the generalized SEIR model proposed in this paper is shown in Figure 1. We divide the total population into six distinct epidemic classes: susceptible, exposed, infectious, quarantined, recovered and insusceptible. We will represent the number of individuals at time t in the above classes by $S(t)$, $E(t)$, $I(t)$, $Q(t)$, $R(t)$ and $P(t)$, respectively. The specific explanations of the above six categories are as follows:

- Susceptible $S(t)$: the number of uninfected individuals at the time t.

- Exposed $E(t)$: the number of infected individuals at the time t but still in incubation period (without clinical symptoms and low infectivity).
- Infectious $I(t)$: the number of infected individuals at the time t (with obvious clinical symptoms).
- Quarantined $Q(t)$: the number of individuals who have been diagnosed and isolated at the time t.
- Recovered $R(t)$: the number of recovered individuals at the time t.
- Insusceptible $P(t)$: the number of susceptible individuals who are not exposed to the external environment at the time t.

The incidence rate plays a very important role in the epidemic, which can describe the evolution of infectious disease. According to the spread of different diseases in different regions, there are many forms of incidence rate. [10–12]. We expect to establish a generalized incidence rate, which can contain most of these specific forms. Therefore, the SEIQRP model with general incidence rate is given by

$$\begin{cases} D^{\alpha}S(t) = \Lambda - \beta_1 f_1(S(t))g_1(I(t)) - \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad - \mu S(t) - \rho S(t), \\ D^{\alpha}E(t) = \beta_1 f_1(S(t))g_1(I(t)) + \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad - \varepsilon E(t) - \mu E(t), \\ D^{\alpha}I(t) = \varepsilon E(t) - (\delta+\mu)I(t), \\ D^{\alpha}Q(t) = \delta I(t) - (\lambda+\kappa+\mu)Q(t), \\ D^{\alpha}R(t) = \lambda Q(t) - \mu R(t), \\ D^{\alpha}P(t) = \rho S(t) - \mu P(t), \end{cases} \tag{2}$$

where Λ is the inflow number of susceptible individuals, β_1 and β_2 denote the infection rates of the infected individuals and the exposed individuals, respectively, ρ is the protection rate, ε represent the incubation rate. δ is the rate at which symptomatic infections are diagnosed and quarantined. λ and κ represent the cure rate of isolated individuals and the death rate caused by the disease, respectively, μ is the natural mortality, the incidence rate $\beta_1 f_1(S)g_1(I)$, $\beta_2 f_2(S)g_2(E)$ are used to describe the transmission of diseases, which satisfy the following conditions [16]:
(i) $f_1(0) = f_2(0) = g_1(0) = g_2(0) = 0$,
(ii) $f_1(S) > 0$, $f_2(S) > 0$, $g_1(I) > 0$, $g_2(E) > 0$, for any S, E, $I > 0$,
(iii) $f_1'(S) > 0$, $f_2'(S) > 0$, $g_1'(I) > 0$, $g_2'(E) > 0$, for any S, E, $I > 0$,
(iv) $g_1'(I) - \frac{g_1(I)}{I} \le 0$, $g_2'(E) - \frac{g_2(E)}{E} \le 0$, for any E, $I > 0$.

2.3 Modified SEIQRP model

In the following, we will combine the actual situation and government mitigation policy to improve the SEIQRP model (2). Firstly, during the outbreak of COVID-19, we need to make reasonable assumptions according to the actual mitigation policies and circumstances.

- During the COVID-19 outbreak, population mobility is strictly controlled by many countries. Especially, the policy of city closure was implemented in Hubei province, China. Therefore, the impact of migration will not be considered in the improved model.
- For the prediction of this short-term virus transmission, the impact of natural mortality will not be considered. It is assumed that novel coronavirus is the only cause of death during the outbreak.
- In order to predict the trend of cumulative confirmed cases, it is necessary to simulate the number of death cases. Therefore, a new class $D(t)$ to the SEIQRP model will be added in the modified model, which denotes the number of death cases at time t.
- The time-varying cure rate, mortality rate and infection rate will be applied to the improved model. This can better simulate the impact of the improvement of medical conditions and government control on individuals in reality.

Based on the above assumptions and analysis, the following improved model for COVID-19 is proposed.

$$\begin{cases} D^{\alpha}S(t) = -\beta_1 f_1(S(t))g_1(I(t)) - \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad - \rho S(t), \\ D^{\alpha}E(t) = \beta_1 f_1(S(t))g_1(I(t)) + \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad - \varepsilon E(t), \\ D^{\alpha}I(t) = \varepsilon E(t) - \delta I(t), \\ D^{\alpha}Q(t) = \delta I(t) - (\lambda(t)+\kappa(t))Q(t), \\ D^{\alpha}R(t) = \lambda(t)Q(t), \\ D^{\alpha}P(t) = \rho S(t), \\ D^{\alpha}D(t) = \kappa(t)Q(t), \end{cases} \tag{3}$$

where $\beta_1(t) = \sigma_1 \exp(-\sigma_2 t)$, $\lambda(t) = \lambda_1(1 - \exp(\lambda_2 t))$ and $\kappa(t) = \kappa_1 \exp(-\kappa_2 t)$. The parameters σ_1, σ_2, λ_1, λ_2, κ_1 and κ_2 are all positive, where σ is used to simulate the intensity of government control. It should be emphasized that the protection rate ρ for susceptible individuals also reflects the intensity of government control.

3 Qualitative Analysis of the SEIQRP Model

3.1 The Existence and Uniqueness of Equilibrium Point

Obviously, the right side of system (2) satisfies the local Lipschitz condition, then there exists a unique solution of system (2) [8, 15].

