Hue University Journal of Science: Techniques and Technology; pISSN 2588-1175 | eISSN 2615-9732

Vol. 131, No. 2B, 2022, 15–32; DOI: 10.26459/hueunijtt.v131i2B.6644

Modeling and forecasting the spread of COVID-19 pandemic: The case of Vietnam

Mai Xuan Thanh, Mai Van Xuan*

University of Economics, Hue University

Abstract. Examining the evolution of the COVID-19 pandemic in many countries since the beginning of the pandemic outbreak in early 2020, we found a common pattern of daily infections with a skewed distribution with different peaks. This trend was observed in Vietnam. Based on those observations, we adapted the skewed distribution function Logistic Growth (Skewed Logistic Growth – SLG) to develop our model for forecasting COVID-19 infections. In the case of Vietnam, we focused on the fourth outbreak – the largest and most complicated pandemic in the country to date. The results depict a clear pattern following closely with the actual development of three infection waves during the fourth outbreak. This confirms that the model can be used to forecast the spread of COVID-19 in the coming time, as the pandemic situation will be more complicated due to the appearance of new variants (i.e., Delta, Omicron, etc.) along with critical adjustments in the government pandemic control and prevention strategies. The model forecasted that the fourth outbreak would peak between the end of December 2021 and the end of January 2022, with about 16,000 new cases per day. The forecasting results are useful for the government and relevant agencies to proactively design timely and effective solutions for prevention. It further proposes various directions for future research to enrich the methodological aspects and empirical evidence of the research domain.

Keywords: COVID–19, forecasting model, pandemic spread, infection waves, Vietnam

1 Introduction

The COVID-19 pandemic has been widespread globally with unpredictable development. This is one of the world's largest pandemics regarding the number of infections and deaths. Since the first outbreak in December 2019 in Wuhan, China, the world has undergone up to the 4th outbreak, with more than 350 million infections and 5.5 million deaths in over 220 countries and territories [1–3]. In Vietnam, despite strict prevention measures, the pandemic has still broken out formidably. According to the statistics of the Ministry of Health, as of Jan. 10, 2022, Vietnam has 1,899,575 infected cases (ranked 28th out of 224 countries and territories) and 34,319 deaths [4]. SARS-CoV-2 has developed more complex variants, such as Alpha, Delta, and Omicron, with a faster spreading speed, causing an increasing infection scale worldwide. In that context, numerous studies have been carried out to predict the extent and infection rate of this pandemic

* *Corresponding: xuanmv@yahoo.com*

Received: 17–12–2021; Accepted: 22–01–2022

to facilitate timely and effective responses to the pandemic. Several forecasting methods have recently been introduced and applied to forecasting COVID-19, such as the use of the R_T infection coefficient by Ha Thanh Trung et al. [5], the Exponential Growth model by Bertozzi et al. [6], or the Susceptible Infected Recovered model (SIR) [6–8], which provided reliable results. However, because of the complicated development of SARS-CoV-2 with many variants having different infection rates, lengthening their affecting time and broadening the transmission scale, there has hardly been a complete model to forecast the pandemic development. This calls for more efforts to improve the existing forecasting models.

Predicting the course of an epidemic is extremely difficult because it is affected by different factors, especially in the case of SARS-CoV-2. The American Centres for Disease Control and Prevention (CDC) have synthesized different methods and models (called Ensemble Model) to forecast the situation and development of the SARS-CoV-2 pandemic [5].

The evolution of the pandemic during the last two years has revealed that its development has a pattern of skewed distribution with numerous outbreaks. The extent and spreading rate of each outbreak depend significantly on many factors, especially the nature and variation of SARS-CoV-2. Based on previous studies on phenomena that had a similar distribution pattern as of the evolution of the COVID-19 pandemic, namely the Logistic Growth Function [9, 10, 13, 14], we aim to develop a mathematical model of Skewed Logistic Growth (SLG) to simulate and forecast the spread of the COVID-19 pandemic over time and by geographical regions. Based on the databases collected by the Ministry of Health of Vietnam and related organizations, such as Our World in Data, Johns Hopkins University, and World Health Organization (WHO), we calibrate to provide reliable forecasting results of the pandemic spread in specific periods. The volatility of the model depends on four key factors: i) the changing state of the Coronavirus variants; ii) the government policy and approach to the pandemic prevention and control; iii) community awareness and attitudes; iv) advances in science and technology, especially in vaccines and therapeutics. The forecasting results are useful for the government and relevant agencies to proactively design timely and effective solutions to the pandemic.

2 Theoretical background for modeling and forecasting

By analyzing the situation of COVID-19 infection in many countries since the initial phase of the global pandemic outbreak (the beginning of 2020 until present), we found that the number of daily infected cases has a skewed distribution pattern with different peaks. To facilitate a clearer observation of this pattern, we highlight some selected countries with a different number of daily infected cases, in particular the United States of America (USA), India and the United Kingdom (UK) (Fig. 1), and Germany, Argentina, Italy and Russia, in the following demonstration (Fig. 2).

