

UC Merced

UC Merced Previously Published Works

Title

RLIM: a recursive and latent infection model for the prediction of US COVID-19 infections and turning points

Permalink

<https://escholarship.org/uc/item/34n3h16w>

Journal

Nonlinear Dynamics, 106(2)

ISSN

0924-090X

Authors

Yu, Xiang
Lu, Lihua
Shen, Jianyi
[et al.](#)

Publication Date

2021-10-01

DOI

10.1007/s11071-021-06520-1

Peer reviewed



RLIM: a recursive and latent infection model for the prediction of US COVID-19 infections and turning points

Xiang Yu · Lihua Lu · Jianyi Shen ·
Jiandun Li · Wei Xiao · Yangquan Chen

Received: 29 January 2021 / Accepted: 4 May 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Initially found in Hubei, Wuhan, and identified as a novel virus of the coronavirus family by the WHO, COVID-19 has spread worldwide at exponential speed, causing millions of deaths and public fear. Currently, the USA, India, Brazil, and other parts of the world are experiencing a secondary wave of COVID-19. However, the medical, mathematical, and pharmaceutical aspects of its transmission, incubation, and recovery processes are still unclear. The classical susceptible–infected–recovered model has limitations in describing the dynamic behavior of COVID-19. Hence, it is necessary to introduce a recursive, latent model to predict the number of future COVID-19 infection cases in the USA. In this article, a dynamic recursive and latent infection model (RLIM) based on the classical SEIR model is proposed to predict the number of COVID-19 infections. Given COVID-19 infection and recovery data for a certain period, the RLIM is able to fit current values and produce an optimal set

of parameters with a minimum error rate according to actual reported numbers. With these optimal parameters assigned, the RLIM model then becomes able to produce predictions of infection numbers within a certain period. To locate the turning point of COVID-19 transmission, an initial value for the secondary infection rate is given to the RLIM algorithm for calculation. RLIM will then calculate the secondary infection rates of a continuous time series with an iterative search strategy to speed up the convergence of the prediction outcomes and minimize the maximum square errors. Compared with other forecast algorithms, RLIM is able to adapt the COVID-19 infection curve faster and more accurately and, more importantly, provides a way to identify the turning point in virus transmission by searching for the equilibrium between recoveries and new infections. Simulations of four US states show that with the secondary infection rate ω initially set to 0.5 within the selected latent period of 14 days, RLIM is able to minimize this value at 0.07 and reach an equilibrium condition. A successful forecast is generated using New York state's COVID-19 transmission, in which a turning point is predicted to emerge on January 31, 2021.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11071-021-06520-1>.

X. Yu (✉) · L. Lu (✉) · J. Shen · J. Li · W. Xiao
School of Electronics and Information, Shanghai Dianji
University, Shanghai 201306, China
e-mail: yux@sdju.edu.cn

L. Lu
e-mail: lulihua@sdju.edu.cn

Y. Chen
Mechatronics, Embedded Systems and Automation Lab, University of California, Merced, CA, USA

Keywords COVID-19 · SEIR · Secondary infections · Recursive time series · Turning point

1 Introduction

Since its first appearance in Hubei, Wuhan, China, in December 2019, a novel virus named COVID-19 has affected millions of people worldwide, causing unpredicted economic losses and public fear. To date, the origin, incubation time, and transmission speed of COVID-19 have not been clarified. Numerous attempts from medical, clinical, and mathematical perspectives have been made to analyze the dramatic increase in infections brought by COVID-19 and predict its transmission trends.

A number of COVID-19-related studies developed their mathematical modeling based on the susceptible–infected–removed (SIR) model, which was originally proposed by Kermack and McKendrick [13] to analyze black death virus transmission occurring in London, the UK, and pestilence in Mumbai, India, in 1666 and 1906, respectively. Theoretically, this model divides the progress of virus transmission into three phases—susceptible, infected, and removed—and relates mathematical parameters with the characteristics of each stage. For example, a mathematical parameter, β , was assumed between susceptible and infected to identify the percentage of the healthy and vulnerable population that transform into a positively infected patient. β has been associated with R_0 , the basic reproductive number, which is widely used by clinical experts to express the average speed of transmission for a specific virus. Another important indicator, γ , has been widely applied to record the percentage that move from infection to recovery or death. The reciprocal of γ indicates the median incubation period of COVID-19 transmission, which has attracted much interest from the scientific community.

Regarding the incubation period of COVID-19, a number of research findings have also been published: Yu et al. [21] investigated COVID-19-infection cases reported in China and other countries and recorded incubation periods ranging from 7 to 14 days; Lai [14] collected exposure periods for 125 Chinese patients, and the estimation indicated that the median incubation period was 4.75. Zhu [22] assumed that the latent period and the infectious period are approximately equal to the incubation period and the length of stay in the hospital and preliminarily concluded that the value of the latent period and the infectious period is 5 and 10 days, respectively. Adhikari [1] asserted that the average incubation duration of COVID-19 was 4.8 ± 2.6 ,

ranging from 2 to 11 days (with 95% confidence interval, 4.1 to 7).

Although numerous mathematical models have been developed to address the dynamics of COVID-19, very few focus on the secondary infections caused by recovery. Many of these models treat COVID-19 as a respiratory disease that requires immediate medical attention but does not last for long or cause secondary effects. However, long-lasting illnesses and secondary outbreaks in the USA, UK, Brazil, and India all indicate that COVID-19 symptoms cannot be treated as terminating in a manner similar to flu. For example, Sabino et al. [18] observed the resurgence of COVID-19 in January 2021 in Brazil and asserted that one of the main reasons behind this resurgence was that immunity against COVID-19 infection had already begun to wane by December 2020. Thus, the recovered group could still be infected or become a virus carrier. According to the COVID tracking project [8], the definition of “COVID-19 recovery” varies among different US regions, ranging from, for example, “symptom improvement” to “hospital discharges” or even “days since diagnosis”. In addition, there is no clear evidence that “recovered” patients are subsequently immune to COVID-19. Thus, it is reasonable and necessary to assume that a portion of them, after a certain period of time, will move from the immune group to the susceptible group. A recent scientific report from Christian Gaebler et al. [9] proves that the humeral memory response to COVID-19 will last between 1.3 and 6.3 months after infection without vaccine support. Okhueuse [16] attempted to estimate the probability of COVID-19 reinfection by searching the equilibrium state of the SEIRUS model. In his simulation report, after 12 days, the rate of recovery and rate of infection will meet and reach an equilibrium state. However, his model merely considered incorrectly executed PCR tests, which is not sufficiently accurate to describe current COVID-19 transmission in the USA. According to Altan and Karasu [2], X-ray images are able to provide better results than RT-PCR tests in the diagnosis of COVID-19 disease.

