

Mathematical prediction of the spreading rate of COVID-19 using entropy-based thermodynamic model

A Ghanbari¹, R Khordad^{1*} and M Ghaderi-Zefrehei²

¹Department of Physics, College of Science, Yasouj University, Yasouj 75918-74934, Iran

²Department of Animal Genetics, Yasouj University, Yasouj 75918-74934, Iran

Received: 12 July 2020 / Accepted: 16 September 2020 / Published online: 2 January 2021

Abstract: In the COVID-19 pandemic era, undoubtedly mathematical modeling helps epidemiological scientists and authorities to take informing decisions about pandemic planning, wise resource allocation, introducing relevant non-pharmaceutical interventions and implementation of social distancing measures. The current coronavirus disease (COVID-19) emerged in the end of 2019, Wuhan, China, spreads quickly in the world. In this study, an entropy-based thermodynamic model has been used for predicting and spreading the rate of COVID-19. In our model, all the epidemic details were considered into a single time-dependent parameter. The parameter was analytically determined using four constraints, including the existence of an inflexion point and a maximum value. Our model has been layout-based the Shannon entropy and the maximum rate of entropy production of postulated complex system. The results show that our proposed model fits well with the number of confirmed COVID-19 cases in daily basis. Also, as a matter of fact that Shannon entropy is an intersection of information, probability theory, (non)linear dynamical systems and statistical physics, the proposed model in this study can be further calibrated to fit much better on COVID-19 observational data, using the above formalisms.

Keywords: COVID-19; Spreading; Entropy; Thermodynamic

1. Introduction

Beginning from late 2019 (December 2019), an outbreak of coronavirus (COVID-19) infection was detected in Wuhan city China [1]. This is the seventh known coronavirus to infect humans and spreads quickly among people [2]. Some forms of coronavirus including severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS) were appeared and caused death to many people around the world [1, 3, 4].

Coronaviruses composed of large RNA viruses that infect humans and animals. In 1966, Tyrell and Bynoe have described coronaviruses, and they have coined the coronaviruses term (Latin: corona = crown) [5]. Four subfamilies of coronaviruses exist, e.g., alpha, beta, gamma and delta. The first and second types originate from mammals, and the third and fourth originate from pigs and birds. Scientists believe that beta-coronavirus can infect humans and may cause severe disease and fatalities [6].

It is unanimously thought that the COVID-19 has an animal origin [7–11]. Clinical symptoms of the COVID-19 are fever and cough in addition to other non-specific symptomatology including dyspnea, headache, muscle soreness and fatigue [12, 13]. The World Health Organization (WHO) has declared a global health emergency on January 30, 2020 [14]. The COVID-19 human-to-human transmission is so fast. This process is like the growth of a tumor or economic growth of a company. Some crucial computational endeavors have been conducted to decipher transmission dynamics of COVID-19 using various mathematical models [15–25]. In this way, Donnelly et al. [15], Riley et al. [16] and Lipsitch et al. [17] have successfully fitted mathematical models to data for SARS.

Recently, some authors have introduced various models to study the COVID-19 in several countries. For example, Wang [26] has considered applications, limitations and potentials of mathematical models for COVID-19. Yang and Wang [27] have proposed a mathematical model to investigate the current outbreak of the coronavirus disease 2019 (COVID-19) in Wuhan, China. Ivorra et al [28] have developed a mathematical model for the spread of the

*Corresponding author, E-mail: khordad@yu.ac.ir

coronavirus disease 2019 (COVID-19). Zeb et al [29] have used a mathematical model for coronavirus disease 2019 (COVID-19) containing isolation class.

It is now evident that behavior of COVID-19 epidemic depends on medical interventions, public-service announcements, isolation of people, restriction of individual and social activities. Therefore, it would be very difficult to find all those details of the epidemic by a mathematical model. The dynamics of an epidemic is an interesting and important topic in medicine, physics and biology, and it is usually modeled by differential equations. The proposed models are based on a modified susceptible–exposed–infectious–recovered (SEIR) computational framework.

