

Estimating the size of COVID-19 epidemic outbreak

Chakrit Pongkitivanichkul,¹ Daris Samart,¹ Takol Tangphati,² Phanit Koomhin,^{3,4} Pimchanok Pimton,^{4,5,6} Punsiri Dam-O,^{4,5} Apirak Payaka,^{4,5,7} and Phongpichit Channuie^{4,5,7}

¹Department of Physics, Faculty of Science, Khon Kaen University, Khon Kaen, 40002, Thailand

²Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

³School of Medicine, Walailak University, Nakhon Si Thammarat, 80160

⁴Research Group in Applied, Computational and Theoretical Science (ACTS),
Walailak University, Nakhon Si Thammarat, 80160, Thailand

⁵School of Science, Walailak University, Nakhon Si Thammarat, 80160, Thailand

⁶Plasmas and Electromagnetic Wave Science (PEwave) Center of Excellence,
Walailak University, Nakhon Si Thammarat, 80160, Thailand

⁷College of Graduate Studies, Walailak University, Nakhon Si Thammarat, 80160, Thailand

(Dated: March 26, 2020)

Abstract

In this work, we analyze the epidemic data of cumulative infected cases collected from many countries as reported by WHO starting from January 21st 2020 and up till March 21st 2020. Our inspection is motivated by the renormalization group (RG) framework. Here we propose the RG-inspired logistic function of the form $\alpha_E(t) = a(1 + e^{-c(t-t_0)})^{-n}$ as an epidemic strength function with n being asymmetry in the modified logistic function. We perform the non-linear least-squares analysis with data from various countries. The uncertainty for model parameters is computed using the squared root of the corresponding diagonal components of the covariance matrix. We carefully divide countries under consideration into 2 categories based on the estimation of the inflection point: the maturing phase and the growth-dominated phase. We observe that long-term estimations of cumulative infected cases of countries in the maturing phase for both $n = 1$ and $n \neq 1$ are close to each other. We find from the value of root mean squared error (RMSE) that the RG-inspired logistic model with $n \neq 1$ is slightly preferable in this category. We also argue that n determines the characteristic of the epidemic at an early stage. However, in the second category, the estimated asymptotic number of cumulative infected cases contain rather large uncertainty. Therefore, in the growth-dominated phase, we focus on using $n = 1$ for countries in this phase. Some of them are in an early stage of an epidemic with an insufficient amount of data leading to a large uncertainty on parameter fits. In terms of the accuracy of the size estimation, the results do strongly depend on limitations on data collection and the epidemic phase for each country.

INTRODUCTION & MOTIVATION

The coronavirus disease 2019 (COVID-19) caused by a novel coronavirus SARS-CoV-2 that emerged in the city of Wuhan, China, last year and has since expanded to a large scale COVID-19 epidemic and spread to all regions as reported by the World Health Organization (WHO) [1] causes a serious situation worldwide. Since the first reports of COVID-19, there are various attempts to estimate the final size of the spread mathematically. Among several approaches [2–9], the logistic growth models are very popular and useful to predict the final size of the epidemic [10, 11]. It is worth noting that the logistic function or logistic curve is a common “S” shape (Sigmoid curve). Logistic functions are often used in statistics and machine learning, medicine, chemistry, physics, material science, linguistics, and even agriculture. However its root somehow seems to be coincidentally originated from the high energy physics and condensed matter physics’ point of view. In terms of the renormalisation group (RG) framework, we define the logistic growth function as an epidemic strength function allowing us to describe the spread of disease. Its derivative with respect to time

yields a new quantity which can be interpreted as the beta function of an underlying microscopic model. This frame work was first proposed in Ref.[12] to describe the underlying dynamics of disease spread by invoking the usual logistic-growth model.