One of the essential questions to be answered by a forecasting algorithm is when and how a turning point will appear. A turning point within COVID-19 transmission contains valuable information to help governments, clinical services, and scientists model the transmission and prepare. Yang et al. [20] successfully predicted the peak of COVID-19’s first wave in China in

late February under public health interventions. Many studies, such as [7] and [21], relate COVID-19 transmission's turning point with political or public affairs, such as city lockdowns and school closure. Recently, some researchers claimed that the turning point and end of an expanding epidemic cannot be precisely forecasted [3] because COVID-19 transmission is highly dynamic and unstable, and the forecasting results are sensitive to small variations in parameters. The recursive and latent infection model (RLIM) algorithm proposed in this paper can provide a reliable estimation of the COVID-19 transmission turning point due to three factors. First, the authors chose a period of 14 days for prediction, which is not long enough for new effects to emerge and affect the results. Second, the turning point in our system was predicted with validated data records, and the trend in these data records was carefully observed to guarantee their smoothness. Finally, a detailed investigation of state-level regulations was made for the target states and dates to ensure that no political events occurred (such as a state lockdown or hospital emergencies).

In this article, we develop and present the RLIM, a novel COVID-19 transmission model. The main contributions of this paper are as follows:

(1) Developed a novel method to forecast the number of infections in the upcoming 14 days based on historical infection and recovery data. This method is able to efficiently locate the relationship between historical data and infection data and optimize the parameters of the RLIM model in a short period. Evidence from our experiment proves that the key parameters converge within a certain period in the optimized RLIM model, thereby locating the optimal parameters.

(2) Given an infection-recovery dataset for COVID-19, this method is able to promptly locate the turning point with an iterative search strategy. Our experiment shows that within a period of 60 days, an RLIM model with an optimized set of parameters based on historical data (see contribution point 1) is able to predict the secondary infection rates for the coming week with an optimization strategy that minimizes the MSE (maximum square error) between the reported number of infections and RLIM predictions.

The remainder of this manuscript is organized as follows. Section 2 discusses the implemented mathematical modeling, equations and algorithms. Section 3 describes the simulation settings, software and scientific packages utilized by the RLIM program. Sec-

tion 4 discusses the data and simulation results for four US states' COVID-19 and provides predictions on their infections between mid-January and mid-February. Section 5 summarizes the work and offers further discussion.

2 Method

This section discusses the algorithm in three steps. The first step introduces the mathematics behind RLIM in detail, explaining how it evolves from the classical SIR model and describes COVID-19's infection process in a series of equations. The next step assigns mathematical symbols to parameters in RLIM and implements these equations into sequential procedures in our algorithm. Finally, a performance measure on RLIM is proposed to evaluate how the algorithm runs on COVID-19 dataset.

2.1 RLIM: mathematics

In this paper, a modified COVID-19 transmission model is proposed based on the original SIR model by Kermack and McKendrick [13]. They proposed the susceptible–infected–removed model and used it to successfully explain the 1665–1666 plague in London and the 1906 pestilence at Mumbai, India. The SIR model diagram is shown in Fig. 1.

The transmission process is described by Eqs. (1), (2), and (3).

$$\frac{dS_t}{dt} = -\beta \times S_t \times I_t \quad (1)$$

$$\frac{dI_t}{dt} = (\beta \times S_t - \gamma) \times I_t \quad (2)$$

$$\frac{dR_t}{dt} = \gamma \times I_t \quad (3)$$

Their SIR model is only feasible in an ideal epidemic transmission environment because it does not consider the time variance in the infection rate β or recovery rate γ . Additionally, it requires no disease control—any political or clinical intervention is forbidden, and such transmission behavior rarely appears. However, based on these theoretical assumptions, many revised

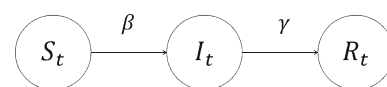


Fig. 1 The SIR model

models, such as SEIR [19], SEIRUS [16], mechanic-statistic SEIR [17], and deep learning SEIR [12], have been proposed and developed by researchers adopting different epidemic transmission characteristics and human interventions. A description of these models can be found in Hethcote’s review [10].

RLIM is inspired by research from Jianping Huang’s team at Lanzhou university [11]. Their model, named as Global Prediction system for COVID-19 Pandemic (GPCP), adds 4 states of disease from SIR model: insusceptible state (P), potentially infected state (E), quarantined state (Q), and mortality state (D). The GPCP disease transmission model is described by Eqs. (4)–(10).

$$\frac{dS_t}{dt} = \frac{-\beta_t \times I_t \times S_t}{N} - \alpha \times S_t \tag{4}$$

$$\frac{dP_t}{dt} = \alpha \times S_t \tag{5}$$

$$\frac{dE_t}{dt} = \frac{-\beta_t \times I_t \times S_t}{N} - \gamma_t \times E_t \tag{6}$$

$$\frac{dI_t}{dt} = \gamma_t \times E_t - \delta_t \times I_t \tag{7}$$

$$\frac{dQ_t}{dt} = \delta_t \times I_t - \lambda_t \times Q_t - K_t \times Q_t \tag{8}$$

$$\frac{dR_t}{dt} = \lambda_t \times Q_t \tag{9}$$

$$\frac{dD_t}{dt} = K_t \times Q_t \tag{10}$$

RLIM, based on our observations and given facts from news reports analysis, adds a symbol ω in the transmission loop. ω represents the probability that a patient who had recovered from COVID-19 for a certain period is again identified by respiratory tests or antibody tests as a virus carrier. Following this definition, ω is used between statuses R and SI , representing the transmission possibility between the recovered group and the secondary infected group (Fig. 2).