In this work, we want to determine the number of hospitalized cases (cumulative number of cases minus the number of deaths and the number recovered). Thus, a new approach has been considered to predict this number. In this approach, all the mechanisms controlling the spread are factored into a single parameter. Also, we suppose that the system controlling the spread of COVID-19 epidemic is a thermodynamic system and define an entropy and determine the only parameter by the principle of extreme rate of entropy production. We apply a simple differential equation model with the spread rate forced to satisfy four constraints. Assuming the system for spread to be thermodynamically, the model has been closed by the use of maximum or minimal rate of entropy production.

Computer simulations of cellular automata [30, 31] were considered as one of possible methodologies for investigation of epidemiological models [32, 33]. The cellular automata are especially suitable for that purpose because it is usually assumed that agents building the society can be in a discrete number of states. A basic group of epidemiological models originates from the SIR (susceptible–infectious–recovered) model [34] which is, on the one hand, often modified by introduction of additional states of the individual [33, 35]. On the other hand, the interplay between agents and the assumed system topology is another field of possible modification of the considered models [36]. It is frequently assumed that transitions of agents between possible states take place with a constant probability [37, 38]. However, some transitions (like infection) are neighborhood dependent and require that a given agent is in contact with at least one infectious agent [39]. If a susceptible agent is in contact with a larger number of infectious individuals, it can become infectious with a higher probability. This originates not from a larger probability of infection in a single interaction but from a larger number of trials. Accordingly, we can adopt the entropic-based model to fit data from COVID-19 and predict its dynamics.

Entropy is one of the significant quantities in statistical mechanics. This quantity is applicable in studying biological systems [40, 41]. Entropy shows missing information on the state of a system, and it is measure disorder of a system. Here, this quantity has been used to predict the behavior of our system. In this way, in Sect. 2, we have presented the entropic-based mathematical model and defined some parameters. Application, validation and discussion of the model in some countries COVID-19 data are explained in Sect. 3. Finally, the conclusion of our proposed model efficiency and generality is presented in Sect. 4.

2. Mathematical framework

2.1. The entropy-based thermodynamic basic model

First, we consider $f(t)$ function to be the number of confirmed daily cases, as the cumulative cases number. This function is subtracted from the daily deaths and recovery cases, since death and recovery cases are also parts of players in the thermodynamic system. In this research, we have considered the epidemic as a thermodynamic system and followed the approach for the rate of vibration excitation and chemical reaction [42, 43]. Thus, we have supposed that the rate of increase (decrease) would be proportional to the cumulative cases at the previous day:

$$\frac{df(t)}{dt} = \alpha(t)f(t), \quad (1)$$

where $\alpha(t)$ denotes a time-dependent parameter. The determination of this parameter depends on the spreading and controlling of the epidemic. Here, we have not considered the dependence of the parameter on other quantities e.g., population.

To proceeding along, here, we impose four constraints on this parameter:

- (i) The parameter $\alpha(t)$ must have the dimension of t^{-1} . We know that Eq. (1) is the master equation, and $\alpha(t)$ shows the transition probabilities for various pairs of states. The parameter is proportional to inverse time. Therefore, we have selected the simplest choice as $\alpha(t) \sim t^{-1}$. With this selection, the following constraint is satisfied ($t \rightarrow 0, \alpha \rightarrow \infty$).
- (ii) There is an exponential enhancement of smooth spread to start at the initial stage, $\alpha(0) \rightarrow \infty$.
- (iii) On a given day, $t = L$, the rate must decrease.

Mathematically, it means that $\frac{d^2f(t)}{dt^2}$ vanishes at $t = L$; this is called inflexion date. Substituting Eq. (1) into

$$\frac{d^2f(t)}{dt^2} = 0 \text{ yields}$$

$$\frac{d\alpha}{dt} + \alpha^2 = 0, \quad t = L \quad (2)$$

(iv) $f(t)$ has a maximum; this means that for $t = D$, $\alpha(D) = 0$.