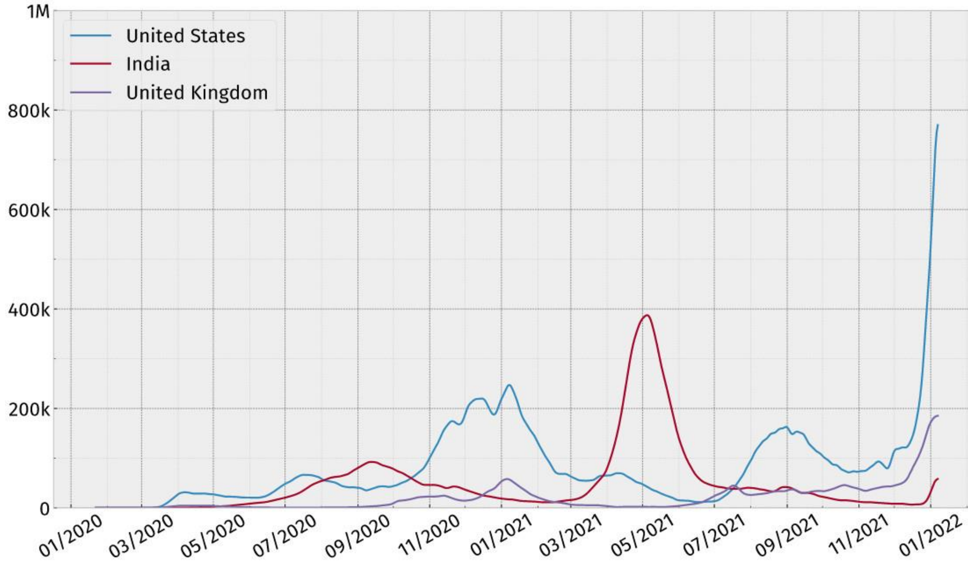

Fig. 1. Daily infected cases in the United States, India, the United Kingdom from 1/2020 to 1/2022

Source: [1–3]

In general, the COVID-19 pandemic is evolving and spreading in all countries on an increasingly large scale. Outbreaks depend largely on the nature of SARS-COV-2 with the occurrence of different variants.

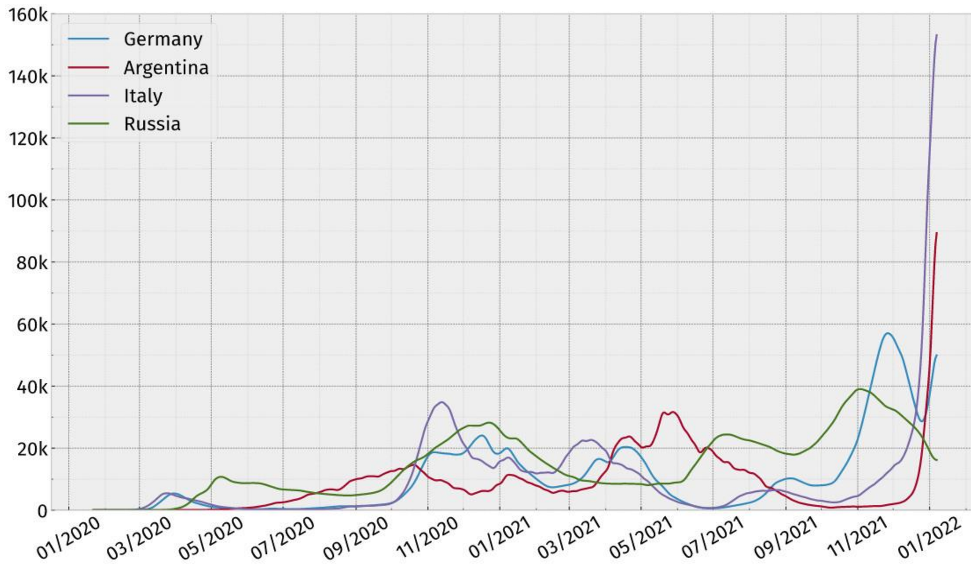

Fig. 2. Daily infected cases in Germany, Argentina, Italy and Russia from 1/2020 to 1/2022

Source: [1–3]

The data show that the world has experienced four different COVID-19 outbreaks (Fig. 3, Table 1). Until Jan. 10, 2022, the total number of infections has reached over 310 million cases and the number of deaths is roughly 5.49 million [1–3]. The first outbreak lasted from January 2020 to Feb. 15, 2021, and the highest number of daily infections was over 773,000 cases. In the second outbreak, from Feb. 15, 2021, to June 21, 2021, the highest number of infections was nearly 906,000 cases. The third outbreak was from June 21, 2021, to Oct. 3, 2021, with a peak of more than 790,000 daily infected cases, aggravated by the appearance of its new variants, such as Alpha and Delta. Especially, the fourth outbreak, lasting from Oct. 3, 2021, to date, with the arrival of the Omicron variant, has recorded the highest number of daily infected cases as compared with other variants (Table 1). The weekly average (Jan. 3–10, 2022) of infected cases in this outbreak is 2,555,375 cases per day [4, 11], although the vaccine coverage rate increased significantly from 0.73 to 49.2% of the world population (Table 2).

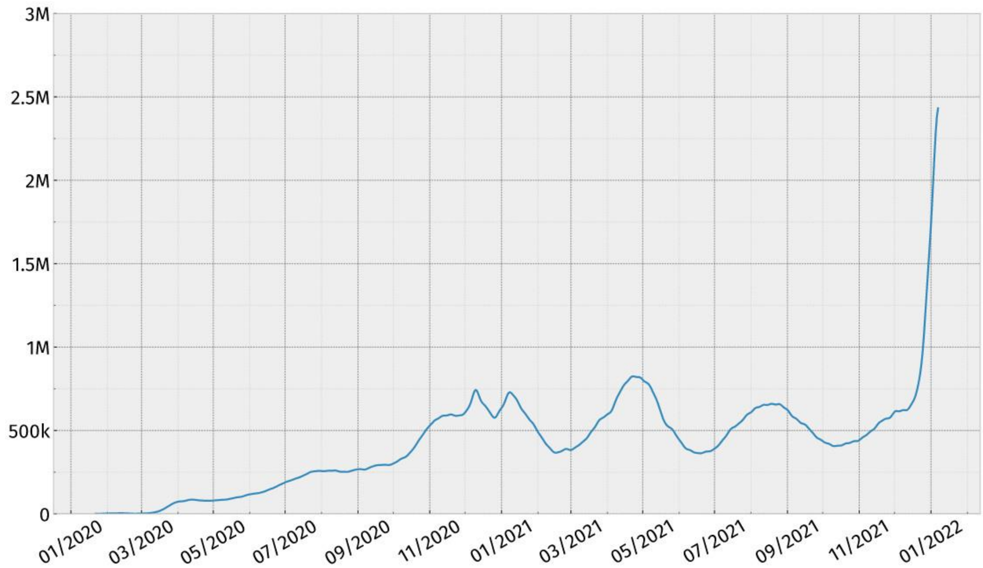

Fig. 3. Daily infected cases worldwide

Source: [1–3]

Table 1. Global outbreaks of COVID-19

Out-break	Time	Total daily infected cases		Weekly average of daily infected cases		Time	
		Peak	Bottom	Peak	Bottom	Peak	Bottom
1st	Jan. 1, 2020– Feb. 15, 2021	773,396	280,002	707,001	383,645	15/1/2021	15/2/2021
2nd	Feb. 15, 2021–Jun. 21, 2021	905,842	299,934	826,973	359,200	28/4/2021	21/6/2021
3rd	Jun. 21–Oct. 3, 2021	790,463	307,747	656,471	437,906	20/8/2021	3/10/2021
4th	Oct. 3, 2021 to date*	3.286.345*	–	2.555.375	–	10/01/2022	–

Source: [4, 11]

Note: * The number of infected cases on Jan. 10, 2022.