According to the reported data, a number of the COVID-19 infected cases grow up slowly at a very early stage of the epidemic spread. The next stage of growth is approximately exponential, then as saturation begins, the growth slows down to linear, and at maturity, the growth stops. Similar behaviour happens in various phenomena in physics, for example, the running coupling constants of particular theories describing phase transitions in high energy physics and in condensed matter physics. The renormalization group (RG) method is a successful framework, and is widely used to describe the dynamics of mentioned phenomena [13, 14]. We will use, in this work, the generalized characteristic function of the epidemic strength of the form

$$\alpha_E(t) = a(1 + e^{-c(t-t_0)})^{-n}, \quad (1)$$

where n is an arbitrary constant, a is the curve’s maximum value, c is the logistic growth rate or steepness of

the curve, and t_0 can be thought of as a starting time at which $\alpha_E(t_0) = a/2^n$. It is worth noting that when $n = 1$ the strength function given in Eq.(1) is just a usual logistic one used in, for example, Refs.[10–12]. Here in the present work, we generalize such a function by considering its power-law form given in Eq.(1) and estimate the size of COVID-19 epidemic outbreak in terms of the state-of-the-art RG framework.

RENORMALIZATION GROUP RECIPE

In this work, the epidemic strength function α_E in Eq.(1) will be identified as the number of the infected cases. This function is also known as the generalized logistic function. Moreover, the function α_E has many interesting features and it can be reduced to several useful function in physics such as Fermi-Dirac distributions [15] or running coupling constant of the four-fermion contact interaction [16] for $n = 1$ and the generalized Woods-Saxson nuclear potential [17] for $n = 2$ as well as the Starobinsky inflation potential [18] for $n = -2$. Turning to the crucial part of this section, we will derive the beta function of epidemic strength coupling with the RG technique in the following and demonstrate how COVID-19 spread underlying the RG flow description. It is well known that the evolution of the $\alpha_E(t)$ function is governed by the beta function. In the present work, we use t as a time of the evolution of the system instead of energy in high energy physics. The time scale in our work is related to the typical energy scale as $t = \ln \mu_0/\mu$ in the standard RG approach where μ_0 and μ are the energy scale limit of our interested and the running energy scale, respectively. The standard beta function of α_E is defined by

$$\begin{aligned} \beta_E &\equiv \frac{d\alpha_E}{d \ln(\mu/\mu_0)} \\ &= -\frac{d\alpha_E}{dt} = cn \left(\alpha_E - \frac{\alpha_E^{1+1/n}}{a^{1/n}} \right), \end{aligned} \quad (2)$$

where a , c and n are arbitrary constants. An evolution (or RG flow) of the epidemic strength is controlled by the beta function in Eq.(2). The fixed points of the beta function are obtained by solving $\beta_E = 0$ which yield two fixed points in the RG equation as

$$\alpha_{E,1}^* = 0, \quad \alpha_{E,2}^* = a. \quad (3)$$

In order to explore the critical phenomena in the RG analysis, one needs to identify the stability and classify asymptotic properties of the fixed points from the the beta function. To do these, the slope of the beta function is very useful and convenient to study the stability for each fixed points. It reads,

$$\theta_E^* \equiv \left. \frac{d\beta_E}{d\alpha_E} \right|_{\alpha_E^*} = cn \left(1 - \left[1 + \frac{1}{n} \right] \frac{\alpha_E^{1/n}}{a^{1/n}} \right) \Big|_{\alpha_E^*}. \quad (4)$$

It is worth noting that the relative signs of the scaling exponent tell us about the fixed point stability. The plus sign reflects an unstable node while the negative value of the θ_E^* gives an attractor fixed point. Having use Eq.(4) for $\alpha_{E,1}^*$ and $\alpha_{E,2}^*$, we obtain,

$$\theta_{E,1}^* = \theta_E^* \Big|_{\alpha_{E,1}^*} = cn, \quad \theta_{E,2}^* = \theta_E^* \Big|_{\alpha_{E,2}^*} = -c, \quad (5)$$

where the parameters c and n are positive constants and well defined for fitting procedure in the latter. On one hand, according to results in Eq.(5), the fixed point $\alpha_{E,1}^*$ is the unstable node where this solution corresponds to the source or the origin of the dynamical system. On the other hand, the $\alpha_{E,2}^*$ is the attractor fixed point which means the solution of the system asymptotically converging to this point. One may interpret the $\alpha_{E,1}^*$ as the zero infected cases or starting point of the disease spreading while the maximum number of the infected cases and staying at this point under infinite time translation are represented by the $\alpha_{E,2}^*$ solution. As the results, we see that the spread of the COVID-19 can be represented by the running epidemic strength effect in the RG flow framework. Interestingly, moreover, the microscopic theoretical method in physics is suitable and useful to study the macroscopic phenomena such as the COVID-19 spread in the present study. Then, we would call the epidemic strength function in Eq.(1) in this work as the RG-inspired model.