According to above imposed constraints, it is supposed that $\alpha(t)$ to be an analytical function of time. The following function would satisfy the four constraints, and it is sufficiently efficient

$$\alpha(t) = -\frac{b}{t} \ln\left(\frac{t}{D}\right) \quad (3)$$

where $b = \frac{1 - \ln(L/D)}{\ln^2(L/D)}$. It should note that above function is not unique, and we can select other functions which satisfy the constraints. To find the function of $f(t)$ analytically, we have chosen $\alpha(t)$ as Eq. (3) that is sufficiently simple. Inserting Eq. (3) into Eq. (1), we obtain the following relation:

$$f(t) = \frac{k}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(\ln(t) - \mu)^2}{2\sigma^2}\right) \quad (4)$$

Here σ and k are proportion constants and $\mu = \ln D + \sigma^2$. It is noteworthy that σ can be computed by of the principle of the extreme rate of entropy production.

2.2. Extreme rate of entropy production

The principle of the extreme rate of entropy production can be employed as a reliable tools to calculate model parameters [44, 45]. Santos et al. [46] have shown that the entropy production of a quantum system undergoing open-system dynamics can be formally split up into a term that only depends on population unbalances. Tsuruyama [47] has considered the locality of the second law of thermodynamics and showed that entropy can be divided into entropy derived from a chemical reaction and entropy produced by the diffusion of signaling molecule. It is to be noted that the width of the curve $f(t) \propto t$ can be characterized by σ . The wider the curve is, the larger is the (Shannon) entropy. The intrinsic spread mechanism of virus and the large mixing activity of the population tend to make the curve wider.

As the maximum dissipation rate is obtained, the width would cease to enhance. Therefore, maximum dissipation rate corresponds to the extreme rate of entropy production. It is related to

$$\frac{d^2 S(\sigma, \eta)}{d\sigma^2} = 0, \quad (5)$$

where $S(\sigma, \eta)$ is the Shannon entropy and it can be written as

$$S(\sigma, \eta) = -\int_0^\infty F(t) \ln F(t) dt^\eta \quad (6)$$

Here, $F(t) = t^{1-\eta} f(t)$ with $\eta = 3$. This value is in the usual entropy definition [44].

The solution of the integral is

$$S(\sigma, \eta) = \eta \left(\ln(\sqrt{2\pi}\sigma) + \eta(\ln D + \sigma^2) + \frac{1}{2} \right) \quad (7)$$

Equation (5) holds if and only if

$$\sigma = \frac{1}{\sqrt{2\eta}} \approx 0.0408, \quad \text{for } \eta = 3 \quad (8)$$

With the value of σ , the intrinsic spread mechanism is balanced by the dissipation mechanism. To obtain the relationship between two typical dates D (maximum number of confirmed daily cases) and L (inflexion date), we should employ the definition of inflexion point. The inflexion point is the date at which the multiple controlling measure takes effect.

We inserted Eq. (4) into $\left. \frac{d^2 f(t)}{dt^2} \right|_{t=L} = 0$; then we have

$$D = L \exp\left(\frac{1}{2}\sigma^2 + \frac{1}{2}\sqrt{4\sigma^2 + \sigma^4}\right) \quad (9)$$

Applying Eq. (4), one can determine a relation between $f(D)$ and $f(L)$ as

$$f(D) = F(L) \exp\left(-\sigma^2 - \frac{1}{2}\sqrt{4\sigma^2 + \sigma^4} + \frac{1}{2}\left(\frac{3}{2}\sigma + \frac{1}{2}\sqrt{4 + \sigma^2}\right)^2\right) \quad (10)$$

According to Eq. (8), the following important relations can be grasped

$$\left. \frac{D}{L} \right|_{\eta=3} = 1.649, \quad \left. \frac{f(D)}{f(L)} \right|_{\eta=3} = 2.120 \quad (11)$$

By applying the above relations, we can predict the maximal number of confirmed daily cases and the day in which this maximum appears. The maximum number of infected individuals and the time at which this maximum occurs can be related to the number and time corresponding to the critical date.