Following the assumptions above, for the RLIM, the equation series is modified as Eqs. (11)–(15).

$$\frac{dS_t}{dt} = \frac{-\beta_t \times I_t \times S_t}{N} \tag{11}$$

$$\frac{dE_t}{dt} = \frac{\beta_t \times I_t \times S_t}{N} - \gamma \times E_t \tag{12}$$

$$\frac{dI_t}{dt} = \gamma_t \times E_t + \omega_t \times R_{t-\tau} - \lambda I_t - K I_t \tag{13}$$

$$\frac{dR_t}{dt} = \lambda_t \times I_t - \omega \times R_{t-\tau} \tag{14}$$

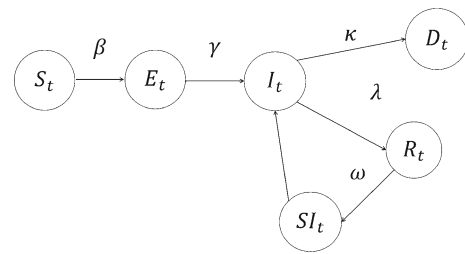


Fig. 2 RLIM model diagram

$$\frac{dD_t}{dt} = K_t \times R_t \tag{15}$$

Compared with GPCP model, the RLIM has the following advantages:

(1) The RLIM simplifies the classical susceptible–infected–quarantined–immune process into a susceptible–infection process due to the maturity of the COVID-19 detection system through PCR tests or other nucleic acid amplification tests approved by CDSE [4]. Given an accurate number of confirmed infections, RLIM focuses on differentiating first-time and secondary infections brought by different groups to achieve more accurate prediction results.

(2) The RLIM improves the GPCP model with recursive state SI and parameter ω to avoid the problem of forward transmission only. Without a recursive state and the existence of parameters, the number of new infections will decrease regardless of the actions taken, and this process would be contradictory to the current US COVID-19 transmission records.

(3) Introduce the latency parameter τ to indicate the median reinfection period. In RLIM, τ is initialized with a value of 14 according to WHO’s instructions and scientific reports. This parameter correlates with the recovery policy in many US states: patients in the hospital will be automatically treated as recovered after a certain period.

To apply Eqs. (11) and (12) in our algorithm, a transform into discrete data series shall be implemented as Eqs. (16) and (17). Replace $I(t)$ and $R(t)$ with fourth stage Eqs. (16) and (17) into Eq. (14) and we have Eq. (18).

$$I(t) = at^3 + bt^2 + ct + d \tag{16}$$

$$R(t) = et^3 + ft^2 + gt + h \tag{17}$$

$$\frac{d(et^3 + ft^2 + gt + h)}{dt} = \lambda \times (at^3 + bt^2 + ct + d) - \omega \times (e(t - \tau)^3 + f(t - \tau)^2 + g(t - \tau) + h) \tag{18}$$

In Eq. (16), a relationship between coefficients $[a, b, c, d]$ and $[e, f, g, h]$ is established; thus, RLIM is able to predict the infected number of cases given historical number of recovery, previous infections, and assumptions of infection rate, recovery rate, and secondary infection period. A detailed description of the corresponding algorithm flow and diagram will be discussed in is given in Fig. 3 and Sect. 2.2.

2.2 RLIM: algorithm

The RLIM algorithm calculates predicted infection numbers according to Eqs. (16)–(18) and optimizes the difference between actual data recorded by the COVID tracking project and the predicted numbers returned from our model. Initially, the predicted recovery numbers R_1, R_2, \dots, R_n are calculated by the fourth-order method based on the real data series of a selected state from the USA between November 2020 and January 2021.

2.2.1 Notations

The notations used throughout this article are described in Table 1. The coefficients (e, f, g, h) associated with this recovery function will then be transformed into other coefficients (a, b, c, d) , with the preassigned recovery rate λ .

2.2.2 RLIM algorithm

The RLIM algorithm calculates predicted infection numbers according to Eqs. (16)–(18) and optimizes the difference between actual data recorded by the COVID tracking project and the predicted numbers returned from our model. Initially, the predicted recovery numbers R_1, R_2, \dots, R_n are calculated by the fourth-order method based on the real data series of a selected state from the USA between November 2020 and January 2021. The coefficients (e, f, g, h) associated with this recovery function will then be transformed into other

coefficients (a, b, c, d) , with the preassigned recovery rate λ (default value 0.01) and secondary infection rate ω (default value 0.01). With coefficients (a, b, c, d) assigned, the number of newly infected cases I_1, I_2, \dots, I_n within this state can be calculated. Comparing these predicted numbers with actual data, one can evaluate and justify whether this round of prediction is accurate or not. Our RLIM model will continuously search for the optimal infection series and then determine the optimal ω associated with this state.

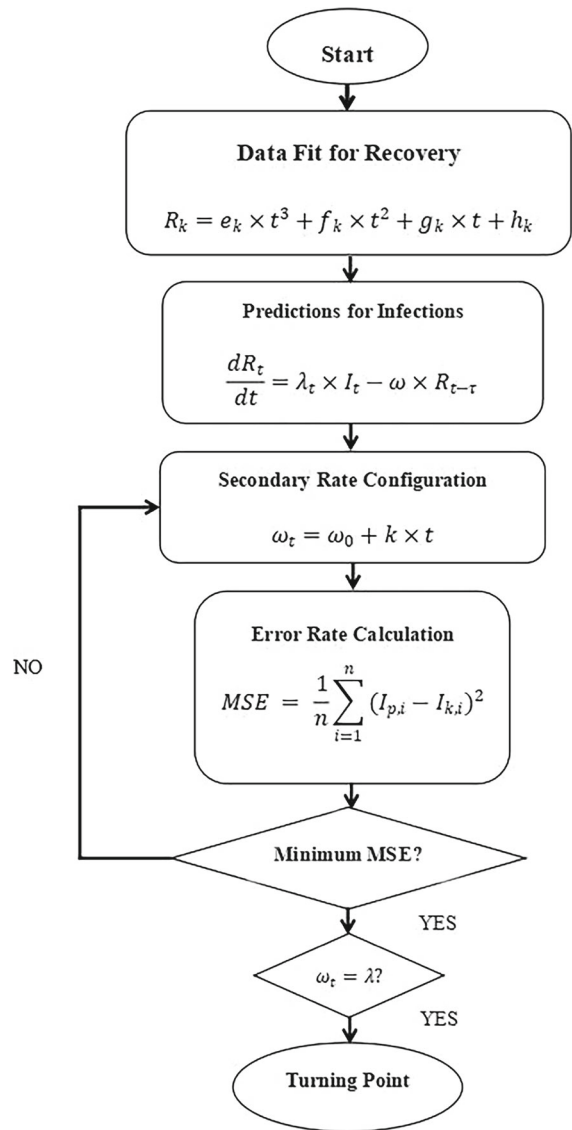


Fig. 3 The RLIM model diagram

Table 1 Notations in RLIM algorithm

I_k	Number of newly infected cases on k^{th} date.
R_k	Number of newly recovered cases on k^{th} date.
$I_{p,k}$	Number of predicted infected cases on k^{th} date.
$R_{p,k}$	Number of predicted recovered cases on k^{th} date.
ω	Probability of secondary infections after δ days.
λ	Probability of recovery from infected cases.
τ	Time interval between recovery and secondary infection in days.
e, f, g, h	Coefficients returned by fourth method on recovered cases.
a, b, c, d	Coefficients calculated by Eq. (16) on infected cases.

