RLIM: a recursive and latent infection model for COVID-19 prediction and turning point in United States

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

Abstract Initially found at Hubei, Wuhan and identified as a novel virus of coronavirus family by WHO, COVID-19 has spread worldwide with an exponentially speed, causing millions of death and public fear. Currently, COVID-19 has brought a secondary wave within U.S., India, Brazil and other parts of the world. However, its transmission, incubation, and recovery processes are still unclear from the medical, mathematical and pharmaceutical aspects. Classical Suspect-Infection-Recovery model has limitations to describe the dynamic behavior of COVID-19. Hence, it becomes necessary to introduce a recursive, latent model to predict the number of future COVID-19 infected cases in U.S. In this article, a dynamic model called RLIM based on classical SEIR model is proposed to predict the number of COVID-19 infections with a dynamic secondary infection rate ω in assumption. An intermediate state called SI is introduced between recovery and infection statues to record the number of secondary infected cases from a latent period of recovery. Compared with other models, RLIM fits historical recovery cases and utilizes them to predict future infections. Because RLIM utilizes multiple information sources, and provides error back propagation schematics, it is reasonable to assert that its predictions are more accurate and persuasive. Projections of four U.S. COVID-19 states show that with the secondary infectious rate ω varies from 0.01 to 0.3 within a latent period of 14 days chosen, RLIM can predict the newly infected number from January 15 to February 15, 2021 with AFER lower to

Xiang Yu, Lihua Lu, Jianyi Shen, Jiandun Li, Wei Xiao
School of Electronics and Information, Shanghai Dianji University, Shanghai, 201306, China
The corresponding authors are Xiang Yu and Lihua lu, Tel.: +86-18616994443, +86-13585804102
E-mail: yux@sdju.edu.cn,lulihua@sdju.edu.cn

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

14%. It also successfully estimates the turning point of New Yorks infections in January 2021, based on current data records.

Keywords COVID-19 · SEIR · secondary infections · recursive time series · turning point

1 Introduction

Since its first appearance in Wuhan, Hubei, China from December 2019, a novel virus named as COVID-19 has affected millions of people all over the world, which has caused unpredictable economic losses and public fear. Until now, COVID-19s origin, its incubation time and transmission speed are still waiting for clarifications. Numerous attempts from medical, clinical, and mathematical aspects have been made around COVID-19, trying to answer the dramatic increase of infections brought by it and predict its transmission trends.

A certain number of COVID-19 related researches developed their mathematical modeling based on the Susceptible-Infective-Removal (SIR) model, which was originally proposed by Kermack and McKendrick [1] to analyze Black Death virus transmission occurred at London, United Kingdom and pestilence at Mumbai India, 1666. Theoretically, this model divides a virus transmission progress into three phases: suspect, infection and removal, and relates mathematical parameters with characteristics of each stage. For example, a mathematical parameter: β, was assumed between suspect and infection to identify the percentage of a healthy and vulnerable person transforming into a positively infected patient. The β has been associated with R_0: basic reproductive number, which express the average speed of a certain virus transmission. Another indicator:γ, was widely applied to record the percentage from infection to recovery status. The reciprocal of γ indicated the median incubation period of COVID-19 transmission, which attracted numerous interests of scientific community.

In response to the COVID pandemic, the White House and a coalition of leading research groups have prepared the COVID-19 Open Research Dataset (CORD-19) on Kaggle [4]. One of the challenges is to explore the COVID-19 transmission, incubation and environment facts from over 400,000 scholarly articles within CORD-19. According to CORD-19s record, many scientists began their investigation of COVID-19 from Wuhan, Hubeis patients records since April 2019, but soon dived into Italy, United States, Brazil and India as well. Anatoly Zhiljavsky et al. [6] developed a mathematical model based on SIR model to analyze COVID-19 United Kingdom epidemics, asserting that a reproduction number R_0 equals 2.5 is suitable for megacities such as London, Manchester and Sheffield. Virginia E. Pitzer et al. [7] delved into the diagnostic testing practices in United States, and estimated that United States had a higher reproduction number R_0 ranging from 4.0 to 7.1 compared with China. Easton R. White et al. [8] claimed that government interventions in the United States resulted into an SIR model with a dynamic, time-variant R_0.

In our modeling, R_0 is also determined as a time-variant parameter, ranging from 1 to 10.

Regarding to the incubation period of COVID-19, a certain number of research findings were also published: Lu [9] investigated 2015 COVID-19 infected cases and recorded their incubation periods ranging from 0 to 33 days. Lai [10] collected 125 Chinese patients exposure periods and the estimation indicated that the median incubation period was 4.75. Zhu [11] assumed that the latent period and the infectious period are approximately equal to the incubation period and the length of stay in hospital, and concluded that the value of the latent period and the infectious period are preliminarily 5 to 10 days, respectively. Adhikari [12] 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). In this article, the value of γ is ranging from 0.03 (33 days) to 0.25 (4 days) based on the scientific findings mentioned above.