First, we compute the following relation

$$\left. \frac{df(t)}{dt} \right|_{t=L} = -\frac{1}{\sigma^2 L} f(L) \ln \frac{L}{D} \quad (12)$$

which yields

$$L = \left(\frac{1}{2} + \frac{1}{2} \sqrt{\frac{4}{\sigma^2} + 1} \right) \frac{f(L)}{\left. \frac{df(t)}{dt} \right|_{t=L}} = \frac{3f(L)}{\left. \frac{df(t)}{dt} \right|_{t=L}} \quad (13)$$

We want to know the initial date for which the spreading

Table 1 Confirmed daily cases and daily deaths of COVID-19 Worldwide, China, and Iran from January 23, 2020, to April 5, 2020 (Data are given from <https://www.worldometers.info/coronavirus/>)

The date	Confirmed daily cases			Daily deaths		
	Worldwide	China	Iran	Worldwide	China	Iran
Jan 23	398	259	–	–	8	–
Jan 24	472	457	–	–	16	–
Jan 25	698	688	–	–	15	–
Jan 26	785	769	–	24	24	–
Jan 27	1781	1771	–	26	26	–
Jan 28	1477	1459	–	26	26	–
Jan 29	1755	1737	–	38	38	–
Jan 30	2010	1981	–	43	43	–
Jan 31	2127	2099	–	46	46	–
Feb 1	2603	2589	–	45	45	–
Feb 2	2833	2825	–	55	57	–
Feb 3	3239	3235	–	65	64	–
Feb 4	3915	3884	–	66	65	–
Feb 5	3721	3694	–	73	73	–
Feb 6	3173	3143	–	73	73	–
Feb 7	3437	3385	–	86	86	–
Feb 8	2676	2652	–	89	89	–
Feb 9	3001	2973	–	97	97	–
Feb 10	2546	2467	–	108	108	–
Feb 11	2035	2015	–	97	97	–
Feb 12	4153	4108	–	146	146	–
Feb 13	5151	5090	–	122	121	–
Feb 14	2662	2641	–	143	143	–
Feb 15	2097	2008	–	143	142	–
Feb 16	2132	2048	–	106	105	–
Feb 17	2003	1888	–	98	98	–
Feb 18	1852	1749	–	136	136	–
Feb 19	516	391	–	117	114	–
Feb 20	977	889	–	121	118	–
Feb 21	996	823	13	131	109	2
Feb 22	978	648	11	100	97	2
Feb 23	554	214	14	158	150	2
Feb 24	882	508	18	81	71	4
Feb 25	741	406	34	64	52	4
Feb 26	992	433	44	37	29	3
Feb 27	1292	327	106	57	44	7
Feb 28	1503	427	143	65	47	8
Feb 29	1989	573	205	54	35	9
Mar 1	1981	202	385	73	42	11
Mar 2	1858	125	523	67	31	12
Mar 3	2573	119	835	85	38	11
Mar 4	2298	139	586	83	31	15
Mar 5	3111	143	591	102	30	16
Mar 6	3625	99	1234	107	28	16
Mar 7	4049	44	1076	105	27	21
Mar 8	3892	40	743	228	22	49

Table 1 continued

The date	Confirmed daily cases			Daily deaths		
	Worldwide	China	Iran	Worldwide	China	Iran
Mar 9	4390	–	595	198	17	43
Mar 10	4567	–	881	271	22	54
Mar 11	7266	–	958	332	11	63
Mar 12	8295	–	1075	353	7	75
Mar 13	10907	–	1289	447	13	85
Mar 14	11059	–	1365	405	10	97
Mar 15	13042	–	1209	687	14	113
Mar 16	12897	–	1053	642	13	129
Mar 17	15745	–	1178	817	11	135
Mar 18	20585	–	1192	972	8	147
Mar 19	26158	39	1046	1079	3	149
Mar 20	30648	41	1237	1356	7	149
Mar 21	29429	46	966	1625	6	123
Mar 22	32480	39	1028	1629	9	129
Mar 23	41371	78	1411	1873	7	127
Mar 24	43744	47	1762	2381	4	122
Mar 25	48461	67	2206	2388	6	143
Mar 26	60830	55	2389	2791	5	157
Mar 27	64501	54	2926	3270	3	144
Mar 28	66761	45	3076	3518	5	139
Mar 29	60263	52	2901	3204	5	123
Mar 30	61351	79	3186	3709	7	117
Mar 31	73620	36	3110	4535	6	141
Apr 1	76871	–	2988	4889	4	138
Apr 2	79864	–	2875	5979	4	124
Apr 3	101566	–	2715	5714	3	134
Apr 4	84821	–	2560	5800	4	158