Theorem 1 *The solutions of system (2) are non-negative, and the closed set* $\Omega = \{(S,E,I,Q,R,P) \in \mathbb{R}^6_+ : 0 \le S+E+I+Q+R+D \le \frac{\Lambda}{\mu}\}$ *is a positive invariant set of system (2).*

Proof In order to investigate the non-negativity of solutions of the system (2), we consider the following system

$$\begin{cases} D^\alpha S_1(t) = -\beta_1 f_1(S(t))g_1(I(t)) - \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad -\mu S(t) - \rho S(t), \\ D^\alpha E_1(t) = \beta_1 f_1(S(t))g_1(I(t)) + \beta_2 f_2(S(t))g_2(E(t)) \\ \qquad\qquad -\varepsilon E(t) - \mu E(t), \\ D^\alpha I_1(t) = \varepsilon E(t) - (\delta+\mu) I(t), \\ D^\alpha Q_1(t) = \delta I(t) - (\lambda+\kappa+\mu) Q(t), \\ D^\alpha R_1(t) = \lambda Q(t) - \mu R(t), \\ D^\alpha P_1(t) = \rho S(t) - \mu P(t), \end{cases} \tag{4}$$

where the initial conditions are $S_{10}=0$, $E_{10}=0$, $I_{10}=0$, $Q_{10}=0$, $R_{10}=0$ and $P_{10}=0$. $(0,0,0,0,0,0)$ is the unique solution of the above system. According to the fractional-order comparison theorem [17], one can deduce that the solutions of system (2) satisfy $S(t)\geq 0$, $E(t)\geq 0$, $I(t)\geq 0$, $Q(t)\geq 0$, $R(t)\geq 0$ and $P(t)\geq 0$. By adding six equations of system (2), one can deduce

$$\begin{aligned} D^\alpha N(t) &= \Lambda - \mu N(t) - \kappa Q(t) \\ &\leq \Lambda - \mu N(t). \end{aligned}$$

By applying the fractional-order comparison theorem, one has

$$D^\alpha N(t) \leq (-\frac{\Lambda}{\mu} + N(0)) E_\alpha(-\mu t^\alpha) + \frac{\Lambda}{\mu}.$$

When $N(0)\leq \frac{\Lambda}{\mu}$, since $E_\alpha(-\mu t^\alpha)\geq 0$, we have

$$S(t)+E(t)+I(t)+Q(t)+R(t)+P(t) \leq \frac{\Lambda}{\mu}$$

Thus we can draw the result of Theorem 1.

System (2) has an obvious disease-free equilibrium point $M_0=(S_0,0,0,0,0,P_0)$, where

$$S_0 = \frac{\Lambda}{\mu+\rho},\ P_0 = \frac{\rho\Lambda}{\mu(\mu+\rho)}.$$

In order to obtain the endemic equilibrium point of the system (2), we set:

$$\begin{cases} \Lambda - \beta_1 f_1(S(t))g_1(I(t)) - \beta_2 f_2(S(t))g_2(E(t)) \\ -\mu S(t) - \rho S(t) = 0, \\ \beta_1 f_1(S(t))g_1(I(t)) + \beta_2 f_2(S(t))g_2(E(t)) \\ -\varepsilon E(t) - \mu E(t) = 0, \\ \varepsilon E(t) - (\delta+\mu) I(t) = 0, \\ \delta I(t) - (\lambda+\kappa+\mu) Q(t) = 0, \\ \lambda Q(t) - \mu R(t) = 0, \\ \rho S(t) - \mu P(t) = 0, \end{cases} \tag{5}$$

which implies

$$S = \frac{\Lambda - (\varepsilon+\mu)E}{\mu+\rho},\ I = \frac{\varepsilon E}{\sigma+\mu}.$$

Combining the above equations and the second equation of (5), one has

$$\begin{aligned} (\mu+\varepsilon)E =& \beta_1 f_1(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_1(\frac{\varepsilon E}{\mu+\delta}) \\ &+ \beta_2 f_2(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_2(E). \end{aligned} \tag{6}$$

Define

$$\begin{aligned} \varphi(E) =& \beta_1 f_1(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_1(\frac{\varepsilon E}{\mu+\delta}) \\ &+ \beta_2 f_2(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_2(E) - (\mu+\varepsilon)E. \end{aligned} \tag{7}$$

Note that $\varphi(0)=0$ and $\varphi(\frac{\Lambda}{\mu+\varepsilon}) = -\Lambda < 0$. In order to show that $\varphi(E)=0$ has at least one positive root in the interval $(0, \frac{\Lambda}{\mu+\varepsilon}]$, we need to prove that $\varphi'(0)>0$. Hence

$$\begin{aligned} \varphi'(E) =& -\frac{(\mu+\varepsilon)\beta_1}{\mu+\rho} f_1'(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_1(\frac{\varepsilon E}{\mu+\delta}) \\ &+ \frac{\varepsilon\beta 1}{\mu+\delta} f_1(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_1'(\frac{\varepsilon E}{\mu+\delta}) \\ &- \frac{(\mu+\varepsilon)\beta_2}{\mu+\rho} f_2'(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_2(E) \\ &+ \beta_2 f_2(\frac{\Lambda-(\mu+\varepsilon)E}{\mu+\rho}) g_2'(E) - (\mu+\varepsilon). \end{aligned} \tag{8}$$