Table 2. World’s vaccine coverage (Percentage of the population receiving two doses)

Continent	**28/2/21**	**30/4/21**	**30/6/21**	**30/8/21**	**30/10/21**	**30/12/21**
World	**0.73**	**3.60**	**8.17**	**27.03**	**38.47**	**49.20**
Europe	1.80	8.80	28.18	47.65	54.88	61.26
North America	5,10	21.0	33.40	42.55	52.19	58.19
South America	0.53	6.20	12.90	30.29	50.62	63.91
Asia	0.24	1.40	3.41	28.63	42.44	56.42
Africa	0.00	0.36	1.14	2.73	5.92	9.23

Source: [3, 11].

Currently, Vietnam is coping with the fourth outbreak of the pandemic. According to the Ministry of Health’s data, the pandemic situation is very complicated not only in terms of extent, speed, and duration (accounted for more than 99% of the cumulative infected cases since the beginning of the pandemic) but also of increasing threat caused by new variants of the virus, such as the Delta and Omicron (Table 3). Therefore, this forecasting model focuses on examining the 4th outbreak in Vietnam.

Table 3. Outbreaks of the COVID-19 pandemic in Vietnam

Out-break	**Time**	**Number of infected cases**			
		Total	**Domestics**	**Immigrated**	**Deaths**
1st	Jan. 23–Jul. 24, 2020	415	106	309	0
2nd	July 25, 2020–Jan. 27, 2021	1,136	554	582	35
3rd	Jan. 28–Apr. 26, 2021	1,301	910	391	0
4th	Apr. 27, 2021, to date*	1,911,541+	1,908,353+	3,188	34,531+

Source: [4], Vietnam Ministry of Health, Jan. 2022

Note: * Estimated up to Jan. 10, 2022

Using the graphic method and the Gaussian smoothing (Fig. 4), we can see that the 4th outbreak of the pandemic in Vietnam exhibits three different waves or three phases. The first wave lasted from Apr. 27, 2021, to Aug. 4, 2021, with the highest number of daily infections at 8,198 cases. The second wave, from Aug. 4, 2021, to Oct. 18, 2021, reached the highest number of daily infections at 13,027 cases. The third wave, from Oct. 18, 2021, to date, has been the longest, most large-scale and most complicated development with new variants of Coronavirus, especially the Omicron (Table 4).

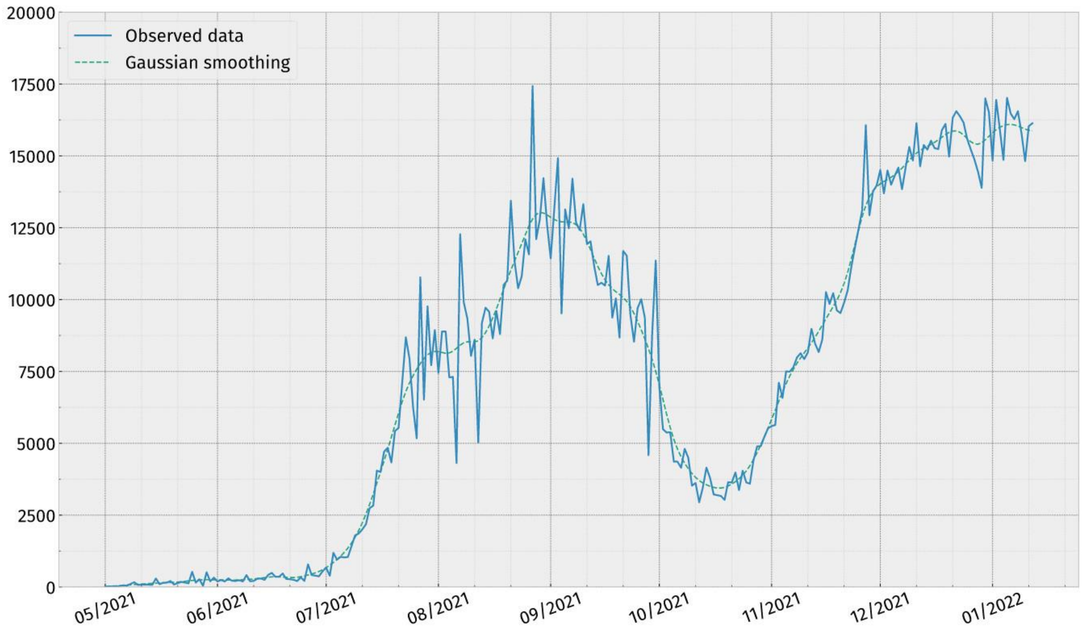

Fig. 4. Pattern of the daily infected cases in Viet Nam in the fourth outbreak (Apr. 27, 2021 to Jan. 2022)

Source: [3, 4]

Table 4. Waves of the forth pandemic outbreak in Vietnam (data adjusted using the method of Gaussian smoothing with σ = 3)

Outbreak	Time	Total daily infected cases		Time	
		Peak	Bottom	Peak	Bottom
1st	Apr. 17–Aug. 4, 2021	8,198	8,134	Jul. 31, 2021	Aug. 3, 2021
2nd	Aug. 4–Oct. 18, 2021	13,027	3,443	Aug. 29, 2021	Oct. 18 2021
3rd	Oct. 2021, to date*	16,000	–	Dec. 2021–Jan. 2022	–