of the epidemic triggers. For this purpose, we must determine the value L . This value has been obtained in Eq. (13).

3. Validation and discussion of the proposed model

Before using the model on real data, we must determine some model parameters. Firstly, we characterized the number $F = f(L)$ at the inflexion date (the date that $\frac{df(t)}{dt}$ tends to reduce) and $\left. \frac{df(t)}{dt} \right|_{t=L}$ using the reported daily cases data. The constant k has been obtained by $f(L) = F$ in Eq. (4). Finally, we used Eq. (13) to find L . Afterward, using Eq. (11), we predicted D and $f(D)$ and plot the curve $f(t) \sim t$ to predicted the number $f(t)$ for $L < t$.

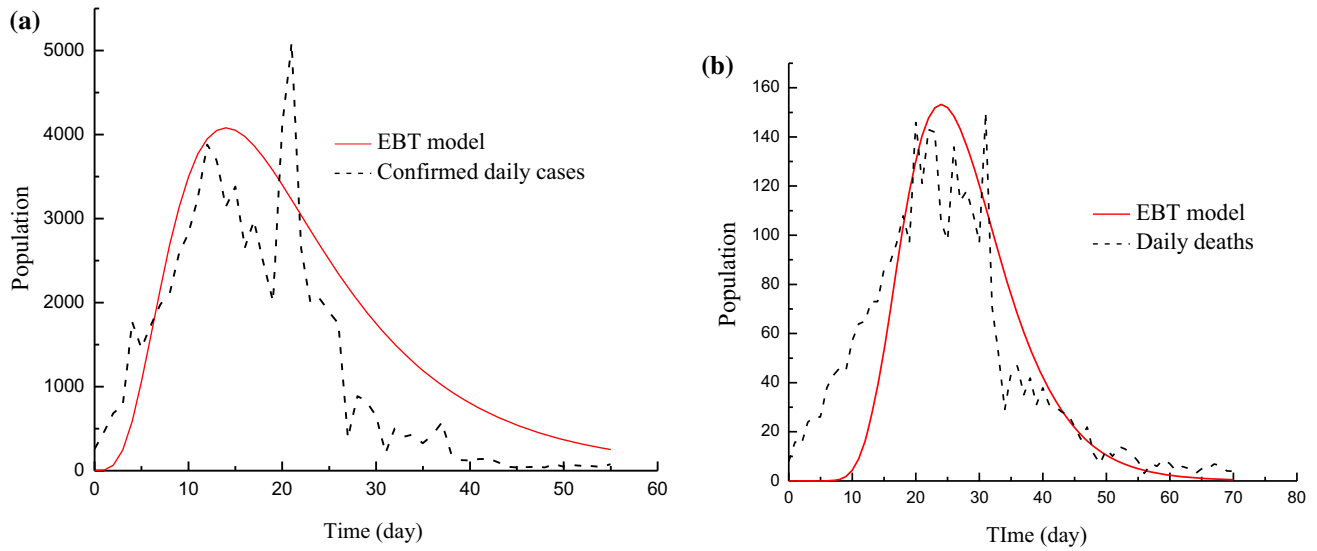


Fig. 1 Comparison of the results of China COVID-19 with entropy-based thermodynamic model: (a) confirmed daily cases and (b) daily deaths