Therefore, we have

$$\begin{aligned} \varphi'(0) &= \frac{\varepsilon\beta_1}{\mu+\delta} f_1(S_0) g_1'(0) + \beta_2 f_2(S_0) g_2'(0) - (\mu+\varepsilon) \\ &= (\mu+\varepsilon)(R_0-1), \end{aligned}$$

where the basic reproduction number is given by

$$R_0 = \frac{\varepsilon\beta_1}{(\mu+\varepsilon)(\mu+\delta)} f_1(S_0) g_1'(0) + \frac{\beta_2}{\mu+\varepsilon} f_2(S_0) g_2'(0). \tag{9}$$

If $R_0>1$, the system (2) has at least one endemic equilibrium point $M_*=(S_*,E_*,I_*,Q_*,R_*,P_*)$, where

$$S_* = \frac{\Lambda-(\mu+\varepsilon)E_*}{\mu+\rho},\ I_* = \frac{\varepsilon E_*}{\mu+\delta},\ Q_* = \frac{\delta I_*}{\mu+\lambda+\kappa},$$

$$R_* = \frac{\lambda Q_*}{\mu},\ P_* = \frac{\rho S_*}{\mu}.$$

In the following section, we show that endemic equilibrium point M_* is unique. According to the above analysis and hypothesis (i)-(iii), one has

$$\begin{aligned}\varphi'(E_*) =& -\frac{(\mu+\varepsilon)\beta_1}{\mu+\rho}f_1'(S_*)g_1(I_*)+\frac{\varepsilon\beta_1}{\mu+\delta}f_1(S_*)g_1'(I_*)\\ &-\frac{(\mu+\varepsilon)\beta_2}{\mu+\rho}f_2'(S_*)g_2(E_*)+\beta_2 f_2(S_*)g_2'(E_*)\\ &-(\mu+\varepsilon)\\ =& -\frac{(\mu+\varepsilon)\beta_1}{\mu+\rho}f_1'(S_*)g_1(I_*)-\frac{(\mu+\varepsilon)\beta_2}{\mu+\rho}f_2'(S_*)g_2(E_*)\\ &+\frac{\varepsilon\beta_1}{\mu+\delta}f_1(S_*)g_1'(I_*)-\frac{\varepsilon\beta_1 f_1(S_*)g_1(I_*)}{(\mu+\delta)I_*}\\ &+\beta_2 f_2(S_*)g_2'(E_*)-\frac{\beta_2 f_2(S_*)g_2(E_*)}{E_*}\\ =& -\frac{(\mu+\varepsilon)\beta_1}{\mu+\rho}f_1'(S_*)g_1(I_*)-\frac{(\mu+\varepsilon)\beta_2}{\mu+\rho}f_2'(S_*)g_2(E_*)\\ &+\frac{\varepsilon\beta_1}{\mu+\delta}f_1(S_*)[g_1'(I_*)-\frac{g_1(I_*)}{I_*}]\\ &+\beta_2 f_2(S_*)[g_2'(E_*)-\frac{g_2(E_*)}{E_*}].\end{aligned} \tag{10}$$

By the hypothesis (iv), this implies $\varphi'(E_*)<0$. If there is another endemic equilibrium point M_{**}, then $\varphi'(E_{**})\geq 0$ holds, which contradicts the previous discussion. Hence, system (2) has a unique endemic equilibrium point M_* when $R_0>1$. Based on the above analysis, the following result can be obtained.

Theorem 2 *System (2) has a unique disease-free equilibrium point M_0, if $R_0<1$. System (2) has a unique endemic equilibrium point M_*, if $R_0>1$.*

3.2 Stability Analysis

In this section, the local asymptotic stability of disease-free equilibrium point M_0 and endemic equilibrium point M_* for system (2) is discussed.

Theorem 3 *With regard to system (2), the disease-free equilibrium point M_0 is locally asymptotic stability, if $R_0<1$; the disease-free equilibrium point M_0 is unstable, if $R_0>1$.*

Proof The Jacobian matrix of the system (2) at the disease-free equilibrium point M_0 is

$$J_{M_0}=\begin{bmatrix}J_{11} & J_{12}\\ J_{21} & J_{22}\end{bmatrix},$$

where

$$J_{11}=\begin{bmatrix}-\mu-\rho & -\beta_2 f_2(S_0)g_2'(0) & -\beta_1 f_1(S_0)g_1'(0)\\ 0 & \beta_2 f_2(S_0)g_2'(0)-\mu-\varepsilon & \beta_1 f_1(S_0)g_1'(0)\\ 0 & \varepsilon & -\mu-\delta\end{bmatrix},$$

$$J_{12}=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix},\ J_{21}=\begin{bmatrix}0&0&\delta\\0&0&0\\\rho&0&0\end{bmatrix},$$

$$J_{22}=\begin{bmatrix}-\mu-\kappa-\lambda & 0 & 0\\ \lambda & -\mu & 0\\ 0 & 0 & -\mu\end{bmatrix}.$$

The corresponding characteristic equation is

$$\begin{aligned}H(s)=&|sE-J_{M_0}|\\ =&(s+\mu)^2(s+\mu+\rho)(s+\mu+\lambda+\kappa)H_1(s),\end{aligned} \tag{11}$$

where

$$\begin{aligned}H_1(s)=&s^2+(\delta+2\mu+\varepsilon-\beta_2 f_2(S_0)g_2'(0))s\\ &+(\mu+\delta)(\mu+\varepsilon)-\beta_2(\mu+\delta)f_2(S_0)g_2'(0)\\ &-\varepsilon\beta_1 f_1(S_0)g_1'(0).\end{aligned} \tag{12}$$