FITTING RESULTS & INTERPRETATIONS

In the present analysis, we have collected data of the number of cumulative infections reported by WHO starting from January 21st 2020 and up till March 21st 2020, see [1]. The data are counted using the first day of an epidemic as day one such that the comparison between countries are feasible. We perform the non-linear least squares to fit with the 4-parameters model (a, t_0, c, n) from Eq.(1). We are also interested in the case $n = 1$ where our results becomes the general model of population growth (S-curve). The uncertainty for model parameters is estimated using the squared root of the corresponding diagonal components of the covariance matrix. We carefully divide countries under consideration into 2 categories based on the estimation of the inflection point, t_0 , which is the point where the growth rate begins to decline, see FIG.1. We label any country where the current situation has already passed an estimated inflection point of the S-curve as currently in the *maturing phase*. A country in this phase will provide a good fit to all model parameters in our model. Whereas any country where the current situation has not reached an estimated inflection point yet is labelled as currently in the *growth-dominated phase*. Generally, a model fit of a country in this phase yields no reliable result. Only exponential function seems to fit a country in this phase.

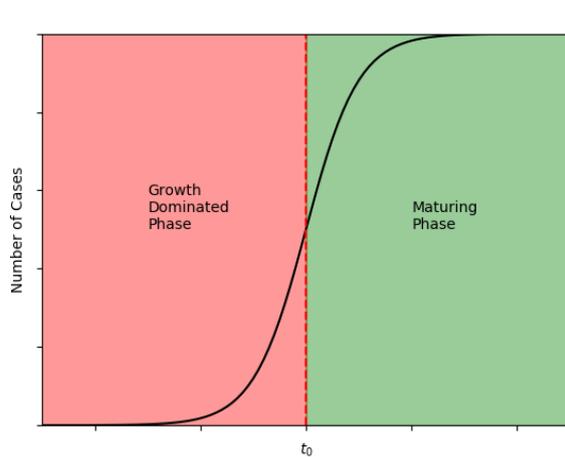


FIG. 1: The classification of countries based on the inflection point. t_0 is the midpoint of the logistic function ($\alpha_E(t_0) = a/2$) and the turning point where the growth rate begins to slow down.

Maturing Phase

A clear example of countries beyond the inflection point is *China* where the outbreak was first reported. Another country where the outbreak has passed beyond the inflection point and the situation is now under control is *Korea*. The results of parameter fit for China and Korea are shown in Fig.2. In the China and Korea cases, The reported data show that the spread have reached the end point of the epidemic. This is the reason in our RG-inspired logistic growth model when both $n = 1$ and $n \neq 1$ give very similar model parameters and very close prediction to each other. The value of root mean squared error (RMSE) from both models suggests that the RG-inspired logistic model with $n \neq 1$ is slightly preferable.

Another set of countries where the situation has already passed the inflection point but has not reached the plateau yet are *Austria*, *Belgium*, *Japan*, and *Norway*. The results of fitting are shown in Fig.3 for Austria and Japan.

Compared with $n = 1$, the RG-inspired logistic model with $n \neq 1$ is again a slightly better fit to data, and the prediction of the size of the outbreak is lower than that of $n = 1$. The values of the model parameters with estimated errors are given in TABLE I. The values of n in RG-inspired logistic model represent asymmetry in the logistic function. When $n < 1$, initial phase of the epidemic is slower than that observed in the normal logistic function ($n = 1$) see Fig.4.

