FITTING TWO SIMPLE PROBABILISTIC MODELS FOR COVID–19 WITH DATA FROM ECUADOR

FELIPE BENAVIDES AND JHOANA P. ROMERO–LEITON*

ABSTRACT. **Background:** In this study, we use public data of COVID–19 diagnosed cases from Guayas province (Ecuador) and use them to evaluate the forecasting ability of some probabilistic models commonly used in the epidemiological and mathematical literature. We highlight the precision of the models in predicting COVID–19 cases, and the evolution of the epidemic curve until 2021. Those are helpful to estimate a possible stabilization date for the epidemic. **Methods:** We use Gompertz and Hill functions to provide robust predictions of the COVID–19 diagnosed cases and the stabilization date of the epidemic curve. We simulate the model errors for predictions by including a probabilistic error term from a Gamma distribution. Non–linear least square method is performed trough the Gauss–Newton algorithm with 5000 iterations to estimate each parameter in the models. The Coefficient of Determination, the Akaike Information Criteria and the Root Mean Squared Error are used as the metrics to evaluate the model performances. **Conclusions:** The accuracy of our mathematical models is high compared to diagnosed cases until December 9, 2020. Medium terms predictions indicate an increase of cumulative COVID–19 cases until 2021. Since to date, the diagnosed cases will be increasing until the stabilization is achieved. According to the Hill model the stabilization will be achieved on February, 2021, while for the Gompertz model the stabilization will be on December, 2020.

1. Introduction

In December (2019), the Wuhan Municipal Health Commission (Hubei Province, China) informed to World Health Organization (WHO) about a group of 27 cases of unknown etiology pneumonia, who were commonly exposed to a wet market in Wuhan City. It was also noticed that seven of these patients were critically serious. The symptoms of the first case began on December 8, 2019. On January 7, 2020, Chinese authorities identified a new type of family virus as the agent causing the pneumonia. The causative agent of this pneumonia was identified as a new virus in the *Coronaviridae* family that was officially named SARS–CoV–2, more broadly known as COVID–19. The clinical picture associated with this virus has been named COVID–19. On March 11, WHO declared the global pandemic. Since the beginning of the epidemic, to December 9, 2020 have been reported more than 70,522,778 cases and more than 1,584,795 deaths in the world [6].

FELIPE BENAVIDES AND JHOANA P. ROMERO–LEITON*

Coronaviruses are a family of viruses that cause infection in humans and some animals. Diseases by coronavirus are zoonotic, that is, they can be transmitted from animals to humans. Coronaviruses that infect humans (HCoV) can produce clinical symptoms from the common cold to serious ones like those caused by the Severe Acute Respiratory Syndrome (SARS) viruses and Middle East Respiratory Syndrome (MERS–CoV) [7]. The transmission mechanisms of SARS–COV–2 are animal–human and human–human. The first one is still unknown, but but some researchers claim that it could be driven by respiratory secretions and/or feces [8]. The second one, is considered similar for other coronaviruses through the secretions of infected people, mainly by direct contact with respiratory drops and hands or fomites contaminated with these secretions, followed by contact with the mucosa of the mouth, nose or eyes [8].

The COVID–19 pandemic in Ecuador is an expansion of the COVID–19 outbreak that started in China. This fact alerted all countries worldwide from the beginning of 2020 and at the end of February it reached Latin America. On February 29, Ecuador reported a first case of COVID–19, corresponding to an Ecuadorian woman who had arrived from Spain on February 14 but with no symptoms. On March 13, the first death due to COVID–19 was reported in the country, who was the first infected to arrive from Spain. By December 9, 2020, Ecuador reported 200,379 cases of COVID–19 and more than 13,850 deaths [2]. According to data from the *Johns Hopkins Coronavirus Resource Center* [1], to date, Ecuador is one of the most affected Latin American countries and it ranks second in number of infections and deaths after Brazil, even though its population is twelve times smaller while its territory is 30 times smaller.