The characteristic equation $H(s)=0$ has four obvious negative characteristic roots, which are denoted by $s_1=s_2=-\mu$, $s_3=-\mu-\rho$ and $s_4=-\mu-\lambda-\kappa$, respectively. The discriminant of $H_1(s)$ in quadratic form is

$$\begin{aligned}\Delta=&[\delta+2\mu+\varepsilon-\beta_2 f_2(S_0)g_2'(0)]^2-4[(\mu+\delta)(\mu+\varepsilon)\\ &-\beta_2(\mu+\delta)f_2(S_0)g_2'(0)-\varepsilon\beta_1 f_1(S_0)g_1'(0)]\\ =&(\delta-\varepsilon+\beta_2 f_2(S_0)g_2'(0))^2+4\varepsilon\beta_1 f_1(S_0)g_1'(0)\\ >&0.\end{aligned}$$

This implies that the other two eigenvalues s_5 and s_6 of characteristic equation $H(s)=0$ are real roots. Hence

$$s_5+s_6=-(\delta+2\mu+\varepsilon-\beta_2 f_2(S_0)g_2'(0)),$$

$$s_5 s_6=(\mu+\delta)(\mu+\varepsilon)(1-R_0).$$

If $R_0<1$, then one can obtain $s_5+s_6<0$ and $s_5 s_6>0$, which imply $s_5<0$ and $s_6<0$. If $R_0>1$, one has $s_5 s_6<0$, which imply $s_5>0$ or $s_6>0$. It follows from Lemma 1 that the proof is completed.

Further, we will show the locally asymptotic stability of the endemic equilibrium point M_* of system (2). Similarly, the corresponding Jacobian matrix of the system (2) at M_* is

$$J_{M_*}=\begin{bmatrix}J_1 & J_2\\ J_3 & J_4\end{bmatrix},$$

where

$$J_1=\begin{bmatrix}-l_3-l_4-\mu-\rho & -l_2 & -l_1\\ l_3+l_4 & l_2-\varepsilon-\mu & l_1\\ 0 & \varepsilon & -\delta-\mu\end{bmatrix},$$

$$J_2=\begin{bmatrix}0&0&0\\0&0&0\\0&0&0\end{bmatrix},\ J_3=\begin{bmatrix}0&0&\delta\\0&0&0\\\rho&0&0\end{bmatrix},$$

$$J_4=\begin{bmatrix}-\mu-\lambda-\kappa & 0 & 0\\ \lambda & -\mu & 0\\ 0 & 0 & -\mu\end{bmatrix},$$

with

$$l_1=\beta_1 f_1(S_*)g_1'(I_*),\ \ l_2=\beta_2 f_2(S_*)g_2'(E_*),$$

$$l_3=\beta_1 f_1'(S_*)g_1(I_*),\ \ l_4=\beta_2 f_2'(S_*)g_2(E_*).$$

Hence, the corresponding characteristic equation is

$$\begin{aligned}L(s)=&|sE-J_{M_*}|\\ =&(s+\mu)^2(s+\mu+\lambda+\kappa)L_1(s),\end{aligned}\tag{13}$$

where

$$L_1(s)=s^3+a_1s^2+a_2s+a_3,$$

with

$$a_1=\varepsilon+3\mu+\delta+\rho-l_2+l_3+l_4,$$

$$\begin{aligned}a_2=&(\mu+\varepsilon-l_2)(2\mu+\delta+\rho)+(\mu+\delta)(\mu+\rho)\\ &+(2\mu+\delta+\varepsilon)(l_3+l_4)-\varepsilon l_1,\end{aligned}$$

$$\begin{aligned}a_3=&(\mu+\delta)(\mu+\rho)(\mu+\varepsilon-l_2)+(\mu+\delta)(\mu+\varepsilon)(l_3+l_4)\\ &-\varepsilon(\mu+\rho)l_1.\end{aligned}$$

By the corresponding results in [13, 14], let

$$\begin{aligned}D_1(L_1(s))=&\begin{vmatrix}1 & a_1 & a_2 & a_3 & 0\\ 0 & 1 & a_1 & a_2 & a_3\\ 3 & 2la_1 & a_2 & 0 & 0\\ 0 & 3 & 2a_1 & a_2 & 0\\ 0 & 0 & 3 & 2a_1 & a_2\end{vmatrix}\\ =&18a_1a_2a_3+a_1^2a_2^2-4a_1^3a_3-4a_2^3-27a_3^2.\end{aligned}\tag{14}$$

Then, the following result can be obtained.

Theorem 4 *With regard to system (2), assume that $R_0>1$,*
(i) If $D_1(L_1(s))>0$, $a_1>0$, $a_3>0$ and $a_1a_2>a_3$, then the endemic equilibrium point M_* is locally asymptotically stable.
(ii) If $D_1(L_1(s))<0$, $a_1>0$, $a_2>0$ and $a_3>0$, then the endemic equilibrium point M_* is locally asymptotically stable for $\alpha\in(0,\frac{2}{3})$.
(iii) If $D_1(L_1(s))<0$, $a_1<0$ and $a_2<0$, then the endemic equilibrium point M_* is unstable for $\alpha\in(\frac{2}{3},1)$.
(iv) If $D_1(L_1(s))<0$, $a_1>0$, $a_2>0$ and $a_1a_2=a_3$, then for $\alpha\in(0,1)$, the endemic equilibrium point M_* is locally asymptotically stable.

Proof Based on the previous discussion, the characteristic equation $L(s)=0$ has three obvious negative roots $s_1=s_2=-\mu$ and $s_3=-\mu-\lambda-\kappa$. In order to investigate the stability of equilibrium point M_*, we only need to discuss the range of the root of $L_1(s)=0$, denoted by s_4, s_5 and s_6.