Although numerous mathematical models were developed to address the dynamics of COVID-19, very few of them focus on secondary infections caused from recovery. Many of these models treat COVID-19 as a respiratory disease, which requires immediate medical attention but will not last for long and cause secondary affections. However, the long-lasting phenomenon and secondary outbreaks from United States, United Kingdom, Brazil and India indicate that COVID-19 symptoms cannot be terminated similarly as flu. For example, Ester C Sabino et al. [21] observed the resurgence of COVID-19 in January, 2021, Brazil and asserted that one of the main reasons behind was that immunity against COVID-19s infection has already begun to wane by December, 2020. Thus, the recovered group can have a possibility to be infected, or become a virus carrier. On the COVID Tracking Project website, the definition of COVID-19 recovery is set as "ymptom improvement", "Hospital discharges", or even "days since diagnosis" [15]. Yet there is no clear evidence that these recovered patients are immune to COVID-19 afterwards. Thus it is reasonable and necessary to assume that a portion of them, after a certain period, will back to the susceptible groupdue to their immune systems vulnerability status. A recent scientific report from Christian Gaebler et al. [16] also proves that the humoral memory response to COVID-19 will last between 1.3 and 6.3 months after infection without vaccine support. Victor Alexander Okhuese [14] attempted to estimate the probability of COVID-19 reinfections by searching the equilibrium state of 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 took consideration of wrongly executed PCR tests, which is not accurate enough for description of current United States COVID-19 transmission.

In this article, we develop and present a novel COVID-19 transmission model named as Recursive Latency Infection Model (RLIM). The main contributions of this model are:

- Taken secondary infections into consideration for United States recovery group;
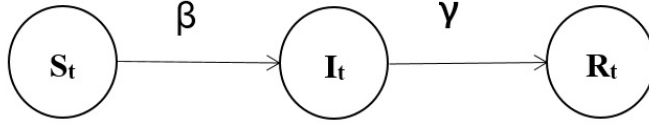
Fig. 1 The SIR model.

– Developed a mathematical model to describe the recover-infection process;
– Provided COVID-19 infected cases and recovered cases prediction with analysis from state level data sources;

The rest of this manuscript is organized as follows. Section 2 discusses necessary mathematical modeling, equations and algorithms implemented. Section 3 describes simulation settings, software and scientific packages utilized by RLIM program. Section 4 takes New Jerseys COVID-19 data and simulation results for discussion, and provides predictions on NJs infections between mid-January and mid-February. Section 5 summarizes the work and brings further discussions.

2 Method

2.1 Recursive Latency Infection Model (RLIM)

In this paper a modified COVID-19 transmission model based on the original SIR model is proposed. As originally proposed by Kermack and McKendrick in 1927 [1]. In this paper, they proposed the Suspectible-Infective-Removed (SIR) model and used their model to explain the 1665-1666s plague at London and the 1906s pestilence at Mumbai, India successfully. The SIR model diagram is shown in Figure 1. And its transmission process is described by equations (1),(2) and (3).

$$\frac{dS_t}{dt} = -\beta \times S_t \times I_t \tag{1}$$

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

$$\frac{dR_t}{dt} = \gamma \times I_t \tag{3}$$

The classical SIR model is only feasible in an ideal epidemic transmission environment, because it takes no consideration on the time-variance of infection rate β and recovery rate γ at all. Also it requires no disease control which means any political or clinical intervention is forbidden. Such transmission behavior rarely appears. However, based on its theoretical assumptions, many revised models, such as SEIR [20], SEIRUS [14], mechanic-statistic SEIR [18], and deep learning SEIR [19], adopted different epidemic transmission characteristics and human interventions, were proposed and developed by researchers. The description of these models can be found in Hethcotes review [2].

The RLIM is inspired by research from Jianping Huangs team, Lanzhou university [3]. Their model, named as GPCP (Global Prediction system for COVID-19 Pandemic), adds four states of disease from SIR model: insusceptible state (P), potentially infected state (E), quarantined state (Q), and mortality state (D). The GPCPs disease transmission process is described by equations (4)-(10).

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

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

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

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

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

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

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

In RLIM, according to our observation and given facts from news reports analysis, we add a symbol ω in the transmission loop. The ω represents a probability of a certain patient, who had recovered from COVID-19 affection for a certain period, which is identified by respiratory tests or antibody tests again as a virus-carrier. Following this definition, ω is used between status R and I, transmitting a certain number of Recovered group into Infected group. Following the assumptions above, the RLIMs equation series shall be modified as equations (11)-(15).