Algorithm 1 RLIM

Input:
 initial parameters ω, λ , maximum iteration step $Nmax$, maximum optimization step $Omax$, error tolerance Te for convergence criterion.

- 1: Load the data set I_k, R_k from target U.S. state,
- 2: **for** $k = 1 \rightarrow Nmax$ **do**
- 3: Calculate coefficients $[e, f, g, h]$ for loaded dataset R_k with equation (17),
- 4: **for** $n = 1 \rightarrow Omax$ **do**
- 5: Calculate the secondary infection rate ω_n ,
- 6: Transform coefficient series $[e, f, g, h]$ into $[a, b, c, d]$ according to equation (18),
- 7: Calculate the fitted infections $I_{p,k}$ with coefficient $[a, b, c, d]$ with equation (16),
- 8: Calculate the MSE according to equation (19) between loaded data I_k and fitted data $I_{p,k}$,
- 9: **if** $MSE_{k,n} \leq Te$ **then**
- 10: **return** current value ω_n ,
- 11: **break**,
- 12: **if** $\omega_n == \lambda$ **then**
- 13: set turning point indicator == TRUE,
- 14: **break**,
- 15: **return** Result

2.3 Performance measure

RLIM’s performance measure is calculated as the difference between its predictive output I_p and actual value I_k . Three performance indicators are given: the

mean square error (MSE), standard deviation, and average forecasting error rate (AFER). Because different US states have quite different numbers of infections, ranging from hundreds to thousands, these indicators will be uniform between 0 and 1 to justify the performance.

- Mean square error (MSE)
 The average of squared difference between RLIM’s predictive output I_p and actual value I_k can be calculated in (19).

$$MSE = \frac{1}{n} \sum_{i=1}^n (I_{p,i} - I_{k,i})^2 \tag{19}$$

- Root mean square error (RMSE)
 The root mean square error is also used to evaluate RLIM’s prediction quality. Its formulation is in Eq. (20).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (I_{p,i} - I_{k,i})^2}{n}} \tag{20}$$

- Average forecasting error rate (AFER) The average forecasting error rate is the percentage of error, which represents the relative difference between the predicted output I_p and the actual value I_t . It is a cumulative statistic capturing deviation between two time series. The $AFER$ is calculated in (21).

$$AFER(\%) = \frac{1}{n} \sum_{i=1}^n \left| \frac{I_{p,i} - I_{k,i}}{I_{k,i}} \right| \times 100; \tag{21}$$

3 Experimental setup

In this section, the data set, the source code, and the software packages which have been used in RLIM are explicitly listed for researchers who are interested in our research and have the intention to re-produce our simulations.

3.1 Data source

The data source directly applied in our simulation is from [6]. This data set contains US state-level data on COVID-19, starting from April 2020 until December 2020. In this article, New Jersey (NJ), New York (NY), South Dakota (SD), and Virginia (VG) are selected because they all have daily tracking recovery reports.

Table 2 RILM simulation report on infections, New York, November 2020–January 2021

No.	Mid-Nov. 2020		Mid-Dec. 2020		Mid-Jan. 2021	
	I_k	$I_{p,k}$	I_k	$I_{p,k}$	I_k	$I_{p,k}$
1	3649	2569	10353	9345	N/A	17324
2	3490	2693	9998	9632	N/A	17480
3	5088	2827	10914	9919	N/A	17626
4	5294	2970	12697	10207	N/A	17762
5	5310	3123	9919	10495	N/A	17889
6	5468	3285	9957	10783	N/A	18005
7	5972	3455	9007	11070	N/A	18112
8	5392	3634	9716	11356	N/A	18207
9	5906	3821	11937	11641	N/A	18291
10	4881	4016	12568	11925	N/A	18365
11	6265	4218	12446	12206	N/A	18426
12	6933	4427	10806	12486	N/A	18475
13	8176	4643	7623	12763	N/A	18512
14	6063	4866	10407	13038	N/A	18537
15	6723	5094	11438	13309	N/A	18549
16	6819	5329	13422	13577	N/A	18547
17	7285	5569	16802	13842	N/A	18532
18	8973	5814	16497	14102	N/A	18503
19	9855	6065	15074	14358	N/A	18460
20	11271	6320	11368	14610	N/A	18403
21	10761	6579	11209	14857	N/A	18331
22	9702	6843	12666	15098	N/A	18244
23	7302	7110	16648	15334	N/A	18141
24	9335	7380	17636	15564	N/A	18023
25	10600	7654	18832	15789	N/A	17889
26	10178	7931	16943	16006	N/A	17738
27	10595	8210	15355	16217	N/A	17571
28	11129	8491	13714	16421	N/A	17388
29	10194	8774	15214	16618	N/A	17186
30	9044	9059	14577	16807	–	–
31	–	–	13661	16988	–	–
MSE	5845242.567		4630069.903			
RMSE	2417.693646		2151.759722			
AFER(%)	29.02461097		14.96828112			

RLIM relies heavily on accurate and reliable recovery case reports, and these states have highly credible recovery data sources.

Parameter initializations for RLIM are the same for all the states:

(1) Data fitting period: November 15, 2020–January; 15, 2021;

(2) Prediction: January 16, 2021–February; 15, 2021;

(3) Recovery rate: $\lambda = 0.07$;

(4) Secondary infection rate: $\omega = 0.5$;

(5) Latency period: $\tau = 14$;

3.2 Software implementation

The programming language inside RLIM is PYTHON version 3.7, and the essential software package used is SCIPY version 1.5.4. Two software modules are inherited from SCIPY: integrate and optimize. RLIM utilizes the integrate function to calculate the MSE and the optimize function to fit the real recovery data into fourth-order parameters.