China, Japan, Austria, Belgium and Norway are fitted into this group. These countries have evidently chosen a relatively strong measure in containing and mitigating the spread of the virus since an early stage. We argue that this value is a reliable indication of an early stage effort on controlling the epidemic. Korea, on the other

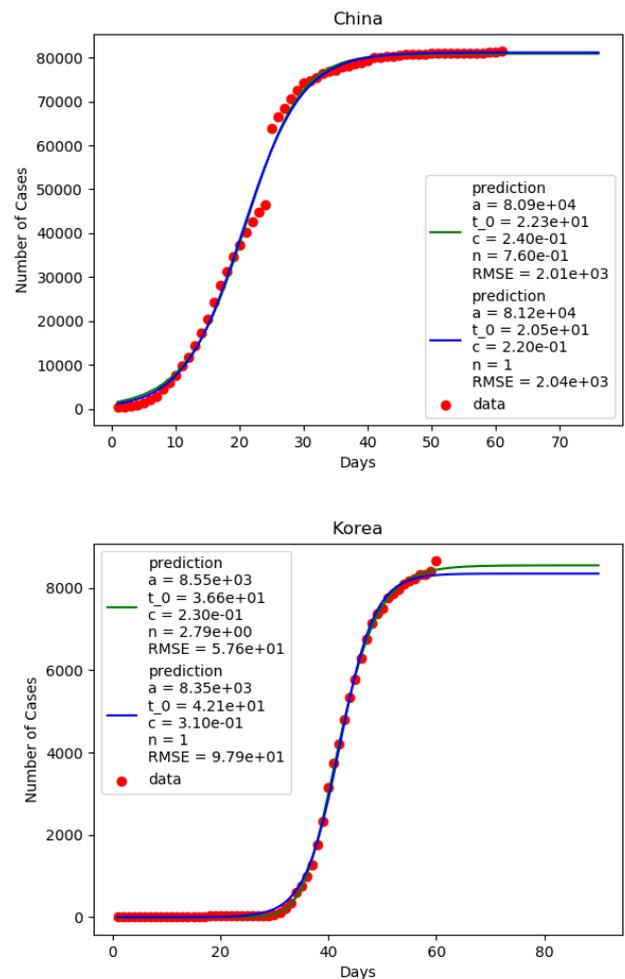


FIG. 2: The parameter fit on data from China and Korea. The green line and blue line are the best fit from RG-inspired logistic model and logistic model respectively. Both models are very close in predicting the size of the epidemic.

Country	a	t_0 (days)	c (day^{-1})	n
China	80914 ± 448	22.3 ± 1.1	0.24 ± 0.02	0.76 ± 0.13
Korea	8548 ± 35	36.6 ± 0.9	0.23 ± 0.01	2.79 ± 0.43
Austria	2179 ± 192	21.9 ± 0.4	0.55 ± 0.15	0.62 ± 0.22
Belgium	2591 ± 806	43.5 ± 1.0	0.36 ± 0.20	0.83 ± 0.61
Japan	988 ± 37	56.0 ± 0.3	0.32 ± 0.05	0.31 ± 0.05
Norway	1725 ± 124	18.5 ± 1.0	0.41 ± 0.10	0.84 ± 0.33

TABLE I: The table shows the best-fitted values of parameters (a , t_0 , c , n) for each country categorized in the maturing phase. Here we observed that $n < 1$ in all cases implying that initial phase of the epidemic is slower than that of the $n = 1$ case, a normal logistic function.

hand, has the value of $n = 2.79 \pm 0.43$ suggesting that the spread of the virus in an early stage was much quicker than the normal logistic function ($n = 1$). This coin-