The *Financial Times*, an international British daily newspaper, claims that Guayas province in Ecuador, is one of the cities most affected by the COVID–19 pandemic in the world and the province of Ecuador most affected by this disease, both in the number of infections and deaths. In fact, Civil Registry Office of Guayas Province, reported that at least 8,600 people died for different reasons during April (2020), while the average number of deaths in Guayas the past years was 2,000 deaths, this means a death increase of 488%. In fact, the number of diagnosed cases until December 9, 2020 was reported in 25,706 (see Figure 1).

According to the previous scenario and the global epidemic situation, it is important for the scientific community to be informed about the behaviour of the disease in short and medium term. In the early stages of an infectious disease outbreak, it is important to comprehend its transmission dynamics and to estimate the changes in transmission over time and space. This can provide better understanding about the epidemiological situation and determine whether social control measures implemented are being effective or not. Such type of analysis can forecast about potential future growth and guide us to design suitable control interventions. To date, many researchers around the world have focused their interests on understanding the transmission dynamics of COVID–19 disease using mathematical and statistical models, see e.g., [3, 4, 12, 9, 10, 13].

In this study, we focus on the implications of two simple empirical functions and probabilistic models, namely, Gompertz and Hill functions, to estimate the future trends of COVID–19 disease in the Guayas province in Ecuador. We estimate

the unknown parameters using public data of diagnosed cases, and compare the results with dataset from February 29 to December 9 (2020). We also attempt to forecast the evolution of number of patients until 800 days after first case. Finally, we compare the precision and the results obtained with our proposed models in order to make some useful recommendations on public health issues.
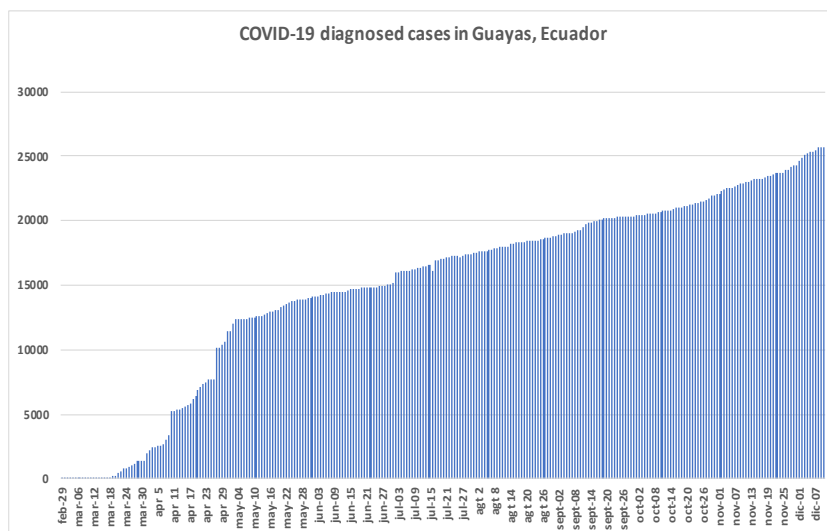

FIGURE 1. Diagnosed cases of COVID–19 in Guayas province, Ecuador from February 29 to December 9, 2020.

2. Methods

Gompertz and Hill functions are two empirical mathematical models highly reliable to provide robust estimations of total cases, active cases, deaths, peak and stabilization times in epidemic dynamics. During the COVID–19 pandemic, these models have showed very accurate estimates of cumulative cases in countries such as Italy, Spain, Austria, Switzerland, Norway, South Korea, and other countries that are already reaching the stabilization stage of the COVID-19 epidemic curve [5, 11]. In this work, these models are used for the following two purposes: (a) to fit models to data of COVID–19 diagnosed cases provided by the Ministerio de Salud Pública of Ecuador for Guayas province from February 29 to December 9, 2020 and (b) to predict its future behaviour at medium–term assuming that the quarantine and confinement measures are kept during almost all 2020. Here, it is important highlight that data are provided with variable error (See Figure 2), whose sources could be several and from different nature, for instance, delays in testing, delays in results, and low and unstable daily sampling rates, among others. We simulated these errors for the model predictions by including a probabilistic error term using a Gamma distribution. Gamma distribution was chosen among others (`normal`, `lognormal`, `Poisson`, etc.) because it is only positive and allow high skewness as those observed in data from infectious cases. Furthermore, it captures the