(i) By the results in [14], if $D_1(L_1(s))>0$, then s_4, s_5 and s_6 are real roots. Further, by Routh-Hurwitz criterion, the necessary and sufficient conditions for $s_i(i=4,5,6)$ to lie in the left half plane are

$$a_1>0,\ \ a_3>0,\ \ a_1a_2>a_3.$$

That is to say, under the above conditions, the roots of $L_1(s)=0$ satisfy

$$|\arg(s_i)|>\frac{\pi}{2}>\frac{\alpha\pi}{2}\ \ (i=4,5,6).$$

Therefore, M_* is locally asymptotically stable and (i) holds

(ii) By the results in [14], if $D_1(L_1(s))<0$, then $L_1(s)=0$ has a real root and a pair of conjugate complex roots, denoted by s_4, $m+ni$ and $m-ni$, respectively. Thus, one has

$$\begin{aligned}L_1(s)=&s^3+a_1s^2+a_2s+a_3\\ =&(s-s_4)(s-m-ni)(s-m+ni).\end{aligned}$$

By calculation,

$$a_1=-s_4-2m,\ \ a_2=2s_4m+m^2+n^2,\ \ a_3=-s_4(m^2+n^2).$$

The conditions $a_1>0$, $a_2>0$ and $a_3>0$ imply that

$$-s_1>2m,\ \ m^2+n^2>-2s_4m,\ \ s_4<0.$$

Further, one has

$$m^2+m^2\tan^2\theta>-2s_4m>4m^2.$$

That is to say $\tan^2\theta>3$, which implies

$$\theta=|\arg(s)|>\frac{\pi}{3}.$$

Therefore, in order to ensure the establishment of $|\arg(s)|>\frac{\alpha\pi}{2}$, we must have $\alpha<\frac{2}{3}$. Thus (ii) holds.

The proof of conclusions (iii) and (iv) is similar to that of conclusion (ii), hence we omit it.

4 Numerical simulations

4.1 Data Sources

The data used in this paper are from the Johns Hopkins University Center for Systems Science and Engineering (https://github.com/CSSEGISandData/COVID-19). the Johns Hopkins University publishes data of accumulated and newly confirmed cases, recovered cases and death cases of COVID-19 from January 22, 2020.

4.2 Analysis of the SEIQRP Model

In the following discussion, the standard incidence rate [18] is used to describe the transmission of COVID-19, and is given by

$$\beta_1 f_1(S)g_1(I) = \frac{\beta_1 SI}{N},\ \beta_2 f_2(S)g_2(E) = \frac{\beta_2 SE}{N},$$

where N represent the total population of the region at the initial time. Hence,

$$\begin{cases} D^\alpha S(t) = \Lambda - \dfrac{\beta_1 S(t)I(t)}{N} - \dfrac{\beta_2 S(t)E(t)}{N} - \mu S(t) - \rho S(t), \\ D^\alpha E(t) = \dfrac{\beta_1 S(t)I(t)}{N} + \dfrac{\beta_2 S(t)E(t)}{N} - \varepsilon E(t) - \mu E(t), \\ D^\alpha I(t) = \varepsilon E(t) - (\delta+\mu)I(t), \\ D^\alpha Q(t) = \delta I(t) - (\lambda+\kappa+\mu)Q(t), \\ D^\alpha R(t) = \lambda Q(t) - \mu R(t), \\ D^\alpha P(t) = \rho S(t) - \mu P(t). \end{cases} \tag{15}$$

The effectiveness of the model (15) in describing the spread of COVID-19 is illustrated by selecting the confirmed, cured and dead cases in the United States. According to the real data provided by the Johns Hopkins University, the outbreak in the United States has not been brought under full control. The data of confirmed cases, cured cases and death cases are selected from January 22, 2020, to April 20, 2020. Assuming that the confirmed individuals will be isolated, then

$$Isolated = Confirmed - Recovered - Death. \tag{16}$$

This hypothesis is in line with the actual situation. Hence, we can obtain the real data of isolated cases. Through the fractional predictor corrector method and the least squares fitting, we can identify the parameters of the model (15) through the real data, which can be found in Table 1.

Table 1: Summary table of the parameter identification for model (15) after using least squares fitting to real data from January 22, 2020, to April 20, 2020

Notation	Parameter Identification
α	0.7182
Λ	3481608
β_1	0.3813
β_2	0.7065
μ	6.7063×10^{-8}
ρ	0.1927
ε	0.2657
δ	0.3352
λ	0.0149
κ	1.1975×10^{-6}

Based on the parameters in Table 1, we can make a simple prediction of isolated cases and recovered cases in the United States, which can be found in Figure 1. We need to emphasize that the peak here represents the number of isolated cases rather than the cumulative number of confirmed cases. Under the parameters in Table 1, it can be calculated by using (9) that $R_0 = 0.2041 < 1$. By the conclusion in Theorem 3, the disease-free equilibrium point is local asymptotic stable. The SEIQRP model has a basic guiding significance for predicting and fitting spreading dynamics of COVID-19. However, the prediction of this model for COVID-19 is relatively rough, we still need to improve model (15) according to actual mitigation policies and research objectives. According to the analysis in Section 2, we choose the SEIQRPD model to predict the trend of the epidemic in the United States under reasonable assumptions.