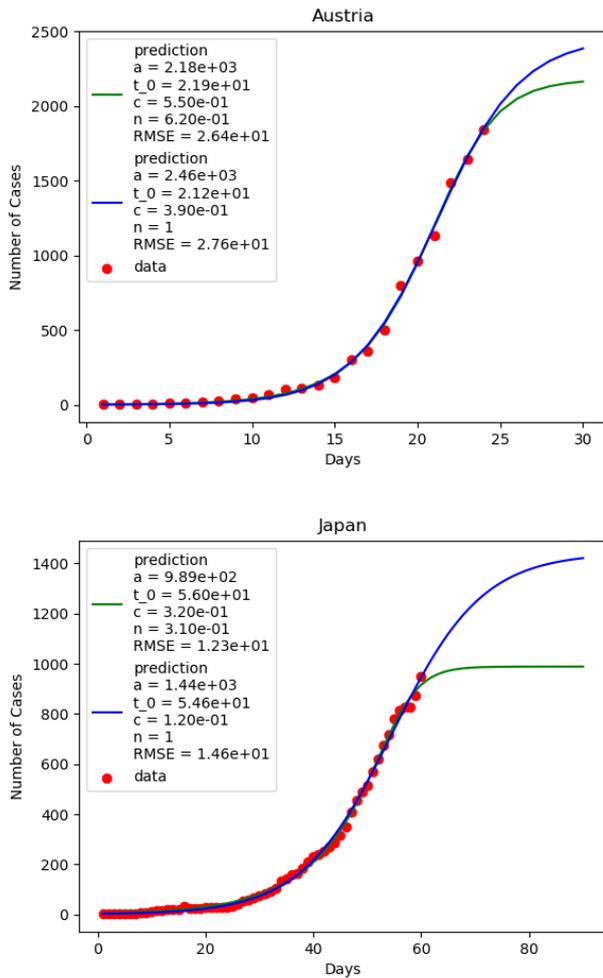


FIG. 3: The parameter fit on data from Austria and Japan. The green line and blue line are the best fit from RG-inspired logistic model and logistic model respectively.

cides with the super-spreader event [19] that happened in Korea on 20 February 2020. Using n as a characterisation on asymmetry, we in the next subsection proceed the estimation analysis on countries in growth-dominated phase.

Growth-Dominated Phase

One of the countries which has not reached the expected inflection point yet is *Italy*. The results of fitting with data for various values of n is shown in Fig.5. Our results suggest that the early stage of epidemic in Italy is not in a controlled stage since the data in early days fit better with a higher value of n . However, we do not pursue the exact value of n of the country before the expected inflection point because of the large uncertainty coming with other parameters especially the size of the epidemic, a which we will discuss later.

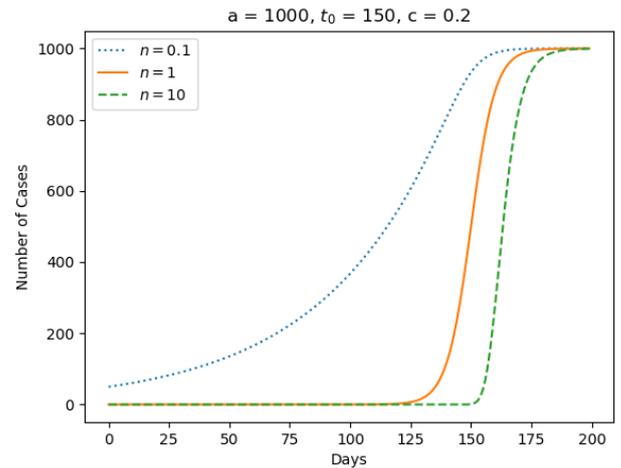


FIG. 4: The plots demonstrate the effect of n on the population growth. $n > 1$ suggests a growth rate higher than typical logistic function ($n = 1$) whereas $n < 1$ suggests a slower growth.

Instead, we first focus on using $n = 1$ for countries in this phase. Some of them are in an early stage of epidemic with insufficient amount of data leading to a large uncertainty on parameter fits. Many of them have an official number of infected cases that might be under investigated due to limited resources in testing kits. We will focus on 5 European countries (*Italy, Spain, France, Germany, and Switzerland*) since their outbreaks have been happening for a while and their health system are somehow capable of reporting data. Examples of the results from parameter fit with $n = 1$ are given in Fig.6 where the error of parameter fits are displayed in TABLE II.