3.3 Code availability

RLIM software is publicly available on GitHub [5], with all codes and implementations available for research. The simulation results are also available upon request.

4 Results and discussion

In this section, the authors present our simulation results in figures and tables and discuss how RLIM located the turning points for four US states. Our discussion starts from New York state, where RLIM successfully located the turning point through COVID-19's epidemic data records. Then, we present a detailed discussion of New York's turning point to show its relationship of the re-infection rate, which we believe is the key. Finally, we describe RLIM's performance on the other US states with predictions of their COVID-19 transmission trends.

4.1 Prediction with MSE/RMSE/AFER

Observations from Fig. 4 indicate that RLIM successfully fits the reported data records from mid-November until mid-January and provides predictions for the upcoming weeks. Delving into the 2 columns of Fig. 4, a conclusion can be drawn that different assignments to the secondary infection rate will result in different outcomes and prediction series. For example, a ω equal to 0.2 will produce a curve of infection numbers with a peak number of 18549, while in the case of 0.3, the peak value is forecast to be 24866. The simulation results suggest that adoption of an iterative search strategy enables RLIM to match real numbers from COVID-19 infection reports. Thus, the relative error between predictions and observations should be minimized. The

other advantage of RLIM is its ability to foretell the turning point within a certain period. Forecasts of the turning point on November 31, 2021, are also marked in the left-column figures. (A discussion of the turning point calculation is presented in Sect. 4.2.)

In Table 2, the number of new infections in New York State caused by COVID-19 is predicted starting on November 15, 2020. The data records of infections were collected from the COVID tracking project during November 2020.

The ranges of infection numbers between November, December, and January are [2600, 9000], [9300, 17,000], and [17,300, 18,500]. Observations from Fig. 4 indicate that the value $\omega = 0.22$ fits infected cases well. Regarding the MSE, RMSE, and AFER indicators, the RLIM reaches 5.8 million absolute errors over 60 days with an AFER of 29.02%. In December 2020, it obtained a better MSE of 4.6 million and a lower AFER of 14.97%. Considering that RLIM's objective is to fit the actual infection numbers and predict the trend, it can be concluded that RLIM achieves a satisfactory result in fitting the actual data and converging the MSE. RLIM's prediction results suggest that COVID-19 transmission in New York state will reach an equilibrium after January 31, 2021, with new infections remaining at a level of 1.8 k per day. New infections will not bring an abrupt change in numbers, so clinical services such as hospitals will not require extra measures.

In Table 3, the number of new recoveries in New York state is also predicted for the months of November, December, and January, with MSE/RMSE/AFER calculated against authentic data records. In November 2020, the MSE of recovery case prediction reached 27893.26, with an AFER of 38.8%. For predictions in December 2020, the MSE increased to 43394.77, but AFER improves to 25.6%. Predictions of recovery data indicate that new recoveries will remain at the level of eight hundred per day, with a flattened tail after mid-January 2021.

4.2 The turning point

From the RLIM's output on New York state's infection and recovery numbers, one can observe that the turning point of this state's COVID-19 transmission occurs around January 30, 2021. Predictions indicate that from mid-January, New York's infections will