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

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

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

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

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

Compared with GPCP model, RLIM has the following advantages:

(1)Simplified the virus transmission process by removing unnecessary states such as P (immune group) and Q (quarantine group). The reason is because no official data records tracking these two states in our dataset.
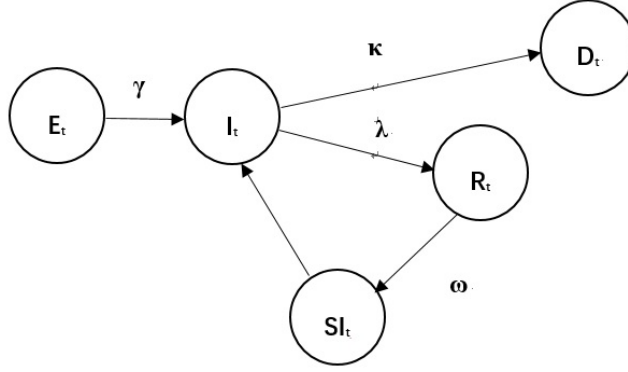
Fig. 2 RLIM model diagram.

(2)Improve the GPCP model with recursive state SI and parameter ω, to avoid the problem of forward transmission only. Without recursive state and parameters existence, number of new infections will decrease no matter what actions to take, and this phenomenon is contradictory to current United States COVID-19 transmission records.

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

In order to apply equations (11) and (12) in our algorithm, a transform into discrete data series shall be implemented as equations (16) and (17). Replace $I(t)$ and $R(t)$ with 4th stage equations (16) and (17) into equation (14) and we have equation (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 equation (16), a relationship between coefficients $[a, b, c, d]$ and $[e, f, g, h]$ is established. Thus, RLIM can predict the infected number of cases given historical number of recovery, previous infections, and assumptions of infection rate, recovery rate and secondary infection period. The description of the corresponding algorithm flow is given in Figure 3, this diagram will be discussed in section 2.2.
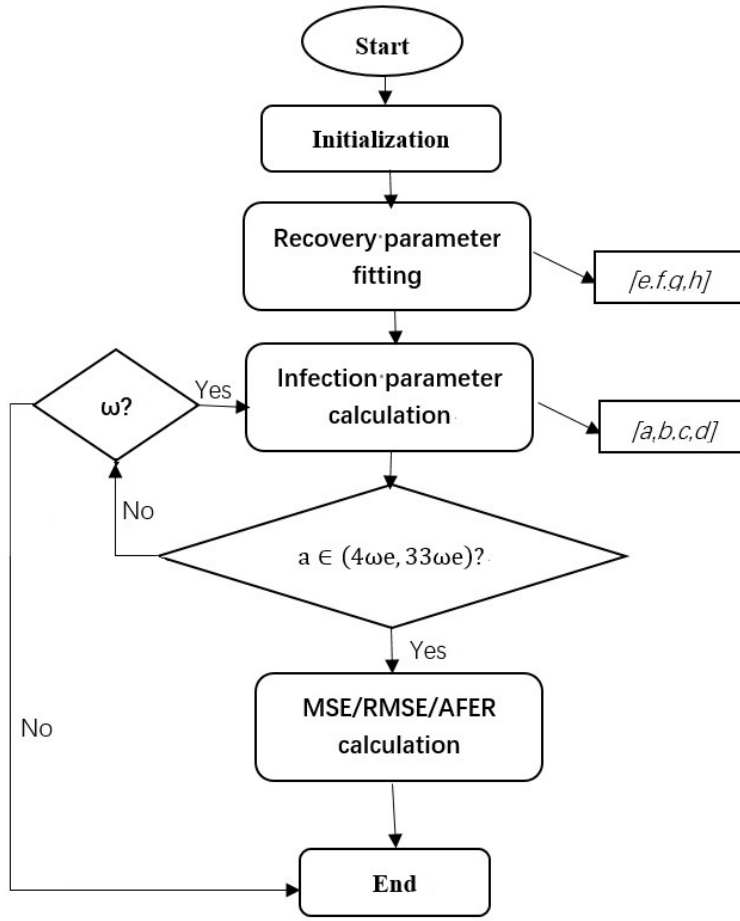
Fig. 3 The RLIM model diagram.

2.2 RLIM: algorithm

2.2.1 Notations

The notations used throughout this article are described in Table 1.