To complete the discussion on the estimation of epidemic size for countries in growth-dominated phase, we find the best fit for each value of n ranging from 1-5 and provide the multiplying factor a/a_0 where a_0 is the prediction of the size of each epidemic from normal logistic function with $n = 1$ (see the second column in TABLE II). The results are given in FIG.7. The multi-

Country	a	t_0 (days)	c (day ⁻¹)
Italy	92968 ± 4379	51.1 ± 0.4	0.206 ± 0.004
Spain	35267 ± 1723	49.2 ± 0.3	0.325 ± 0.008
France	22894 ± 1077	56.5 ± 0.3	0.262 ± 0.004
Germany	52579 ± 17033	58.6 ± 1.9	0.242 ± 0.010
Switzerland	6692 ± 1209	23.2 ± 1.2	0.301 ± 0.026

TABLE II: The table shows the best-fitted values of parameters (a, t_0, c) for each country categorized in the growth-dominated phase. Here the results are of the parameter fit with $n = 1$. In this case, we observed that the best fits still contain higher uncertainty compared with the countries in the maturing phase.

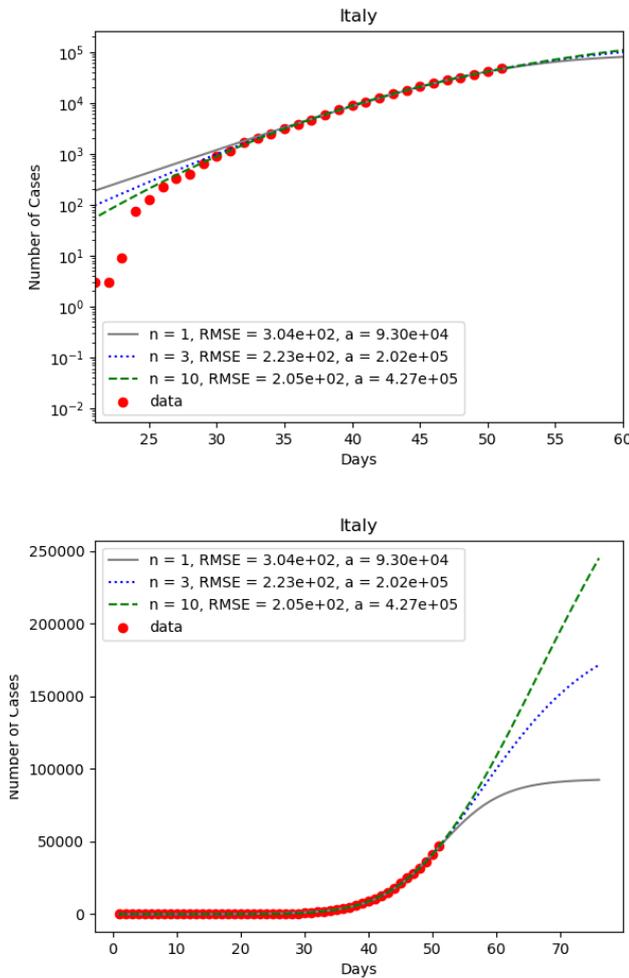


FIG. 5: The parameter fit on data of cumulative infected cases of Italy for $n = 1, 3, 10$.

plying factor can be interpreted as an uncertainty in the final size of each epidemic. In order to provide a reliable prediction of model, we need more data from these countries.

DISCUSSION & CONCLUSION

In this section, we concluded our findings on analyzing the epidemic data of cumulative infected cases from many countries as reported by WHO. In terms of a renormalisation group approach, we considered the RG-inspired logistic growth function, a.k.a. the power-law logistic growth function. The non-linear least squares regression is performed to obtain parameters from the model. We computed the uncertainty for model parameters using the squared root of the corresponding diagonal components of the covariance matrix. We carefully divided countries under consideration into 2 categories based on the estimation of the inflection point, t_0 : the maturing phase