Fig. 4 RILM simulation outputs, New York, USA

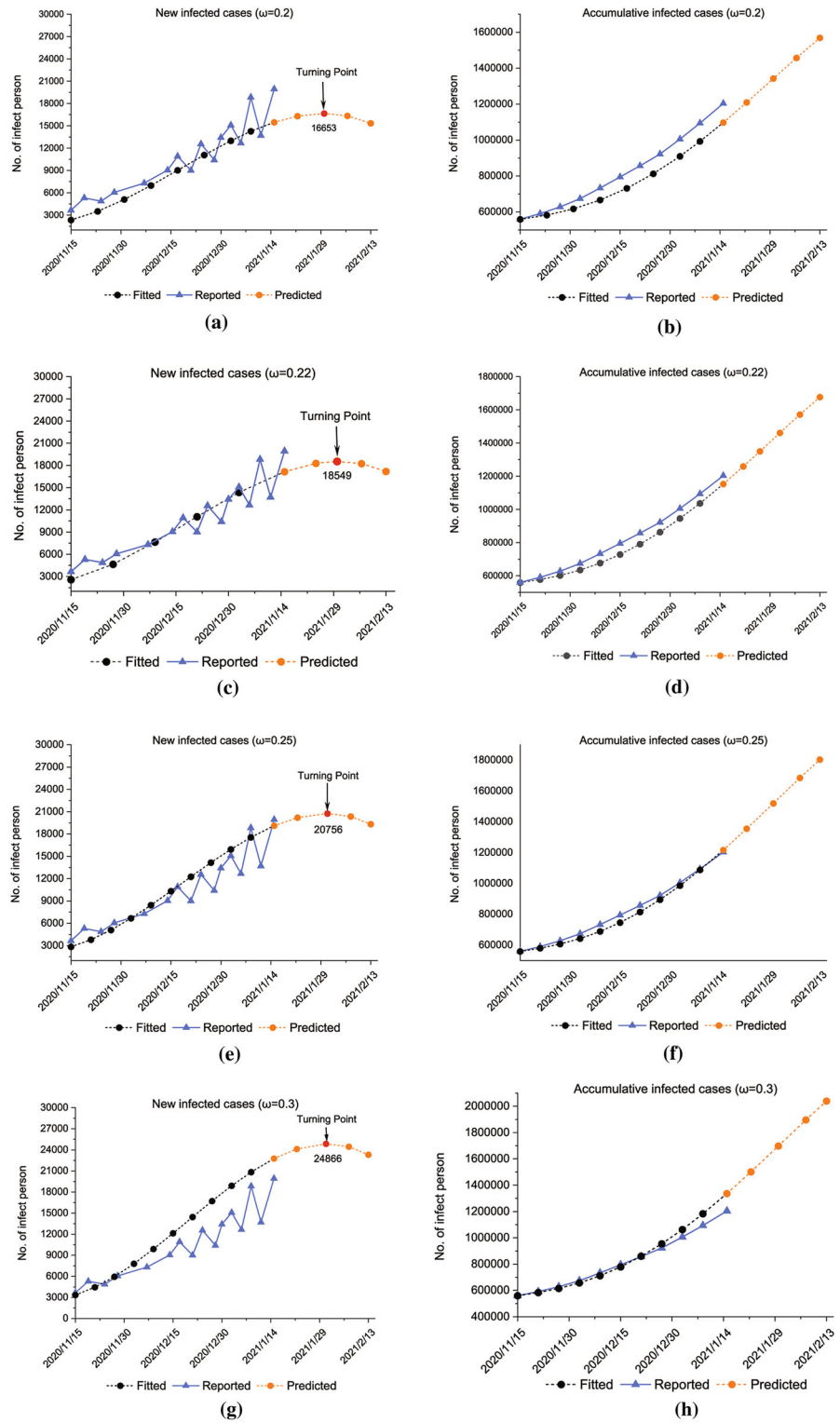


Table 3 RILM simulation report on recoveries, New York, November 2020–January 2021

No.	Mid-Nov 2020		Mid-Dec. 2020		Mid-Jan. 2021	
	\bar{R}_k	$R_{p,k}$	\bar{R}_k	$R_{p,k}$	\bar{R}_k	$R_{p,k}$
1	120	101	599	363	N/A	737
2	114	104	683	375	N/A	746
3	200	108	639	388	N/A	754
4	285	112	522	400	N/A	762
5	259	117	600	413	N/A	769
6	265	122	600	426	N/A	776
7	276	128	406	438	N/A	783
8	200	134	672	451	N/A	789
9	194	141	743	464	N/A	794
10	300	148	750	477	N/A	800
11	338	155	706	490	N/A	805
12	384	163	527	502	N/A	809
13	215	171	425	515	N/A	813
14	349	179	434	527	N/A	816
15	269	188	853	540	N/A	819
16	252	197	834	552	N/A	821
17	393	206	839	565	N/A	823
18	300	216	860	577	N/A	825
19	337	226	574	589	N/A	825
20	635	236	537	601	N/A	825
21	376	247	640	613	N/A	825
22	400	257	800	624	N/A	824
23	335	268	864	636	N/A	822
24	505	280	901	647	N/A	819
25	511	291	891	658	N/A	816
26	552	303	947	669	N/A	813
27	595	314	618	680	N/A	808
28	619	326	541	690	N/A	803
29	300	338	882	700	N/A	797
30	470	351	956	710	–	–
31	–	–	940	719	–	–
MSE	27893.26667		43394.77419			
RMSE	167.012774		208.3141238			
AFER(%)	38.80627859		25.62371705			

slowly increase from 17,324 to 18,549 and then fall back to 17,816 in mid-February. This turning point appears because the secondary infection rate ω , whose range is $[0.1, 0.55]$, experiences high fluctuations in early November 2020 and then drops below 0.4 during December, while after Christmas 2020, it becomes stable around a value of 0.25.

The process of how the RLIM algorithm begins from an initial value and quickly evolves within one month to reach a turning point is clearly illustrated in Fig. 5. The secondary infection rate ω is given a value of 0.5 on November 15, 2020, and the goal is to reach an equilibrium state where ω decreases to 0.07, the same as the value of λ . The optimistic case is that it will

follow a straight pathway to reach 0.07 on a certain date. However, its path needs to be justified according to the actual secondary infection data reported from New York state, recorded as the series of blue points in Fig. 5. Our model produces a series of fitted measurements, which are marked in black in Fig. 5. These measurements represent optimal secondary infection values, chosen by an iterative search algorithm, with a near-optimal error based on the real values. The model forecast starts on January 14, 2021, and successfully reaches an equilibrium state after 2 weeks. It is reasonable to assert that the RLIM model is able to find an optimal curve for the secondary infection rate with an acceptable MSE and follows the curve to predict the forthcoming secondary infection rate. According to the simulation reports in Table 2, RLIM’s prediction results reduce the absolute error of new infections from 5.8 to 4.6 million, with an AFER from 29 to 15%.

4.3 New Jersey, South Dakota and Virginia

The simulation results shown in Fig. 6 indicate three scenarios: moderate increase (New Jersey) at the top, moderate decrease (South Dakota) in the middle, and exponential increase (Virginia) at the bottom. The optimal secondary infection rate ω for these states is marked above (0.13 for New Jersey, 0.056 for South Dakota and 0.19 for Virginia). Observations from these states’ infection and recovery data indicate no strong correlations between ω and COVID-19 transmission trends. Revisiting Eqs. (11) and (15) from Sect. 2.1

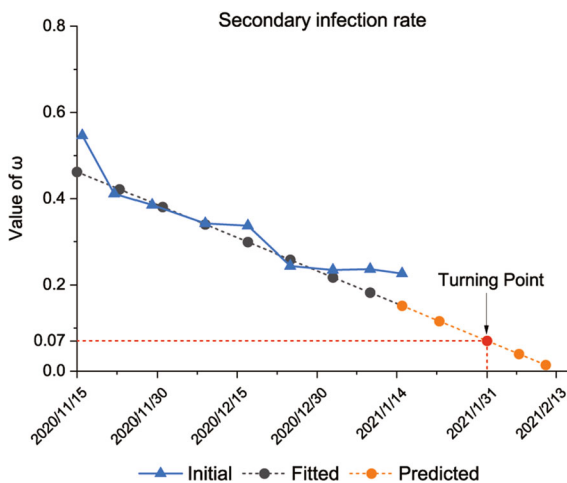


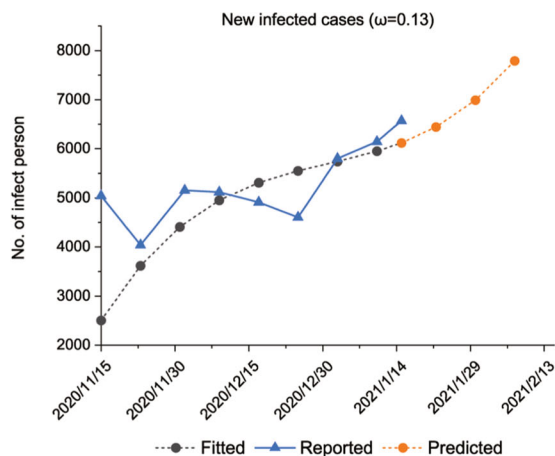
Fig. 5 New York predictions with turning points

explains that in RLIM, ω affects the incremental steps of infection cases positively and recovery cases negatively. However, it is still valuable to predict the turning point when it approaches the value of the recovery rate. Thus, one can conclude that if the recovery rate λ remains stable during the periods of τ (in RLIM, τ equals 14), then RLIM will approach it during a period of time and lock down the turning point. This will greatly reduce the time needed for scientists to elaborate on COVID-19’s behavior.