2.2.2 RLIM algorithm

The RLIM algorithm is used to calculate of predicted infected numbers according to equations from (16) to (18) and optimize the difference between actual data recorded by COVID tracking project and the predictive numbers returned from the model. Initially, the predicted recovery numbers $R_1, R_2, \cdots, R_n$ are calculated using the $4th$ R-K order method, based on the real data series of a state from United States, between November 2020 and January 2021. The

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.
T_0	time interval for data fitting in days.
T_1	time interval for prediction in days.
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 4th R-K method on recovered cases.
a, b, c, d	coefficients calculated by equation (16) on infected cases.

coefficients (e, f, g, h) associated with this recovery function will then be transformed into another coefficients (a, b, c, d), with the pre-assigned recovery rate λ (default value 0.01) and secondary infection rate ω (default value 0.01). With coefficients (a, b, c, d) assigned, number of newly infected cases: $I_1, I_2, \cdots, I_n$ within this state can be calculated. Comparing these predictive numbers with actual data, one can evaluate and justify if this round of prediction is accurate or not. Our proposed RLIM model will continuously search for the optimal infection series and then determine the optimistic ω associated with this state.

2.3 Performance measure

RLIMs performance measure is calculated as difference between its predictive output I_p and actual value I_k. Three performance indicators are defined: mean square error (MSE), standard deviation, and average forecasting error rate (AFER) which are shown equations from (19) to (21). Because different states of America have quite different number of infections, ranging from hundreds to thousands, these indicators will be uniformed between 0 to 1 to justify the performance as well.

2.3.1 Mean square error(MSE)

The average of squared difference between RLIMs predictive output I_p and actual value I_k can be calculated in 19.

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

Algorithm 1 RLIM

Input:

(1)$I_1, I_2, \cdots, I_n$: infected case time series of target state;

(2)$R_1, R_2, \cdots, R_n$: recovered case time series of target state;

(3)λ: recovery rate of target state;

(4)ω: initial value of secondary infection rate;

(5)τ: latency period between recovery and secondary infection;

Output:

(1)$I_{1,p}, I_{2,p}, \cdots, I_{n,p}$: infected case time series of target state;

(2)$R_{1,p}, R_{2,p}, \cdots, R_{n,p}$: recovered case time series of target state;

(3)$MSE, RMSE, AFER$;

(4)Optimal value for secondary infection rate: ω;

1: Initial value assignment for T_0 and T_1;
2: RLIM model initialization with proper λ and τ;
3: Secondary infection rate ω set as linear function of time $\omega(t) = (mt + n)/max(\omega(t))$;
4: Use 4th order method to fit the curve of recovered cases of the target state and calculate $[e, f, g, h]$ series;
5: Apply the mathematical equations from 1 to 4 to transform recovered coefficients $[e, f, g, h]$ into infected coefficients $[a, b, c, d]$
6: Use 4th order method to transform $[a, b, c, d]$ into Infected values;
7: 4th order value verification: λ is restricted into range $[0.03, 0.25]$, so 4th order $\frac{\omega e}{\lambda}$ is in range $[4\omega e, 33\omega e]$. If predicted value of a from step 5 is out of this range, then algorithm needs to return to step 3 and choose another ω.
8: Output: $[a, b, c, d, \omega]$ with a in range $[4\omega e, 33\omega e]$.
9: Termination: if all ω from step 3 have been applied and cannot generate a which satisfies the requirement $4\omega e < a < 33\omega e$, then the algorithm will terminate.
10: Calculate the MSE, RMSE and AFER with equation (19), (20), (21) and present output in Table 2 and Table 3.

2.3.2 Root mean square error (RMSE)

The root mean square error is also used to evaluate RLIMs prediction quality. Its formulation is in 20.

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

2.3.3 Average forecasting error rate(AFER)

Average forecasting error rate is the percentage of error, which represents the relative difference between the predictive output I_p and actual value I_t. It is a cumulative statistic deviation between two time-series.

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

3 Experimental setup

3.1 Data source

The data source directly applied in our simulation is from [13]. This dataset contains the United States state-level data of COVID-19, starting from April 2020. In this article, New Jersey (NJ), New York (NY), South Dakota (SD), and New Mexico (NM) are selected because they all have daily tracking recovery reports. RLIM relies heavily on accurate and reliable recovery case reports, and these states own recovery data sources of high credibility.

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: $\lambda = 0.01$;
(4)Secondary infection rate: $\omega = 0.01$;
(5)Latency period: $\tau = 14$;

3.2 Software implementation

The programming language inside RILM is PYTHON with version 3.7, and the essential software package used is SCIPY with version 1.5.4. Two software modules are inherited from SCIPY: integrate and optimize. RILM utilizes the integrate function to calculate the MSE and the optimize function to fit the real recovery data into the 4th order R-K parameters.

3.3 Code availability

RLIM software is publicly available on Github [17], with all codes and implementations available for research. Simulation results are also available via Github.