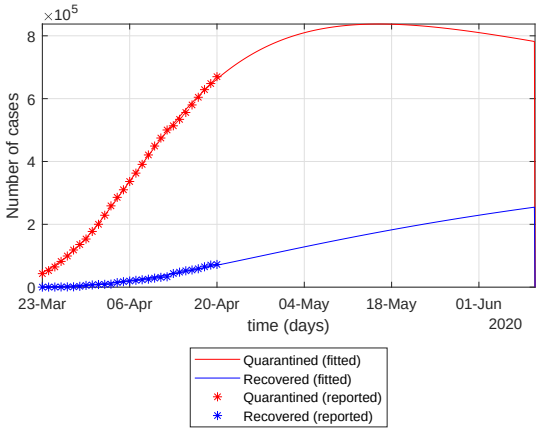

Fig. 2: Number of isolated cases predicted and recovered cases predicted by the model (15) for the United States.

4.3 The SEIQRPD Model for the Prediction of COVID-19

Similarly, the standard incidence rate is used to describe the transmission of COVID-19, the fractional SEIQRPD model can be expressed as

$$\begin{cases} D^\alpha S(t) = -\dfrac{\beta_1 S(t)I(t)}{N} - \dfrac{\beta_2 S(t)E(t)}{N} - \rho S(t), \\ D^\alpha E(t) = \dfrac{\beta_1 S(t)I(t)}{N} + \dfrac{\beta_2 S(t)E(t)}{N} - \varepsilon E(t), \\ D^\alpha I(t) = \varepsilon E(t) - \delta I(t), \\ D^\alpha Q(t) = \delta I(t) - (\lambda(t)+\kappa(t))Q(t), \\ D^\alpha R(t) = \lambda(t)Q(t), \\ D^\alpha P(t) = \rho S(t), \\ D^\alpha D(t) = \kappa(t)Q(t). \end{cases} \tag{17}$$

According to the data provided by Johns Hopkins University, by April 20, 2020, the outbreak in China has been basically controlled. In many provinces of China, the number of new cases per day is increasing in single digits. This

means that the data in China contains more information about the spreading dynamics of COVID-19. Therefore, the data in Hubei, Guangdong, Hunan and Zhejiang are selected to research the fitting effect of the model (17). According to the real data of these four regions in China, the parameters of the model (17) are identified, and the results are shown in Table 2. The model (17) successfully capture the trend of the outbreak, which can be seen in Figure 3.

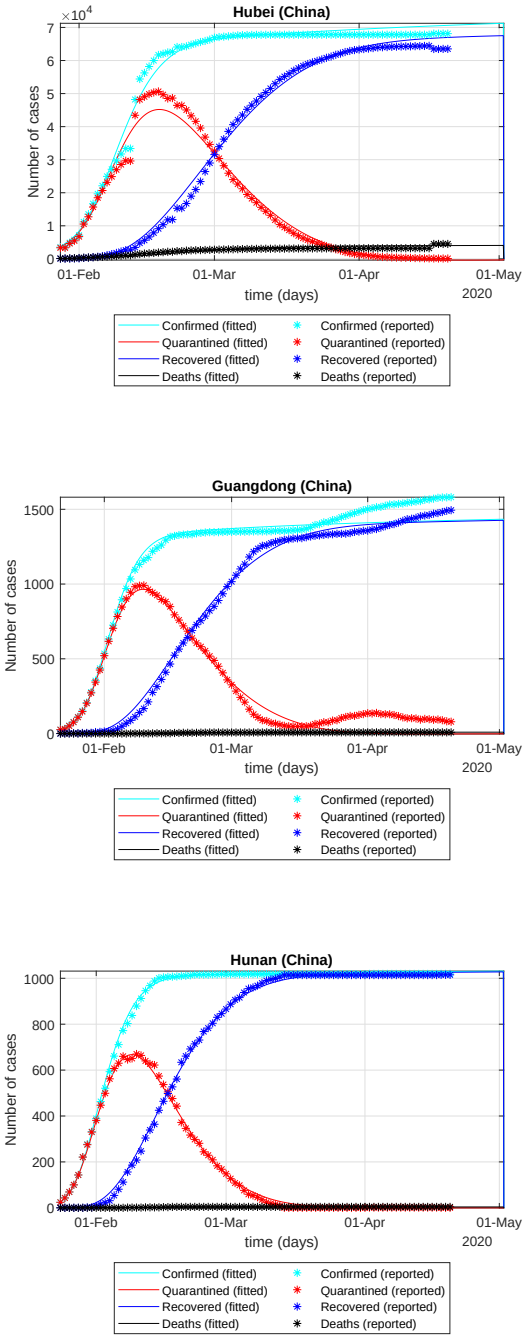

Fig. 3: The fitting effect of the improved model (17) on the outbreak in Hubei, Guangdong, Hunan and Zhejiang.

Table 2: Summary of parameter identification of model (17) (data used from January 22, 2020, to April 20, 2020).

Notation	Hubei	Guangdong	Hunan	Zhejiang
α	1.0729	1.0446	1.0135	1.0482
σ_1	0.3954	0.7057	1.4004	1.9985
σ_2	0.1345	0.0783	0.4752	0.4592
β_2	0.4132	0.6594	0.6908	0.7608
ρ	0.0913	0.1033	0.0702	0.1611
ε	0.2595	0.4095	0.4264	0.3457
δ	0.2283	0.3796	0.9589	0.3665
λ_1	0.6509	0.9695	0.9989	0.9142
λ_2	0.002	0.0019	0.0032	0.0025
κ_1	0.002	2.8571×10^{-4}	0.0012	0.0098
κ_2	6.3088×10^{-6}	2.7547×10^{-14}	0.1017	1.0968

At present, we can obtain the epidemic data of the United States from January 22, 2020 to April 20, 2020. We need to preprocess the data to remove the data smaller than 0.5% of the current highest number of confirmed cases. In order to test the prediction ability of the SEIQRPD model (17) for the development process of the epidemic in the United States, we select the data before April 5 to identify the parameters of the SEIQRPD model (17). Furthermore, in order to illustrate the ability of the SEIQRPD model (17) in predicting the outbreak, we compared the real data and fitted data after April 5, which can be seen in Table 3, Table 4 and Figure 4. According to the results in Table 3 and Table 4, it can be found that the real values of current isolated cases and cumulative confirmed cases fall within the range of 95% - 105% of the predicted values. This shows that the SEIQRPD model (17) can effectively predict the data in the next two weeks.