Source: [3, 4] and the authors' calculation

Note: * Estimated up to Jan. 10, 2022

The epidemic wave, such as in the Covid-19 pandemic, is a widely accepted term but lacks a concrete academic definition. To enable a common ground of understanding, in this study, we follow the definition of wave proposed by Zhang et al. A wave is defined based on two critical characteristics: 1) a wave consists of upward and downward periods of infection cases; 2) the increasing number in the upward trend and the decreasing number in the downward one must be substantial and sustain over a while [20]. This differentiates the trends of a wave from other pandemic movements, such as an uptick, a downtick, reporting errors, or volatility in new cases

[20]. The time frame of a wave is relatively short, and the upward/downward trends can occur several times during a specific outbreak. Therefore, there can be multiple waves in a pandemic/outbreak.

Based on the previous studies [9, 10, 13, 14] on phenomena that shared similar distribution patterns to those of the pandemic development in the observed countries (Fig.s 1, 2, 3 and 4), we formulated our model using the Logistic Growth function [14] to forecast the daily number of COVID-19 infections as follows:

$$\mathrm{F}(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}} \qquad (1)$$

where x is the observed time; μ is the location parameter (determining the mid-point of the curve); σ is the scale parameter (determining the elasticity and slope of the curve).

This function has an S-shape and is often used to encapsulate growth phenomena with respect to time, consisting of three following phases: i) the initial growth phase is approximately exponential; ii) after approaching the saturation phase, it slows down; and iii) at maturity phase, the growth ceases (Fig. 5).

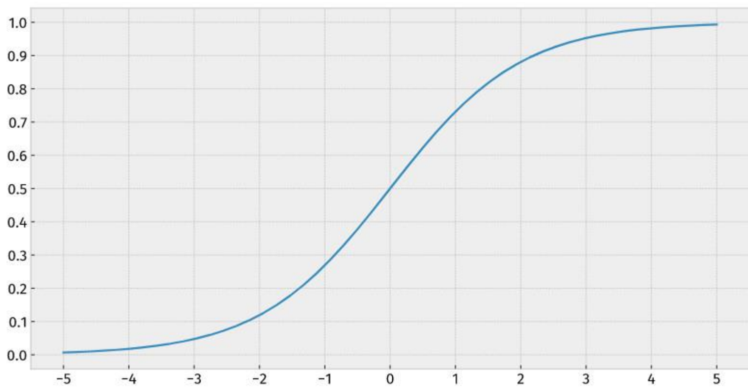

Fig. 5. Logistic Growth (Source: Verhulst, 1838 [14])

The S-shape curve is derived by taking the accumulative summation of observed phenomena over time. The growth rate of the phenomenon can be expressed as the logistic distribution function as follows:

$$\mathrm{f}(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma(1 + e^{-(x-\mu)/\sigma})^2} \qquad (2)$$

This function is derived by taking the derivative of the function in (1), and it has a shape analogous to a standard normal distribution curve (Fig. 6).

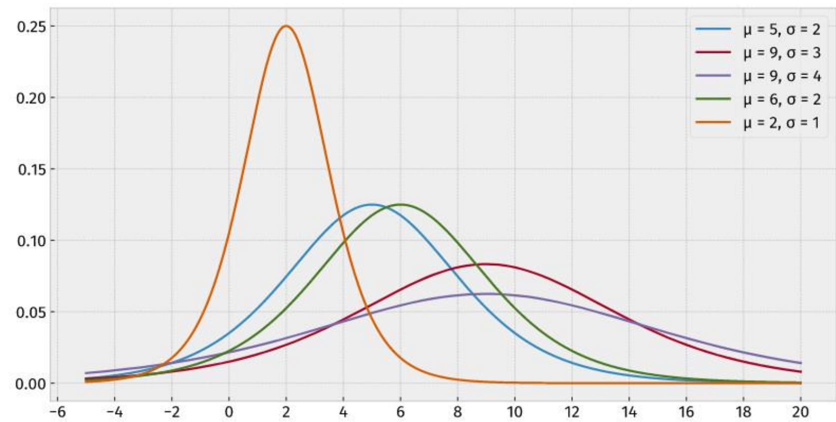

Fig. 6. Logistic distribution curves with different parameters

The skewed logistic distribution function takes the following form [9, 10, 13]

$$\varphi(x) = 2\mathrm{f}(x)\int_{-\infty}^{\alpha x}\mathrm{f}(t)\,\mathrm{d}t = 2\mathrm{f}(x)\mathrm{F}(\alpha x) \quad (3)$$

where $\mathrm{f}(x)$ is derived from (2) and $\mathrm{F}(x)$ derived from (1).

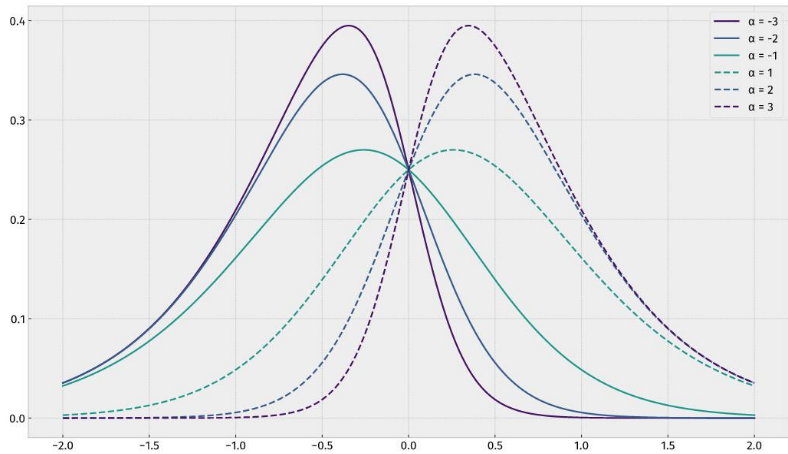

Fig. 7. Skewed logistic distribution curves with different α parameters [9]

Here, α is the shape parameter, determining the direction and slope of the distribution function. When $\alpha > 0$, the distribution will be skewed to the right (positive skewness), and vice versa, when $\alpha < 0$ the distribution will be skewed to the left (Fig. 7). If $\alpha = 0$, hence $\varphi(x) = 2\mathrm{f}(x)\mathrm{F}(0)$, of which $\mathrm{F}(0) = \frac{1}{1+e^{-0x}} = \frac{1}{2}$ (function (1)); then, $\varphi(x)$ is the logistic distribution curve similar to normal distribution $f(x)$ in the function in (2).