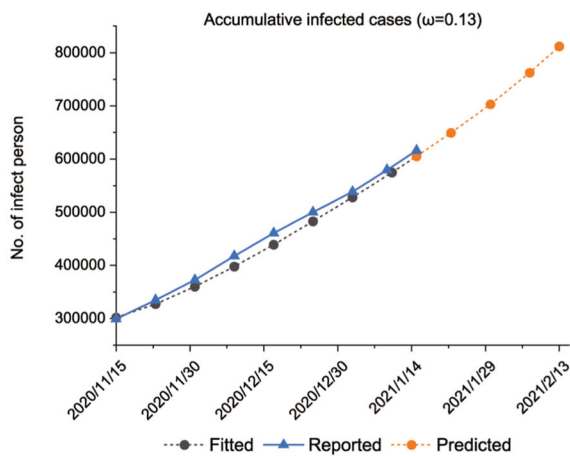
5 Conclusion and future work

This research proposes a recursive, latent, dynamic virus transmission model based on the classical SEIR model. This model, named RLIM, is able to fit the COVID-19 transmission data of the USA and efficiently locate the transmission turning point. Introducing a new parameter ω into the classical SEIR model, RLIM is able to predict newly infected cases based on recovered data and historical COVID-19 records. Experimental results for New York, New Jersey, South Dakota, and Virginia prove that given a reasonable initial value of ω , RLIM is able to predict 30-day infections and recoveries with a reasonable error rate. RLIM also provides an estimation of ω in the time domain, suggesting that it is valuable to explore its approximation and locate the future turning point in COVID-19 transmission. Simulations on New York, dated from mid-November 2020 until the end of January 2021, provide valuable information for ω ’s curve and predict that it reaches an equilibrium state on the 31st of January. Our conclusion, based on the RLIM results, indicates that starting in February, New York state’s COVID-19 transmission will enter an equilibrium state.

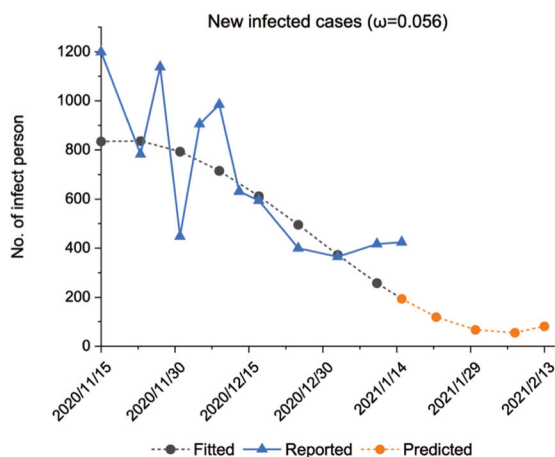
One of RLIM’s advantages is that it does not include environmental factors, such as weather changes, hospital capacity, or city lockdowns. Thus, it is suitable for the prediction of COVID-19 transmission without additional information. Our model is effective for virus modeling including a second wave of COVID-19 epidemic transmission, with key factors such as incubation period and infection rate statistically determined in advance. Compared with other predictive algorithms, RLIM predicts infection numbers based on optimal parameter set from 14-day historical records. The authors believe that this time span would be beneficial to avoid data overfit issue as mentioned in [3].



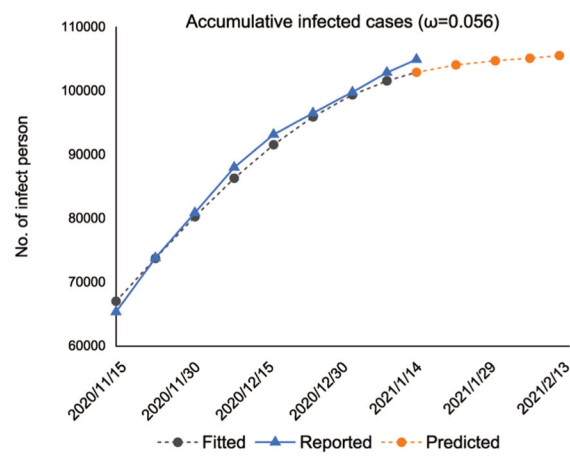
(a) New Jersey



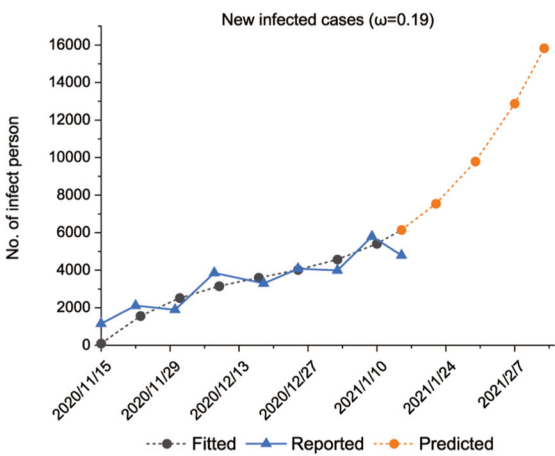
(b) New Jersey



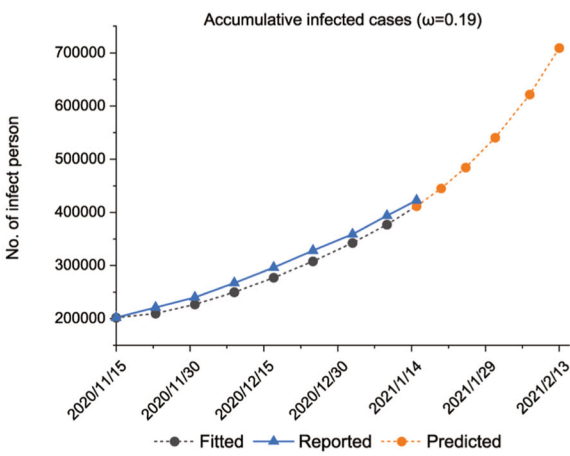
(c) South Dakota



(d) South Dakota



(e) Virginia



(f) Virginia

Fig. 6 RLIM simulation outputs, New Jersey (above), South Dakota (middle), and Virginia

In a word, RLIM is believed to propose a novel yet effective solution in COVID-19 prediction and turning point estimation.