The data before April 20, 2020 are selected to identify the parameters of the improved model (17), and the results

Table 3: Summary of real and fitted data for the cumulative confirmed cases in the United States from April 6, 2020 to April 20, 2020.

Data	Cumulative confirmed cases		Relative error (%)
	Reported	Predicted	
6, April	366667	370220	0.97
7, April	397505	403441	1.49
8, April	429052	436940	1.84
9, April	462780	470492	1.67
10, April	496535	503882	1.48
11, April	526396	536905	2
12, April	555313	569374	2.53
13, April	580619	601117	3.53
14, April	607670	631983	4
15, April	636350	661839	4.01
16, April	667592	690572	3.44
17, April	699706	718090	2.63
18, April	732197	744318	1.66
19, April	759086	769200	1.33
20, April	784326	792699	1.07

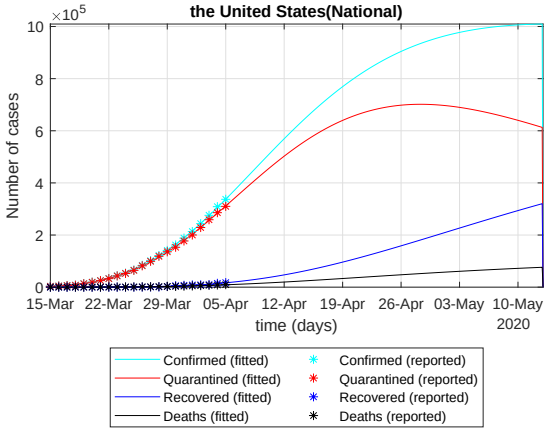

Fig. 4: Based on the data of the United States from January 22 to April 20, 2020 to verify the accuracy of the forecast for the next 15 days.

Table 4: Summary of real and fitted data for the isolated cases in the United States from April 6, 2020 to April 20, 2020.

Data	Isolated cases		Relative error (%)
	Reported	Predicted	
6, April	339897	336303	1.06
7, April	368221	362948	1.43
8, April	396348	390798	1.4
9, April	424049	420826	0.76
10, April	451108	449159	0.43
11, April	477327	474664	0.56
12, April	502526	500306	0.44
13, April	526545	513609	2.46
14, April	549250	534076	2.76
15, April	570525	555929	2.56
16, April	590279	579973	1.75
17, April	608442	604388	0.67
18, April	624968	628693	0.6
19, April	639826	648088	1.29
20, April	653007	669903	2.59

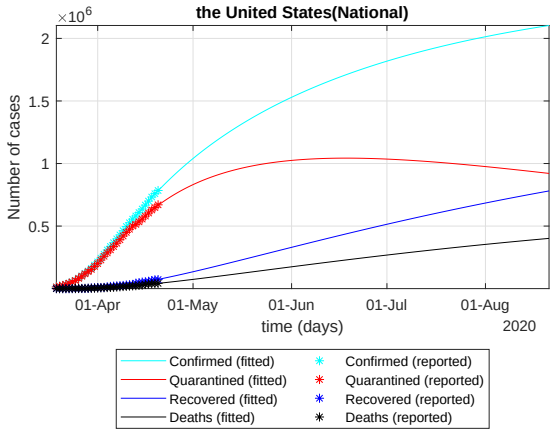

Fig. 5: Based on the data of the United States from January 22 to April 20, 2020, to verify the accuracy of the forecast for the next 15 days.

Table 5: Summary table of the parameter identification for model (17) (data used from January 22, 2020, to April 20, 2020)

Notation	Parameter Identification
α	0.8477
σ_1	0.8601
σ_2	0.1553
β_2	0.3659
ρ	4.3455×10^{-6}
ε	0.3803
δ	0.4024
λ_1	0.0121
λ_2	0.1321
κ_1	0.0064
κ_2	3.7372×10^{-4}

are shown in Table 5. The cumulative number of confirmed cases and the number of isolated cases after two weeks are predicted, which can be seen in Table 6.

Table 6: Summary of predicted data for the United States from April 21, 2020 to May 5, 2020.

Data	Cumulative confirmed cases Predicted	Isolated cases Predicted
21, April	807947	684148
22, April	833465	701725
23, April	858444	718615
24, April	882890	734836
25, April	906816	750409
26, April	930232	765352
27, April	953150	779686
28, April	975584	793432
29, April	997546	806610
30, April	1019050	819239
1, May	1040108	831340
2, May	1060734	842931
3, May	1080941	854031
4, May	1100740	864658
5, May	1120145	874829

The isolated cases in the United States will peak on June 18, 2020, with the peak of 1.0431×10^6, which can be seen in Figure 5.

5 Conclusion

We first propose the fractional SEIQRP model with generalized incidence rates. Some qualitative properties of the SEIQRP model are discussed. In order to predict COVID-19 effectively, we propose an improved SEIQRPD model according to the actual mitigation situation. According to the data of the United States before April 5, 2020, the trend of the outbreak in the United States from April 6 to April 20 is successfully predicted as compared to the real records. Then, using the data before April 20, 2020, we forecast the trend of the outbreak in the United States in the next two weeks, and estimate the peak of isolated cases and the date of the peak.