4 Results and discussion

4.1 Prediction with MSE/RMSE/AFER

Observations from Figure 4, the data fitting and prediction of New York state indicate that RLIM successfully fits current data records from mid-November till mid-January. Different values of the ω, the secondary infection rate, results into different shape of the curve, and the maximum threshold of infections as well. For example, the value of ω equals 0.2, 0.22, 0.25 and 0.3 results in a prediction of newly infected numbers equaling 17,500, 18,000, 20,000 and 25,000. These values also correspond with cumulative infected cases equaling 1.6 million, 1.6 million, 1.8 million and 2 million. To summarize, a higher value of secondary infection rate will result in a faster infection increase pace.
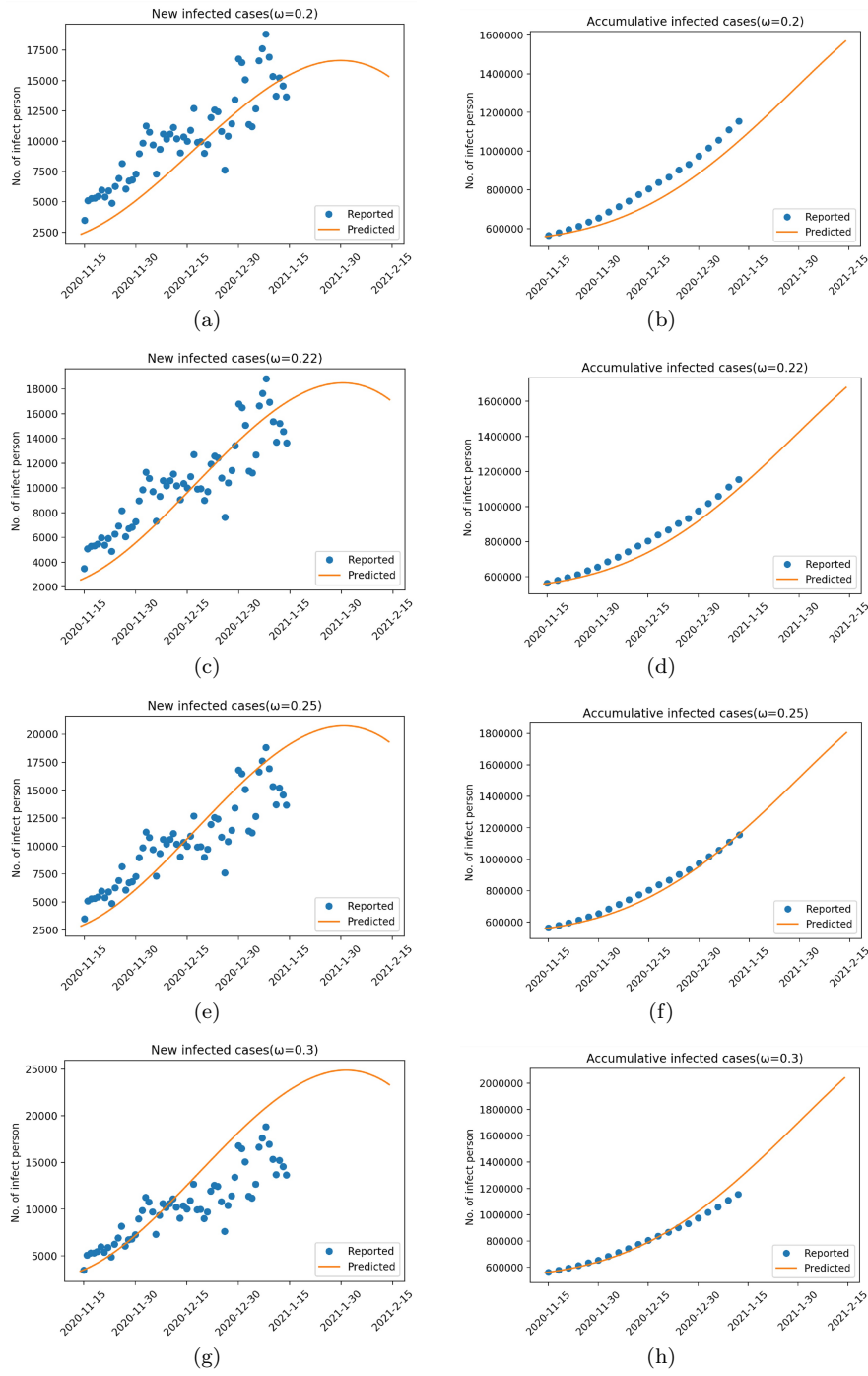
Fig. 4 RILM simulation outputs, New York, United States.

Table 2 RILM simulation report on infections, New York, Nov.2020-Jan.2021

No.	mid-Nov.2020		mid-Dec.2020		mid-Jan. 2021	
	RI	PI	RI	PI	RI	PI
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			

Note: Reported Infection:RI, Predicted Infection: PI.