3 Specification of the SLG forecasting model

From functions (1), (2), and (3), we can develop the model to forecast the transmission of the COVID-19 pandemics in the coming time. From the function of skewed distribution (3), we can specify as follows:

$$\varphi(x) = 2f(x)\int_{-\infty}^{\alpha x} f(t)\ dt = 2f(x)F(\alpha x)$$

where f is the Logistic Growth function from (1); F is Logistic distribution function from (2); then,

$$\varphi(x) = 2\frac{e^{-(x-\mu)/\sigma}}{\sigma(1+e^{-(x-\mu)/\sigma})^2}\frac{1}{1+e^{-\alpha(x-\mu)/\sigma}}$$

$e^{-(x-\mu)/\sigma}$ is repeated many times in the function; therefore, we need to define η to shorten it

$$\eta \triangleq e^{-(x-\mu)/\sigma}$$

$$\varphi(x;\ \mu;\ \sigma;\ \alpha) = \frac{2\eta}{\sigma(1+\eta)^2(1+\eta^{\alpha})} \qquad (4)$$

Function (4) can be used to forecast the number of infected cases for any regions and the whole country. This equation can be written in the Python programming language:

```
def LogisticCurve(X, k, sigma, mu, alpha):
        r = (X – mu) / sigma
        eta = np.where(np.abs(r) > 10, 0, np.exp(-r))
return k * eta / ((1 + eta)**2 * (1 + eta**alpha))
```

Note:

In reality, there may be different scenarios that x is large and/or σ is small, if setting $r = -(x-\mu)/\sigma$, hence r may take a huge value. When r exceeds 710 then $\eta > 2^{1024}$, which can be taken as the computational limit of float64 data in Python. Moreover, when $|r|$ increases, then $\varphi(x)$ decreases to 0. Therefore, when $|r| > 10$, we set $\varphi(x) = 0$.

Function (4) is multiplied by k to calibrate the magnitude of the distribution curve. Besides, when running regression, $2/\sigma$ will be excluded from function (4) to facilitate a simpler and faster computation with no effect on the forecasting result thanks to the compensation from scalar k. The adjusted function is named $\widehat{\varphi}(x)$, where

$$\widehat{\varphi}(x,\mu,\sigma,\alpha,k) = k\frac{\eta}{(1+\eta)^2(1+\eta^{\alpha})} \qquad (5)$$

The scalar k has the following property:

$$\int_{-\infty}^{\infty}\widehat{\varphi}(x)\,\mathrm{d}x = k\frac{\sigma}{2}$$

In other words, the total number of infected cases in each wave is approximately $k\sigma/2$.

The function $\widehat{\varphi}(x)$ represents a single wave of an outbreak. Thus, as observed in Fig. 4 and the data in Table 4, the 4th outbreak has three waves. To represent the movement of the 4th outbreak, we add up three functions with different parameters

$$\widehat{\varphi}_{\Sigma}(x) = \sum_{n=1}^{N} k_n \widehat{\varphi}_n(x, \mu_n, \sigma_n, \alpha_n)$$

$$= k_1\widehat{\varphi}_1(x, \mu_1, \sigma_1, \alpha_1) + k_2\widehat{\varphi}_2(x, \mu_2, \sigma_2, \alpha_2) + k_3\widehat{\varphi}_3(x, \mu_3, \sigma_3, \alpha_3) \qquad (6)$$

In function (6), parameter n (n = 1÷N; N is the number of waves) represents the nature of the nth wave, where x is the time (in days, counting from Jan. 1, 2021); k is the adjusting parameter of magnitude; μ is the location parameter; σ is the scale parameter; α is the shape parameter, determining the skewness of the curve.

Function (6) is the sum of the skewed logistic distributions, known as the skewed logistic growth (SLG) model. SLG model is built on the basis of applying studies on skewed logistic distribution patterns and the actual evolution of the COVID-19 epidemic in countries around the world. Function (6) is a temporal continuous function from –∞ to +∞. The model is assessed to have some advances compared with other forecasting models about the COVID-19 pandemic in the past.

In the context of Vietnam, the fourth outbreak has three waves (Fig. 4). Therefore, n is identified as 3 in the equation, making a sum of three skewed logistic distributions (N = 3) to forecast the epidemic movement during this outbreak.

4 Forecasting results from the SLG

Using data on COVID-19 infections from the pandemic databases of the Vietnam Ministry of Health [4] and Johns Hopkins University [3], we present the results of the forecasting model in Fig. 8.