A promising field of application is to integrate RLIM with machine learning techniques. RLIM's recursive, latent status is suitable for description with a back propagation process inside a neural network, so it can be easily equipped with self-learning abilities. Another interesting yet unexplored subject is to use RLIM in prediction of COVID-19's vaccine impact. Researchers may use RLIM to evaluate different kinds of vaccines' impact on COVID-19 transmission. RLIM is able to generate optimal set for pre and post vaccine inoculation group, and with these parameters' visualized, researchers and governments are able to justify the certain kind of vaccine's effectiveness on COVID-19 transmission.

Acknowledgements This work is part of a project supported by the National Science Foundation for Young Scientists of China (Grant No. 61703267).

Data Availability Statement The data that support the findings of this study are available from The COVID Tracking Project: <http://covidtracking.com>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Adhikari, S.P., Meng, S., Wu, Y.-J., Mao, Y.-P., Ye, R.-X., Wang, Q.-Z., Sun, C., Sylvia, S., Rozelle, S., Raat, H., Zhou, H.: Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infect. Dis. Poverty* (2020). <https://doi.org/10.1186/s40249-020-00646-x>
- Altan, A., Karasu, S.: Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos Solitons Fractals* (2020). <https://doi.org/10.1016/j.chaos.2020.110071>
- Castro, M., Ares, S., Cuesta, J.A., Manrubia, S.: The turning point and an expanding epidemic cannot be precisely forecast. *PNAS* **42**, **117**, 26190–26196 (2020). <https://doi.org/10.1073/pnas.2007868117>
- Coronavirus Disease 2019 (COVID-19) 2020 Interim Case Definition, Approved August 5, 2020, National Notifiable Diseases Surveillance System (NNDSS). [https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/\(2020\)](https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/(2020)). Accessed 15 January (2020)
- d RILM project code repository, <https://github.com/secondary-infection/covid-19>
- Dong, E., Du, H., Gardner, L.: An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf. Dis.* (2020). [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Easton, R.: White, laurent hebert-dufresne: state-level variation of initial COVID-19 dynamics in the United States: the role of local government interventions. *PLoS ONE* (2020). <https://doi.org/10.1371/journal.pone.0240648>
- French, A., Nguyen, Q.P.: Recovery data and the COVID Tracking Project. [https://covidtracking.com/analysis-updates/why-its-hard-to-count-recovered/#footnote-3\(2021\)](https://covidtracking.com/analysis-updates/why-its-hard-to-count-recovered/#footnote-3(2021)). Accessed 31 Jan (2021)
- Gaebler, C., et al.: Evolution of antibody immunity to SARS-CoV-2. *Nature* (2020). <https://doi.org/10.1038/s41586-021-03207-w>
- Hethcote, H.W.: The mathematics of infectious diseases. *Siam Rev.* **42**(4), 599–653 (2000)
- Huang, J., Liu, X., Zhang, L., Yang, K., Chen, Y., Huang, Z., Liu, C., Lian, X., Wang, D.: The amplified second outbreaks of global COVID-19 pandemic. *medRxiv*.(2020) <https://doi.org/10.1101/2020.07.15.20154161>
- Jo, H., Son, H., Hwang, H.J., Jung, S.Y.: Analysis of COVID-19 spread in South Korea using the SIR model with time-dependent parameters and deep learning. *medRxiv* (2020). <https://doi.org/10.1101/2020.04.13.20063412>
- Kenmack, W.O., McKendrick, W.G.: A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A* **115**, 700–721 (1927)
- Lai, C.C., Liu, Y.H., Wang, C.Y., Wang, Y.H., Hsueh, S.C., Yen, M.Y., Ko, W.C., Hsueh, P.R.: Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. *J. Microbiol. Immunol. Infect.* (2020). <https://doi.org/10.1016/j.jmii.2020.02.012>
- Mezzetti, D.: COR-19 Transmission, incubation, environment,; <https://www.kaggle.com/davidmezzetti/cord-19-transmission-incubation-environment> (2020), Accessed 15 June (2020)
- Okhue, V.A.: Estimation of the probability of reinfection with COVID-19 coronavirus by the SEIRUS model. *JMIR Public Health Surveil.* (2020). <https://doi.org/10.1101/2020.04.02.20050930>
- Roques, L., Klein, E., Papaix, J., Sar, A., Soubeyrand, S.: Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology* (2020). <https://doi.org/10.3390/biology9050097>
- Sabino, E.C., Buss, L.F., Carvalho, M.P.S., Prete, C.A., Crispim, M.A.E., Fraiji, N.A., Pereira, R.H.M., Parag, K.V., da Silva Peixoto, P., Kraemer, M.U.G., Oikawa, M.K., Salomon, T., Cucunuba, Z.M., Castro, M.C., de Souza Santos, A.A., Nascimento, V.H., Pereira, H.S., Ferguson, N.M., Pybus, O.G., Kucharski, A., Busch, M.P., Dye, C., Faria, N.R.: Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* (2021) [https://doi.org/10.1016/s0140-6736\(21\)00183-5](https://doi.org/10.1016/s0140-6736(21)00183-5)
- Sun, P., Li, K.: An SEIR model for assessment of current COVID-19 pandemic situation in the UK. *medRxiv* (2020). <https://doi.org/10.1101/2020.04.12.20062588>

20. Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J.: Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* (2020). <https://doi.org/10.21037/jtd.2020.02.64>
21. Yu, Y., Liu, Y.-R., Luo, F.-M., Tu, W.-W., Zhan, D.-C., Yu, G., Zhou, Z.-H.: COVID-19 asymptomatic infection estimation. *medRxiv* (2020) <https://doi.org/10.1101/2020.04.19.20068072>
22. Zhu, H., Li, Y., Jin, X., Huang, J., Liu, X., Qian, Y., Tan, J.: Transmission dynamics and control methodology of COVID-19: a modeling study. *Appl. Math. Model.* (2020). <https://doi.org/10.1016/j.apm.2020.08.056>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.