Another interesting observation is about the turning point, which shows no correspondence with either λ or ω. Simulations with RLIM on different states with recovery rates ranging from 0.01 to 0.5 will not affect the turning point, and we may delve into this topic in section 4.2. In Table 2, the number of New York states new infections caused by COVID-19 is predicted since November 15, 2020. The data records of infections have been collected from the COVID tracking project during November 2020.

The range of infection numbers between November, December and January will be [2600, 9000], [9300, 17000], and [17300, 18500]. Observations from figure 4 indicate that the value $\omega = 0.22$ fits infected cases well. Considering this parameter and our RLIM model, an average of 22 patients will return to infection status from a recovery group of 100 14 days ago. According to the MSE, RMSE and AFER indicator, RLIMs performance reaches 5845242.567

Table 3 RILM simulation report on recoveries, New York, Nov.2020-Jan.2021.

No.	mid-Nov 2020		mid-Dec. 2020		mid-Jan. 2021	
	RI	PI	RI	PI	RI	PI
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			

Note: Reported Infection:RI, Predicted Infection: PI.

in Nov. 2020, with AFER 29.02%. While in December 2020, it obtains a better MSE of 4630069.903 and a lower AFER of 14.97%.

In Table 3, the numbers of New York states new recoveries are also predicted among Nov.2020, Dec.2020 and Jan.2021, with MSE/RMSE/AFER calculated against authentic data records. In Nov. 2020, the mean square error of recovery case prediction reaches 27893.26, with AFER 38.8%. And for predictions in December 2020, the mean square error increases into 43394.77, but with a better AFER of 25.6%.

4.2 The turning point

From RLIMs output on New York states infected numbers and recovery numbers, one can observe that the turning point of this states COVID-19 transmis-
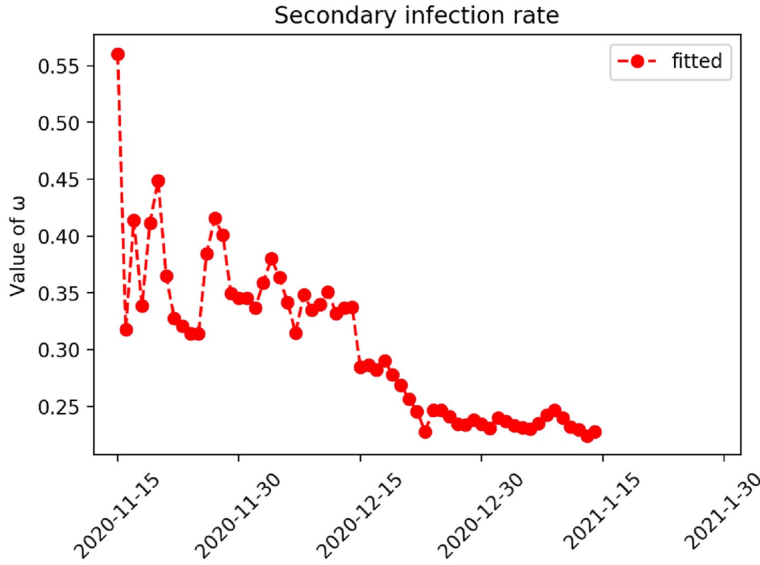
Fig. 5 New York predictions with turning points.

sion is around January 30, 2021. Predictions indicate that from mid-January, New Yorks infections will slowly increase from 17,324 to 18,549, and then fall back to 17,816 at mid-February. The main reason that this turning point appears is the secondary infections ω, whose range is [0.1, 0.55], experiences high fluctuations at early November, 2020, and then drops below 0.4 during December, and after Christmas, 2020, becomes stable around value of 0.25.

Figure 5 illustrates the variance of secondary infection rate ω since mid-November, 2020. These dynamic, inconsistent points are reflections from true COVID-19 data records, calculated from equation 18. In mathematics, RLIM will reach a turning point when recovery rate λ equals secondary infection rate ω. The recovery rate λ for RILM is a constant value of 0.01, thus, given the output from RLIMs prediction of New York state as January 31, 2021, the secondary infection rate ω shall drop to a value of 0.01 and reach an optimal state. Then, RLIMs predictive sequences of new infections and new recoveries will be linear, unless receives real data record from reliable data sources.