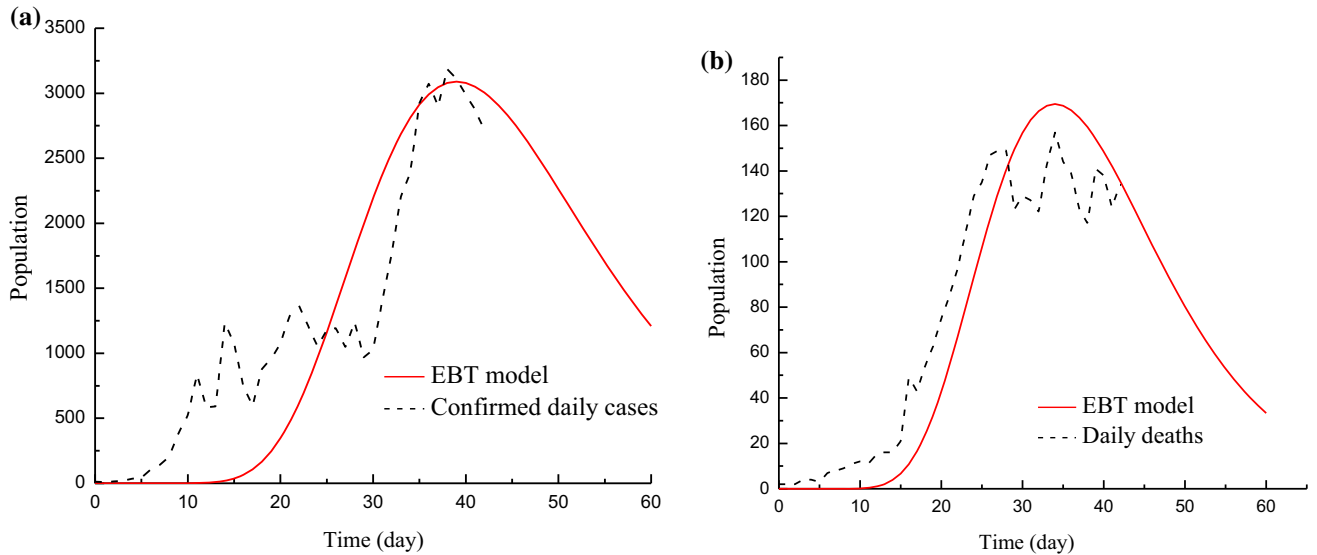


Fig. 2 Comparison of the results of Iran COVID-19 with entropy-based thermodynamic model: (a) confirmed daily cases and (b) daily deaths

Then, we find the inflexion point as follows. We record the reported the number $f(t)$ for each day and plotted the curve $g(t) = f(t) - f(t-1)$. When we observed that $g(t)$ reaches a peak (denoted as G) at $t = L$, then L is considered as the inflexion point.

Also L is calculated numerically as

$$L = \frac{3\bar{F}}{G} \quad (14)$$

Here, $\bar{F} = (2F - G)/2$ is the number of f averaged daily cases over two consecutive dates.

After determining the model parameters, we test the model for China and Iran confirmed COVID-19 data, since as it can be seen, only these two countries have a peak in

data reported. Table 1 shows the reported data from January 23, 2020, to April 5, 2020. In Fig. 1(a) and (b), we have compared our analytical model with confirmed daily cases and daily deaths of China, respectively. As can be seen, there is fairly good agreement between our model and reported data. At initial parts of the curve, our model has a good agreement with reported data. Then, nearly peaks of curves this agreement gets biased as time passes, and this agreement gets disappeared. Figure 2(a) and (b) shows comparison of the model with confirmed daily cases and daily deaths of Iran, respectively. The figures show there is a rather good agreement between data and proposed model. In this case, we don't have enough data so we just can predict the future of cases in Iran. Figure 3(a) and