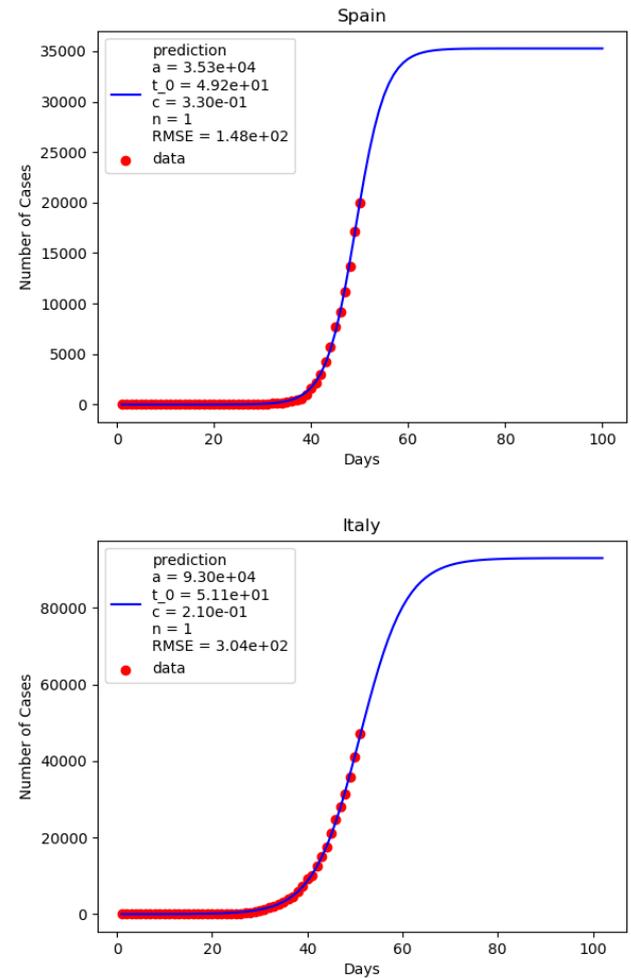


FIG. 6: The parameter fit on data of cumulative infected cases of Spain (upper panel) and Italy (lower panel) for $n = 1$.

and the growth-dominated phase. We found that the outbreak has happened in a large scale and passed beyond the inflection point include China and Korea, while countries where the situation has already passed the inflection point but has not reached the plateau yet are Austria, Belgium, Japan, and Norway with Japan being close to the asymptotic number of infected cases. We observed that long-term estimations of all countries in the maturing phase for both $n = 1$ and $n \neq 1$ are close to each other. Based on the values of root-mean-square error, the RG-inspired logistic model is slightly preferable. Our further investigation shows that n characterizes an early stage of the epidemic. $n < 1$ represents countries with a strong measure at the beginning of the outbreak, where $n > 1$ indicates an uncontrollable event.

In the second category, we consider all countries that the cumulative number of infected cases does not reach the expected inflection point, yet. The European countries that we consider fall in this category include Italy, Spain, France, Germany, and Switzerland. Due to the

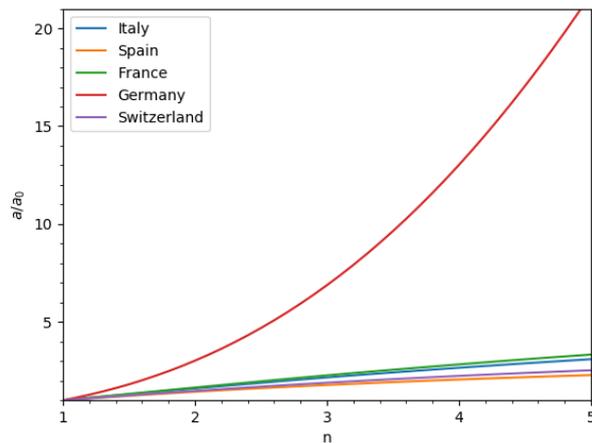


FIG. 7: The multiplying factor a/a_0 as a function of n of 5 European countries. We observed that up till $n \leq 5$ the factor a/a_0 is not greater than 2 apart from Germany. This factor is useful for the estimation of the size of epidemic.

large uncertainty coming from other parameters especially the size of the epidemic, a , in this category we focus on using $n = 1$ for countries in this phase. We estimate the asymptotic number of cumulative infected cases of countries in the growth-dominated phase. We also provide an estimate of the epidemic size as a function of n . However, in order to obtain a reliable prediction of model, we need more data from these countries in the second category. In conclusion, we provide an alternative way to describe the spread of COVID-19 using the renormalization group method originated from physics' point of view.