4.3 New Jersey, South Dakota and Virginia

Simulation results shown in Figure 6 describe three scenarios: moderate increase (New Jersey) above, moderate decrease (South Dakota) in the middle, and exponential increase (Virginia) below. The optimal secondary infection rate ω for these states are 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 trans-
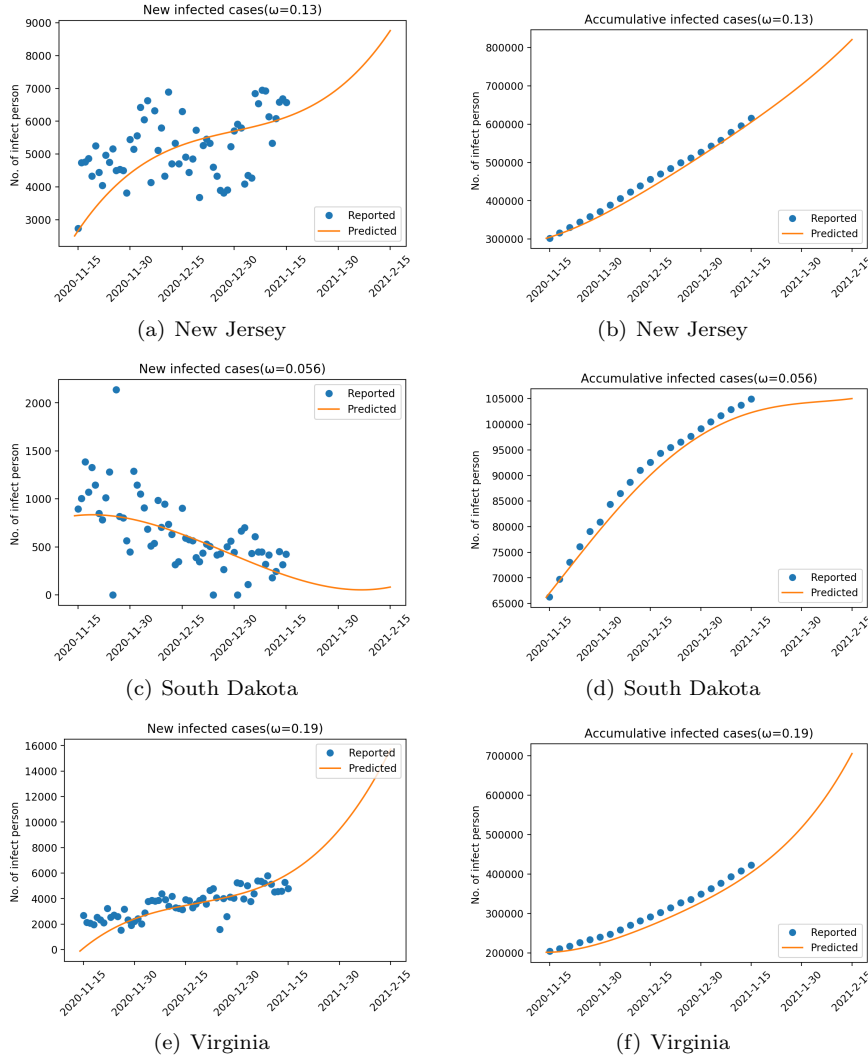
Fig. 6 RLIM simulation outputs, New Jersey (above), South Dakota (middle) and Virginia.

mission trends. Revisiting equations from 11 to 15 from section 2.1 will also explain that in RLIM, ω only affects the incremental steps of infection cases positively and recovery cases negatively. However, it is still valuable in prediction of the turning point, when it approaches the value of recovery rate. Thus, one can conclude that if the recovery rate remains stable during the periods of τ (in RLIM, τ equals 14), then RLIM will approach it during a period and lock down the turning point. The advantage will significantly reduce the time for scientists elaboration on COVID-19s behavior.

5 Conclusion and future work

This research proposes a recursive, latent, dynamic virus transmission model from classical SIR model. This model, named as RLIM, attempts to explain how COVID-19 transmits within different regions of United States. Introducing a new parameter ω into classical SIR model, RLIM is able to predict newly infected cases based on recovered data and historical COVID-19 records. Simulation outputs with RLIM on New York, New Jersey, South Dakota and Virginia proves that given reasonable initial value of ω , this model is able to predict a 30 days infections and recoveries with a reasonable error rate.

As stated in section 1, RLIMs performance relies heavily on accurate recovery reports and statistic data from reliable information sources. Some states of America measured recovery as "length of periods from hospitalization", "days from the first symptom appears", which resulted errand data of recovery. With accurate, precisely defined recovery data, RLIM is able to provide better predictions.