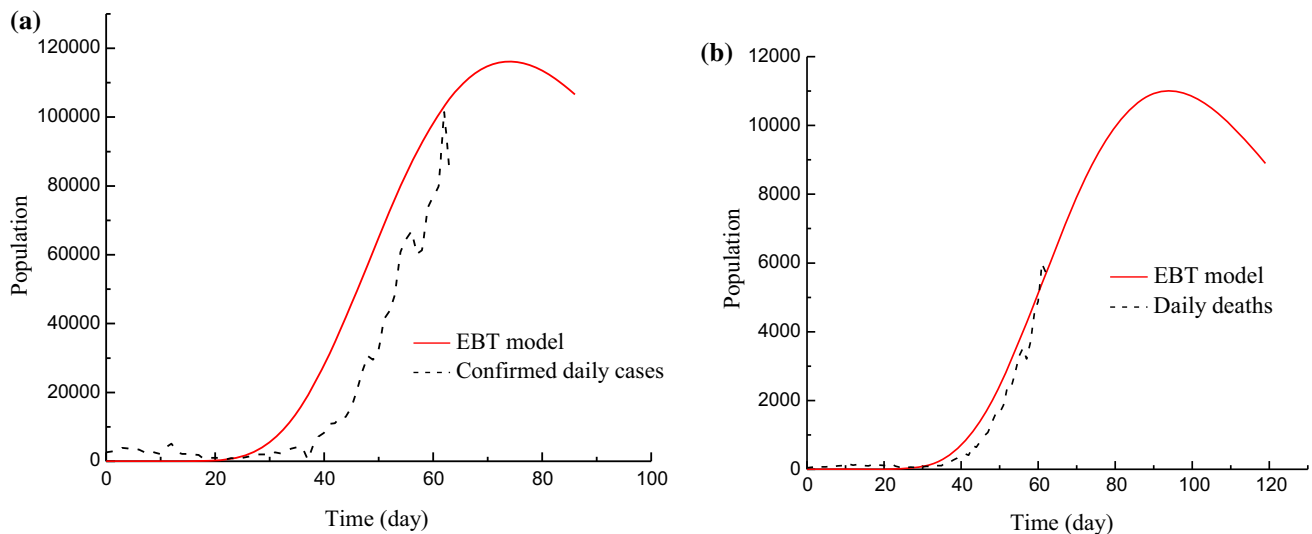


Fig. 3 Comparison of the results of worldwide COVID-19 with entropy-based thermodynamic model: (a) confirmed daily cases and (b) daily deaths

(b) shows confirmed daily cases and daily deaths of worldwide. In this case, no sign of curve peak could be observed in reported data. According to the model, we can predict the behavior of worldwide diagram and find behavior dynamics for it. Finally, we can predict the behavior of cases for other countries and regions. The proposed mathematical perspective gives a central position to entropy and relative entropy in tracking down the dynamic behavior of COVID-19 and provides insights into the notions of randomness, typicality and disorder in COVID-19. However, relevance of in life science and COVID-19 modeling is yet to be demonstrated and worked out.

4. Conclusions

In this work, we have studied the entropic model of coronavirus (COVID-19). To reach the goal, we have used the entropy-based thermodynamic principle and have introduced a mathematical model that almost fitted well to the reported data. Accordingly, the proposed model can be used to predict the future of COVID-19 propagation. It should note that the simplicity of the model is an advantage point. The details of epidemic have been ignored in our model. We have used the model for the first peak in the infection. More comparison needs to be done to further assess the validity. This can be considered in future studies.

It is needless to say that many factors play pivotal roles in spreading and propagating of COVID-19 in the community. The proposed model layouts a template for those factor to be well calibrated in it, given accessibility to

enough amount of data and having semantic conceptual relations among them.