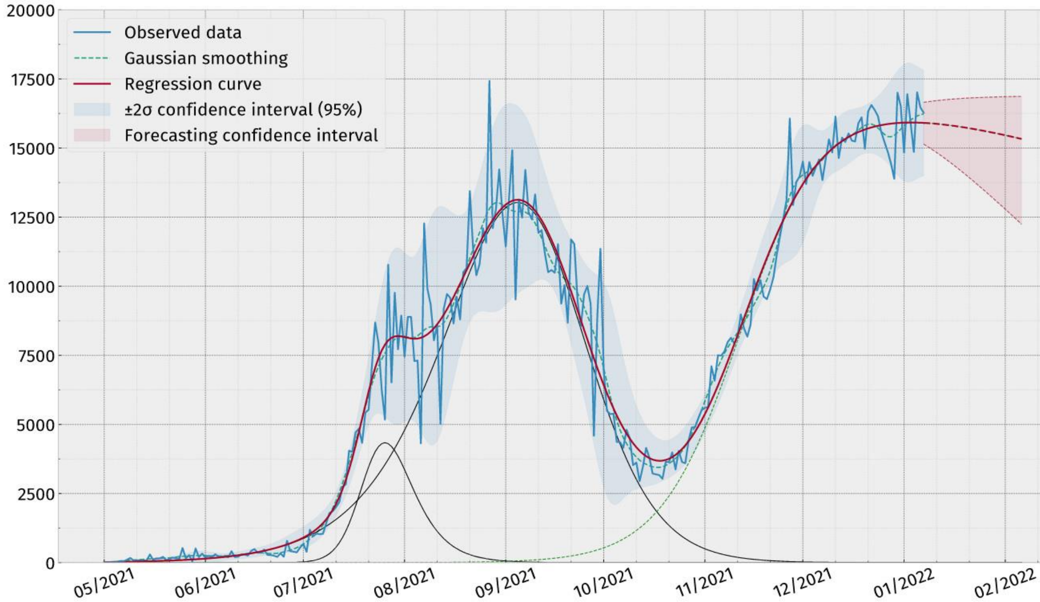

Fig. 8. Forecasting the number of infected cases in Vietnam during the 4th outbreak, by SLG

Note: The blue curve is our observed data; the red curve is the resultant function with parameters acquired from regression and forecasting results; the dashed red curve is the Gaussian smoothing of our observed data (with σ = 3). Three thin curves separately represent each of the waves in the 4th outbreak, that sum up to form the red curve. R_2 = 0.9679.

Table 5. Parameters of the regression model

Wave	k	σ	μ	α	Total cases ($k\sigma/2$)
1	27,259.1213	6.8984	200.8116	1.3308	94,023
2	94,458.8837	16.2070	255.9468	-0.6957	765,446
3	66,980.4926	134.6801	315.0427	10.8127	4,510,470*

Source: Author's computation, 2021

Note: *Forecasted number of infected cases at the end of the 3r wave

The parameters k, σ, μ, α in Table 5 are determined through regression by using the Levenberg-Marquardt algorithm. These parameters indicate the size, shape and position of the forecasted curve. The larger the value of k associated with the larger the corresponding wave, and vice versa. However, the magnitude of each wave also depends further on the value of σ. The larger the values of k or/and σ mean, the larger the corresponding wave, implying the larger the number of infections of that wave and vice versa. To emphasize, the model has not defined a breakpoint for the third wave; hence, the parameter values of wave 3 have a larger difference than those of waves 1 and 2. Interference in the values of the parameters, for example, determining their limits, makes the model's prediction results less objective.

According to the above forecasting model, the 4th outbreak can be divided into three waves:

1) *The first wave started on Apr. 27, 2021, and peaked on Jul. 30, 2021, with a forecast of 8,197 cases per day, while the actual data was 12,275 cases (Fig. 8).*

Thanks to the positive effects of Directive No. 16/CT-TTg [17], the number of infections was more effectively controlled, and the 3rd outbreak in our country had ended with a total of 910 cases of infection by the end of April 2021 (Table 1). However, due to the long holidays (April 30 and May 1), fueled by the reopening of tourism businesses in many localities during this time, the number of infections promptly increased.

Facing the complicated development of the pandemic, especially in Ho Chi Minh City and southern provinces, Ho Chi Minh City was forced to enact Directive 16 once again throughout the city from 0:00 on July 9, 2021. The nationwide medical force was mobilised to support Ho Chi Minh City, and many field hospitals were set up in the region to increase the capacity of healthcare services. The strong government commitments and community consensus created synergy for more effective pandemic control and prevention. However, from the end of July 2021 to early August 2021, the massive migration of thousands of workers from Ho Chi Minh City and the southern provinces (Dong Nai, Binh Duong, etc.) to other localities caused huge challenges in pandemic prevention and control in the whole country.

2) *The second wave of pandemics started on Aug. 4, 2021, and peaked on Sept. 5, 2021, with the forecasted number of infections at 13,129 cases. The actual data at this time was 17,428 cases.*

During this wave, the pandemic occurred with a larger scale and a higher infection rate, especially in Ho Chi Minh City and its vicinities. The government had to mobilize more resources (including the army) to cope with the pandemic spreading, which was extremely complicated because of the appearance of the Delta variant, while vaccination remained very limited. By Sept. 1, 2021, the rate of vaccination had reached only 18% of the population. The percentage of the population receiving two doses of vaccine was about 3% [1, 4]. In that context, the government had urged the localities (especially Ho Chi Minh City and its vicinities) to strictly implement Directive 16 and Directive 16+ until the end of Sept. 30, 2021, and to promote the execution of the vaccine strategy. As a result, the vaccination coverage was quickly improved and had reached 34% of the population by Oct. 1, 2021. The percentage of the population receiving two doses increased to 10% [1, 4]. Thanks to these efforts, the pandemic was gradually put under control.

Although the number of infections tended to decrease significantly and the vaccination coverage increased, right after easing social distancing following Directive 16, tens of thousands of workers left the region. They returned to their hometowns across the country. This larger scale of migration again induced difficulties in disease control and prevention in many provinces. Along with the impacts of other contextual factors, the third wave began.

3) The third wave started on Oct. 18, 2021, with a surge in daily infections. The regression model forecasted that this wave would last from the end of Dec. 2021 to the end of Jan. 2022, with the number of infections ranging from 15,500 to 16,000 cases per day (Fig. 8). The forecasted number was very close to the actual one during the last seven days (Jan. 4–10, 2022) at 15,935 cases.