Furthermore, the other field of application is to integrate RLIM with machine learning techniques. The recursive, latent status can be modified into a back propagation process inside a neural network, so RLIM can be equipped with self-learning abilities.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Kenmack WO, McKendrick WG. A contribution to the mathematically theory of epidemics. Proceedings of the Royal Society of London, Series A, 1927, 115:700-721.
2. Hethcote HW. The mathematics of infectious diseases. Siam Review,2000,42(4):599-653.
3. Huang J., X. Liu, L. Zhang, et al. 2020: The amplified second outbreaks of global COVID-19 pandemic. medRxiv. DOI: 10.1101/2020.07.15.20154161.
4. COVID-19 Open Research Dataset Challenge (CORD-19): An AI challenge with AI2, CZ-I, MSR, Georgetown, NIH & The White House. https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/.
5. David   Mezzetti,   CORD-19   Transmission,   Incubation,   Environment, https://www.kaggle.com/davidmezzetti/cord-19-transmission-incubation-environment.
6. Anatoly Zhiljavsky, Ivan Fesenko, Henry Wynn, Roger Whitaker, Kobi Kremnizer, Jack Noonan, Jonathan Gillard, A prototype for decision support tool to help decision-makers with the strategy of handling the COVID-19 UK epidemic, http://doi.org/10.1101/2020.04.24.20077818.
7. Virginia E. Pitzer, Melanie Chitwood, Joshua Havumaki, Nicolas A.Menzies, Stephanie Perniciaro, Joshua L.Warren, Daniel M.Weinberger, Ted Cohen, The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States, http://doi.org/10.1101/2020.04.20.20073338.

8. Easton R. White, Laurent Hebert-Dufresne, State-level variation of initial COVID-19 dynamics in the United States: The role of local government interventions. http://doi.org/10.1101/2020.04.14.20065318.

9. Yang Yu, Yu-Ren Liu, Fan-Ming Luo, Wei-Wei Tu, De-Chuan Zhan, Guo Yu, Zhi-Hua Zhou, COVID-19 Asymptomatic Infection Estimation, https://doi.org/10.1101/2020.04.19.20068072.

10. Chih-Cheng Lai, Yen HungLiu, Cheng-YiWang,Ya-HuiWang,Shun-ChungHsueh,Muh-YenYen,Wen-ChienKo, Po-Ren Hsueh. Asymptomatic carrier state, acute respiratory disease, and pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Facts and myths. https://doi.org/10.1016/j.jmii.2020.02.012.

11. Hongjun Zhu, Yan Li, Xuelian Jin, Jiangping Huang, Xin Liu, Ying Qian, Jindong Tan. Transmission Dynamics and Control Methodology of COVID-19: A Modeling Study. published in Applied Mathematical Modelling doi: 10.1016/j.apm.2020.08.056.

12. Sasmita Poudel Adhikari, Sha Meng,Yu-Ju Wu, Yu-Ping Mao, Rui-Xue Ye, Qing-Zhi Wang, Chang Sun, Sean Sylvia, Scott Rozelle, Hein Raat,Huan Zhou Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. doi: 10.1186/s40249-020-00646-x.

13. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534. doi: 10.1016/S1473-3099(20)30120-1.

14. Victor Alexander Okhuese, Estimation of the probability of reinfection with COVID-19 coronavirus by the SEIRUS model, https://doi.org/10.1101/2020.04.02.20050930.

15. Amanda French, Quang P. Nguyen, Recovery data and the COVID Tracking Project, https://covidtracking.com/analysis-updates/why-its-hard-to-count-recovered/#footnote-3.

16. Gaebler, C. et al. Evolution of antibody immunity to SARS-CoV-2. Nature https://doi.org/10.1038/ s41586-021-03207-w (2021).

17. RILM project, https://github.com/secondary-infection/covid-19.

18. Lionel Roques, Etienne Klein, Julien Papaix, Antoine Sar, Samuel Soubeyrand, Mechanistic-statistical SIR modeling for early estimation of the actual number of cases and mortality rate from COVID-19, https://doi.org/10.1101/2020.03.22.20040915.

19. Hyeontae Jo, Hwijae Son, Hyung Ju Hwang, Se Young Jung, Analysis of COVID-19 spread in South Korea using the SIR model with time-dependent parameters and deep learning, https://doi.org/10.1101/2020.04.13.20063412.

20. Peiliang SUN, Kang Li, An SEIR Model for Assessment of Current COVID-19 Pandemic Situation in the UK, https://doi.org/10.1101/2020.04.12.20062588.

21. Ester C Sabino, Lewis F Buss, Maria P S Carvalho, Carlos A Prete Jr, Myuki A E Crispim, Nelson A Fraiji, Rafael H M Pereira, Kris V Parag, Pedro da Silva Peixoto, Moritz U G Kraemer, Marcio K Oikawa, Tassila Salomon, Zulma M Cucunuba, Mrcia C Castro, Andreza Aruska de Souza Santos, Vtor H Nascimento, Henrique S Pereira, Neil M Ferguson, Oliver G Pybus, Adam Kucharski, Michael P Busch, Christopher Dye, Nuno R Faria, Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. https://doi.org/10.1016/s0140-6736(21)00183-5.