References

- [1] W H Organization WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China (Beijing: WHO) **9** 12 (2020)
- [2] N Zhu, et al. *New England Journal of Medicine* **385** 543 (2020)
- [3] C W Lam, M H Chan and C K Wong *Clinical Biochemist Reviews* **25** 121 (2004)
- [4] E I Azhar, et al. *Infectious Disease Clinics* **33** 891 (2019)
- [5] D Tyrrell and M Bynoe *Lancet* **12** 76 (1966)
- [6] W Tan, et al. *China CDC Weekly* **2** 61 (2020)
- [7] J F W Chan, et al. *Lancet* **395** 514 (2020)
- [8] C Huang, et al. *Lancet* **395** 497 (2020)
- [9] L T Phan, et al. *New England Journal of Medicine* **382** 872 (2020)
- [10] Q Li, et al. *New England Journal of Medicine* **382** 123 (2020)
- [11] S Chatterjee, et al. *Indian. J. Phys.* (2020) <https://doi.org/10.1007/s12648-020-01766-8>
- [12] C P E R E Novel *Zhonghua liuxingbingxue zazhi* **41** 145 (2020)
- [13] C Wang, et al. *Lancet* **395** 470 (2020)
- [14] E Mahase *Bmj* **368** m408 (2020)
- [15] C A Donnelly, et al. *Lancet* **361** 1761 (2003)
- [16] S Riley, et al. *Science* **300** 1961 (2003)
- [17] M Lipsitch, et al. *Science* **300** 1966 (2003)
- [18] C Dye and N Gay *Science* **300** 1884 (2003)
- [19] N Jia and L Tsui *North American Actuarial Journal* **9** 28 (2005)
- [20] D J Watts, et al. *Proceedings of the National Academy of Sciences* **102** 11157 (2005)
- [21] E Kenah and J M. Robins *Physical Review E* **76** 036113 (2007)
- [22] L Bian and D Liebner *Transactions in GIS* **11** 155 (2007)
- [23] X Yu, et al. *Physica A* **387** 1421 (2008)
- [24] B Dybiec *The European Physical Journal B* **67** 377 (2009)
- [25] B Dybiec, A Kleczkowski and C A Gilligan *Journal of the Royal Society Interface* **6** 941 (2009)
- [26] J Wang *J. Public. Health. Emerg.* **4** 9 (2020)
- [27] C Yang and J Wang *Math. Biosci. Engin.* **17** 2708 (2020)

- [28] B Ivorra et al *Commun. Nonlinear Sci. Numer. Simulat.* **88** 105303 (2020)
- [29] A Zeb et al *Biomed Res. Int.* **7** 3452402 (2020)
- [30] S Wolfram *Rev. Mod. Phys.* **55** 601 (1983)
- [31] S Wolfram *A New Kind of Science* (Champaign: Wolfram Media) (2002)
- [32] C Rhodes and R M Anderson *Theor. Popul. Biol.* **52** 101 (1997)
- [33] B Dybiec, A Kleczkowski and C Gilligan *Phys. Rev. E* **70** 066145 (2004)
- [34] R M Anderson, B Anderson and R M May *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press) 1992.
- [35] E Ahmed and H Agiza *Physica A* **253** 347 (1998)
- [36] N Boccara and K Cheong *J. Phys. A* **26** 3707 (1993)
- [37] B Dybiec, A Kleczkowski and C Gilligan *Acta Phys. Polo. B* **36** 12 (2005)
- [38] A Kleczkowski, B Dybiec and C Gilligan, arXiv preprint physics/0608141 (2006)
- [39] P Grassberger *Math. Biosci.* **63** 157 (1983)
- [40] R Khordad and H R Sedehi *The European Physical Journal Plus* **130** 246 (2015)
- [41] H Husi *Feature Selection in Microarray Data Using Entropy Information—Computational Biology* (2019)
- [42] Z N Wu *Probabil. Eng. Mech.* **18** 241 (2003)
- [43] J D Anderson Jr *Hypersonic and high Temperature Gas Dynamics*, 2nd edn. (Reston: American Institute of Aeronautics and Astronautics, Inc.) (2006)
- [44] H Ziegler *An Introduction to Thermomechanics* (Amsterdam: NorthHolland Publisher) (1983)
- [45] W B Wang Z N Wu, C F Wang and R F Hu *Sci. Chin. Phys. Mech. Astron.* **56** 2143 (2013)
- [46] J P Santos, et al. *npj Quant. Inform.* **5** 1 (2019)
- [47] T Tsuruyama *J. Theor. Biol.* **472** 84 (2019)

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