The improved SEIQRP model proposed in this paper successfully captures the trend of COVID-19. The long-term prediction needs to adjust the model appropriately according to the change of policy and medical level. We will discuss in the future work.

Conflict of Interest

We declare that we have no conflict of interest.

References

1. Lin, Q.Y., Zhao, S. Gao, D.Z., Luo, Y.J., Yang, S., Musa. S.S., Wang, M.H., Cai, Y.L., Wang, W.M., Yang, L., He, D.h.: A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. International Journal of Infectious Diseases. **93**, 211-216 (2020).
2. Peng, L., Yang, W., Zhang, D., Zhuge, C., Hong, L.: Epidemic analysis of COVID-19 in China by dynamical modeling. Cold Spring Harbor Laboratory. (2020). arXiv: 2002.06563.
3. Amjad, S. S., Iqbal, N. S., Kottakkaran, S.N.: A Mathematical model of COVID-19 using fractional derivative: outbreak in India with dynamics of transmission and control. (2020). doi: 10.20944/preprints202004.0140.v1.
4. Chen, Y., Cheng, J., Jiang, X., Xu, X.: The reconstruction and prediction algorithm of the fractional TDD for the local outbreak of COVID-19. (2020). arXiv:2002.10302.
5. Cheng, Z.J., Shan, J.: 2019˙Cnovel Coronavirus: Where We are and What We Know. Infection. (2020). DOI 10.1007/s15010-020-01401-y.
6. Yang, Z.F., Zeng, Z.Q., Wang, K., Wong, S., Liang, W.H., Zanin, M., Liu, P., Cao, X.D., Gao, Z.Q., Mai, Z.T., Liang, J.Y., Liu, X.Q., Li, S.Y., Li, Y.M., Ye, F., Guan, W.J., Yang, Y.F., Li, F., Luo, S.M., Xie, Y.Q., Liu, B., Wang, Z.L., Zhang, S.B., Wang, Y.N., Zhong, N.S., He, J.X.: Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease **12**(2), 165-174 (2020).
7. Zhao, S., Lin, Q, Ran J., Musa, S.S., Yang, G.P., Wang, W.M., Lou, Y.J., Gao, D.Z., Yang, L., He, D.H., Wang, M.H.: Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases. 214-217 (2020).
8. Podlubny, I.: Fractional differential equations, Academic Press, New York, (1999).
9. Diethelm, K.: The analysis of fractional differential equations: an application-oriented exposition using differential operators of Caputo type, Springer Science, Business Media, (2010).
10. Kuniya, T.: Hopf bifurcation in an age-structured SIR epidemic model. Applied Mathematics Letters. **92**, 22-28 (2019).
11. Zhang, X.B., Huo, H., Xiang, H.F., Xiang, H., Meng, X.Y.: An SIRS epidemic model with pulse vaccination and non-monotonic incidence rate. Nonlinear Analysis: Hybrid Systems. 13-21 (2013).
12. Cai, Y., Kang, Y., Wang, W.: A stochastic SIRS epidemic model with nonlinear incidence rate. Applied Mathematics and Computation. 221-240 (2017).
13. Ahmed, E., Elgazzar, A.S.: On fractional order differential equations model for nonlocal epidemics. Physica A. **379**(2), 607-614 (2007).
14. Rocca, A., West, B.J.: Fractional calculus and the evolution of fractal phenomena. Physica A. **265**(3-4), 535-546 (1999).
15. Jalilian, Y., Jalilian, R.: Existence of solution for delay fractional differential equations. Mediterranean Journal of Mathematics. **10**(4), 1731-1747 (2013).
16. Yang, Y., Xu, L.: Stability of a fractional order SEIR model with general incidence. Applied Mathematics Letters, **105**, (2020).
17. Wang, H., Yu, Y.G., Wen, G.G., Zhang, S., Yu, J.Z.: Global stability analysis of fractional-order Hopfield neural networks with time delay. Neurocomputing. **154**15-23, (2015).
18. Hu, Z., Liu, S., Wang, H.: Backward bifurcation of an epidemic model with standard incidence rate and treatment rate. Nonlinear Analysis-real World Applications. **9**(5), 2302-2312 (2008).
19. Sun, H.G., Zhang, Y., Baleanu, D., Chen, W., Chen, Y.Q.: A new collection of real world applications of fractional calculus in science and engineering. Communications in Nonlinear Science and Numerical Simulation. **59**(5), 213-231 (2018).
20. Cao, K.C., Chen, Y.Q.: Fractional order crowd dynamics: Cyber-Human systems modeling and control. (Invited book project. Volume 4 of the De Gryuter Monograph Series 'Fractional Calculus in Applied Sciences and Engineering'). ISBN 978-3-11-047398-8.

21. West, B.J.: Fractional Calculus View of Complexity: Tomorrow's Science, CRC Press, (2015).
22. Li, Y., Chen, Y.Q, Podlubny, I.: Mittag-Leffler stability of fractional order nonlinear dynamic systems. Automatica. **45**(8), 1965-1969 (2009).
23. Lu, Z.Z., Yu, Y.G., Chen, Y.Q., Ren, G.J., Xu, C.H., Yin, Z.: A fractional-order SEIHDR model for COVID-19 with inter-city networked coupling effects. Nonlinear Dynamics (Special Issue on 'Nonlinear dynamics of COVID-19 pandemic: modeling, control, and future perspectives'). (April 2020).