-
- [1] Coronavirus. (2020). Retrieved 21 March 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
- [2] Danby, J. (1985). Computing applications to differential equations. Reston, Virginia: Reston.
- [3] Brauer, F. (2018). Early estimates of epidemic final sizes. *Journal Of Biological Dynamics*, 13(sup1), 23-30. doi: 10.1080/17513758.2018.1469792
- [4] Miller, J. (2012). A Note on the Derivation of Epidemic Final Sizes. *Bulletin Of Mathematical Biology*, 74(9), 2125-2141. doi: 10.1007/s11538-012-9749-6
- [5] Murray, J. (2003). *Mathematical biology*. Berlin [Germany]: Springer-Verlag.
- [6] Fisman, D., Khoo, E., & Tuite, A. (2014). Early Epidemic Dynamics of the West African 2014 Ebola Outbreak: Estimates Derived with a Simple Two-Parameter Model. *Plos Currents*. doi: 10.1371/currents.outbreaks.89c0d3783f36958d96ebbae97348d571
- [7] Pell, B., Kuang, Y., Viboud, C., & Chowell, G. (2018). Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics*, 22, 62-70. doi: 10.1016/j.epidem.2016.11.002
- [8] Wu, J., Leung, K., & Leung, G. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395(10225), 689-697. doi: 10.1016/s0140-6736(20)30260-9
- [9] Nesteruk, I. (2020). Statistics-Based Predictions of Coronavirus Epidemic Spreading in Mainland China. *Innovative Biosystems And Bioengineering*, 4(1), 13-18. doi: 10.20535/ibb.2020.4.1.195074
- [10] Milan Batista, "Estimation of the final size of the second phase of the coronavirus epidemic by the logistic model," medRxiv 2020.03.11.20024901. doi: <https://doi.org/10.1101/2020.03.11.20024901>
- [11] Milan Batista, "Estimation of the final size of the COVID-19 epidemic," medRxiv 2020.02.16.20023606. doi: <https://doi.org/10.1101/2020.02.16.20023606>
- [12] M. M. Della and F. Sannino (2020). "On the COVID-19 Epidemic Outbreak,"
- [13] Wilson, K. (1971). Renormalization Group and Critical Phenomena. I. Renormalization Group and the Kadanoff Scaling Picture. *Physical Review B*, 4(9), 3174-3183. doi: 10.1103/physrevb.4.3174
- [14] Wilson, K. (1971). Renormalization Group and Critical Phenomena. II. Phase-Space Cell Analysis of Critical Behavior. *Physical Review B*, 4(9), 3184-3205. doi: 10.1103/physrevb.4.3184
- [15] R. K. Pathria, "Statistical Mechanics," Butterworth-Heinemann (1996).
- [16] Braun, J. (2012). Fermion interactions and universal behavior in strongly interacting theories. *Journal Of Physics G: Nuclear And Particle Physics*, 39(3), 033001. doi: 10.1088/0954-3899/39/3/033001
- [17] Fakhri, H., & Sadeghi, J. (2004). Supersymmetry approaches to the bound states of the generalized Woods-Saxon potential. *Modern Physics Letters A*, 19(08), 615-625. doi: 10.1142/s0217732304013313
- [18] Starobinsky, A. (1980). A new type of isotropic cosmological models without singularity. *Physics Letters B*, 91(1), 99-102. doi: 10.1016/0370-2693(80)90670-x
- [19] "Superspreader" in South Korea Infects Nearly 40 People with Coronavirus - Interest News." Interest News, 20 Feb. 2020, [interestnews.com/health/superspreader-in-south-korea-infects-nearly-40-people-with-coronavirus.html](https://www.interestnews.com/health/superspreader-in-south-korea-infects-nearly-40-people-with-coronavirus.html).