This third wave appeared to be different from the previous ones. First of all, on Sept. 30, 2021, Ho Chi Minh City issued Directive 18 on easing social distancing and reopening many businesses and commercial activities [18]. Again, thousands of workers left the region for their hometowns, which was deemed to exacerbate the pandemic pressure across the country. Secondly, the government strategy on pandemic prevention changed dramatically, from strict and massive lockdown, or "Zero COVID-19", to "living with the virus". Accordingly, many business and service activities were allowed to reopen. The vaccination acceleration, in combination with the implementation of the 5K prevention measures, was regarded as the key approach for pandemic control and prevention [19]. These strategic measures made remarkable progress. By Jan. 10, 2022, 100% of the population aged 18 and above had been vaccinated, of which 92,6% had been vaccinated with two doses. Meanwhile, these rates in the group of 12 and 17 years old had been 89 and 65,7% [4]. This result pushed Vietnam amongst the top countries that had high vaccination rates in the world. That also set an important premise for implementing the government policy of "living with COVID-19" and continuing to reopen the economy. It can be said that the third wave spread on a larger scale and in unpredictable ways; however, people seemed to have more confidence in the government's strategy of disease control and prevention.

Obviously, the forecasting model simulates a quite clear trend of the COVID-19 pandemic spread, which closely follows the actual development of the three waves in this 4th outbreak. That allows us to confirm that this model can be used to forecast COVID-19 infections in the time to come. As the current situation of the pandemic is getting more complicated, along with changes in the government's approaches and measures for pandemic control and prevention, the model predicts that the peak of this wave would occur between the end of December 2021 and the end of January 2022, with the daily infection about 16.000 cases.

Considering the impacts of different contextual factors, such as the generations of new variants of COVID-19, the changes in government policies, and the enhanced awareness of the citizen on pandemic control and prevention, we regularly update this forecasting model to predict the number of infections. The model can produce an accurate prediction of COVID-19 infections 20–30 days in advance, depending on the policy changes and actual developments of the COVID-19 pandemic.

5 Discussion on several forecasting models

Recently, several studies on forecasting the spread of COVID-19 have been published, of which the following models should be mentioned.

1) The forecasting method using parameter R_T

This method was based on two important parameters, namely Basic Reproduction Number (R_0) and Serial Interval (T) [8, 11], from which the number of infected cases was expressed as $I(t) = R$^(n+1)/R_0, where $n = N/T$ (N is the time interval of forecasting). This method has several strengths, including ease of performance, good results forecasted for a specific time interval and a certain location. Ha Thanh Trung et al. [5] also applied this method to forecast the COVID-19 pandemic in some localities in Viet Nam during the third quarter of 2021 with reliable results. However, a key limitation of this method is that it is hard to accurately determine R_0 and T because they depend on many factors, such as the implementation of social distancing, vaccination, community awareness, and especially infectivity of different virus variants. For example, the value of R_0 of SARS-CoV-2 is 2; those of Alpha is 3.4; R_0 for Delta is 5 [1–4]; and with the Omicron, this factor can be much higher. Given that different virus variants coexist, such as in the case of the COVID-19 pandemic, the R and T parameters may differ, leading to limited forecasting results of this method.

2) Exponential growth model

This model was recently used by Bertozzi et al. [6] to forecast the number of COVID-19 infections in some countries, such as Italy, the UK, Germany, the USA, Japan, and South Korea, in April 2021. The number of infections I at time t is I(t) is determined by the following equation

$$\mathrm{I}(t) = I_0 e^{\propto t}$$

where α is the infection rate; I_0 is the number of infections at the start time.

If we start from $\bar{I}$ cases, then $T_d = \ln 2/\alpha$ is the time it takes for the number of cases to double by $2\bar{I}$ (same as 'the 70/x rule of thumb', where x is the growth rate in percentage of average GDP during the period) in the GDP growth forecast. The key advantage of this model is that it provides the forecasted number of daily infected cases in a short period, and when the pandemic just occurred, the number of I_0 cases is still low. However, this is an exponential predictive model, and as α is constant, the number of infections increases over time exponentially and does not tend to slow down. Furthermore, since $\delta I/t = \alpha \mathrm{I}(t)$, the growth rate is always positive, and the number of infections tends to increase faster. This means the model is only effective for a short time, and the number of infections is small.

3) Compartmental models

The compartmental model (SIR – Susceptible, Infectious, or Recovered) is a very general model, which is usually used for forecasting infectious diseases. The model has been used by many researchers, such as Ross and Hudson, Kermack and McKendrick, and Kendall.

The functions of the SIR are as follows:

$$\frac{\partial S}{dt} = -\beta\frac{IS}{N}; \quad \frac{\partial I}{dt} = \beta\frac{IS}{N} - \gamma I; \quad \frac{\partial R}{dt} = \gamma I; \quad R_0 = \frac{\beta}{\gamma}$$

where R_0 is the Basic Reproduction Number; β is the infectivity (supposed to be unchanged); γ is the recovery duration of patients (supposed to be unchanged).

The advantage of this forecasting model is that the coefficients β and γ can be determined through a sample survey. Therefore, it is possible to predict *S*, *I* and *R* quickly [12, 6–8]. However, since β and γ are constant, they seem not to be entirely relevant to the actual situation of the COVID-19 pandemic. This is because the Coronavirus itself is always mutating to reproduce new variants, which have quite different rates of transmission and recovery. In addition, the government's efforts (social distancing, vaccination, etc.) can also change *R*, β and γ. Hence, the predicted results of this model would be relevant to infectious diseases having one variant or the rate of disease transmission and recovery is less varied.

4) The SLG model

In comparison with the above-mentioned models, the SLG model developed in this study has several strengths. Firstly, as it is built based on the Skewed Logistic distribution function, it is conformable to the actual infection situation of SARS-CoV-2. Therefore, this model generalizes and simply quantifies the factors (variables) affecting the transmission of the COVID-19 pandemic over time. Secondly, its forecast results are based on the investigation and synthesis of actual data, encapsulating the factors affecting the change in the infection situation. It reflects quite accurately the number of daily infected cases over time. Thirdly, this model can capture the complex development of SARS-CoV-2 variants as well as synthesize the government's efforts in implementing pandemic prevention policies and vaccination. Finally, the model is based on the Skewed Logistics Growth function, so it is possible to identify the pandemic peak a short time in advance (about 20–30 days), which provides policymakers with a scientific basis for planning and implementing socio-economic development and effectively coping with the pandemic.

6 Conclusion, limitations and implications for future research on the SLG model

6.1 Conclusion

After two years of coping with the COVID-19 pandemic, it has been assured that the government's policy on disease control and prevention and the consensus of the public are critical to prevent the spread of COVID-19. In this context, daily prediction of infected cases is always important to provide information not only for policymakers but also for the public. This study introduces a forecasting model that was developed based on examining the distribution patterns similar to patterns of disease transmission in many countries. The forecasting results are intended to i) provide updated predictions on disease transmission and ii) help relevant agencies to implement proactively and timely solutions for preventing and controlling the spread of COVID-19.

Using time-series data of COVID-19 infections from several pandemic databases of the Vietnam Ministry of Health, Johns Hopkins University and Our World in Data, and WHO, the forecasting model proposed in this study depicted a clear pattern of pandemic transmission in Vietnam. The results show the pattern that closely follows the actual development of three infection waves during the fourth pandemic outbreak. This supports the theoretical and practical significance of the proposed model, which helps to enrich the literature of this research domain.

6.2 Limitations

Development of the COVID-19 pandemic is very complicated and unpredictable because of various factors, such as the evolution of the Coronavirus, changes in the government's approach and strategy to disease control and prevention methods and policy, community awareness and consensus, and especially scientific progress in vaccines and drugs to prevent Coronavirus. As a result, it is challenging to encapsulate every factor into a mathematical model. Although the results of the forecasting model proposed in this study show potential for further application, the model still has some limitations. First, it is difficult to quantify some contextual factors influencing the number of infected cases. Secondly, as well-recognized in multivariate analysis, the theoretical foundation of the regression method is often subjected to some weaknesses from a mathematical perspective that may affect the forecasting results [6, 15, 16]. Further studies on the mathematical theory are needed to improve the forecasting model.

6.3 Implications for future research

Several future studies can be made to enrich the body knowledge of this research domain:

There is a need to incorporate the SLG model with other forecasting methods, especially the SIR model, to enhance the model's forecast results.

The influence of several contextual factors on disease infection needs to be quantified.

Future research should pay more attention to the relationship between the number of infected cases, the rate of seriously infected patients and fatality.

It is possible to build a forecasting model according to different disaggregation criteria:

age, vaccination rate, region, etc.

References

1. Our World in Data. Retrieved from https://ourworldindata.org/covid-cases
2. The World Organization of Health. Retrieved from https://www.who.int/covid-19
3. COVID-19 Data Repository at Johns Hopkins University. Retrieved from https://github.com/CSSEGISandData/COVID-19 /blob/master/csse_covid_19_data/
4. Official website of Ministry of Health, Vietnam. Retrieved from https://covid19.gov.vn/
5. Hà Thành Trung, Nguyễn Khương Tuấn, Võ Trí Hảo. Dự báo dịch bệnh Covid-19 ở nước ta. http://onyx.vn/covid/CovidVietnam.Report.21903.pdf
6. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. (2020). The challenges of modelling and forecasting the spread of COVID-19. Retrieved from https://arxiv.org/pdf/2004.04741.pdf on Dec. 10th 2021.
7. Schlickeiser, R. & Kröger M. (2021). "Analytical solution of the SIR-model for the temporal evolution of epidemics. Part B: Semi-time case". *Journal of Physics A. 54 (17): 175601. Bibcode:2021JPhA...54X5601S. doi:10.1088/1751-8121/abed66*
8. SIR models of epidemics Level 1 module in "Modelling course in population and evolutionary biology" (701-1418-00). *Module author: Florence D´ebarre; Course director: Sebastian Bonhoeffer Theoretical Biology Institute of Integrative Biology ETH Z¨urich*
9. Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics, 12*(2), 171–178.
10. Nadarajah, S. (2009). The skew logistic distribution. AStA Advance Statistical Analysis. 93, 187–203. https://doi.org/10.1007/s10182-009-0105-6
11. Ying, L., Gayle, A. A., Wilder-Smith, A. and Rocklov, J. (2020). The reproductive number of covid-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine, 27(2), 02 2020. taaa021*
12. Cori, A., Ferguson, N.M., Fraser, C. and Cauchemez, S. (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology,* Vol.178 (9), pp. 1505–1512, https://doi.org/10.1093/aje/kwt133
13. , S., Hazarika, P. J. and Ali, M. M. (2012) A new skew logistic distribution and its properties. *Pakistan Journal of Statistics*, 28(4), pp. 513-524.
14. Verhulst, Pierre-François (1838). Correspondance Mathématique et Physique. 10: 113–121.
15. Harko, T., Lobo, F. S. N. & Mak, M. K. (2014) Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Applied Mathematics and Computation*, 236,184–194.
16. Yan, P. (2008). Distribution theory, stochastic processes and infectious disease modelling. In *Mathematical epidemiology*, pp. 229–293. Springer, 2008
17. Chỉ thị số 16/CT-TTg ngày 31/3/2020 của Thủ tướng Chính phủ về thực hiện các biện pháp cấp bách phòng, chống dịch COVID-19.

18. Chỉ thị số 18/CT-UBND ngày 30/9/2021 của UBND thành phố Hồ Chí Minh về việc "tiếp tục kiểm soát, điều chỉnh các biện pháp phòng, chống dịch Covid-19 và từng bước phục hồi phát triển kinh tế - xã hội trên địa bàn thành phố".

19. Nghị quyết số 128-NQ/CP ngày 11/10/2021 của Chính phủ ban hành quy định tạm thời "Thích ứng an toàn, linh hoạt, kiểm soát hiệu quả dịch COVID-19".

20. Zhang, S. X., Arroyo, M. F., Gao, R. & Wang, S. A (2021). Second Wave? What Do People Mean by COVID Waves? – A Working Definition of Epidemic Waves. *Risk Management Healthcare Policy. 14,* pp. 3775-3782. *Doi: https://doi.org/10.2147/RMHP.S326051*