increasing error variance in observed data as the epidemic advances. This approach allowed us to consider a reasonable measure of uncertainty in predictions, based on the uncertainty of the reported data.

Let N representing the predicted number of diagnosed cases in a moment t. Thus, the formulation of the Gompertz and Hill models is as follows

$$N(t) = \alpha e^{-\beta e^{-Rt}} + \gamma, \qquad \textbf{Gompertz model}$$

$$N(t) = d + \frac{\alpha - d}{1 + \left(\frac{t}{c}\right)^R} + \gamma, \quad \textbf{Hill model}.$$

(2.1)

In above formulation, α represents the maximum number of diagnosed cases at the end of the epidemic, β and c represent parameters controlling the horizontal displacement (time to stabilization) of the curve, R is the epidemic growth rate of the population , d is a correction factor and γ is an error term from a Gamma distribution. The probability density function of the Gamma distribution was estimated to build the error terms

$$\gamma(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)},$$

where k and λ are parameters greater than zero and Γ is the gamma function. Non–linear least square method was performed trough the Gauss–Newton algorithm with 5000 iterations to estimate the values of the parameters involved in the models. The Coefficient of Determination (R^2), the Akaike Information Criteria (AIC) and the Root Mean Squared Error (RMSE) were used as the metrics to evaluate the models performance (See Table 1).

Parameters λ and k were set to 0.1 and 0.001 respectively and 10,000 random simulations of gamma distributions with these fixed parameters were ran to build a distribution of 10,000 error terms (γ) for each model. This was performed both for the observed data (Feb 29 to December 9, 2020) as well as for the prediction time window (800 days after the first case). Gompertz and Hill models were fitted with every different γ to produce a distribution of predictions aimed to capture uncertainties in the future behaviour of cumulative diagnosed cases, as much as possible. Therefore, final predictions of COVID–19 diagnosed cases 800 days after the first case are presented with a region of uncertainty according with these errors (See Figure 2).

3. Results

The results of the modelling procedures are shown in Table 1 and Figure 2. The fit of both models to the observed data was good (95%–97%), and their performances were similar during the first 100 days after the first case, whereas after it, the predicted cases with the Gompertz model are greater with the Hill model. This suggests that for the first stage of the epidemic both models could be reliable in their predictions, agreeing with Yang et al [13]. The Gompertz model predicted a minimum of 18,000 and a maximum of 30,000 cases in Guayas province at the end of 2021, and the stabilization of the curve around days 300–400 after the first case (December, 2020). The Hill model predicted a minimum of 16,000 and

a maximum of 28,000 cases by the end of 2021 and the stabilization around days 400–500 after the first case (February, 2021). The predictions of the Gompertz model probably underestimate the potential number of cases by December, 2020. We must stress the fact that these results could dramatically change if increased social interaction (even if small) led to an increase of parameter R in the models.

Parameters and model performance metrics	Gompertz	Hill
α	$2.24 \times 10^4 \pm 235.9$	$-2.08 \times 10^3 \pm 416.8$
β	$2.94 \times 10^4 \pm 416.8$	—
R	$1.84 \times 10^{-2} \pm 0.69 \times 10^3$	1.32 ± 0.077
c	—	$1.001 \times 10^2 \pm 4.93$
d	—	$2.94 \times 10^4 \pm 1.07 \times 10^3$
R^2	95%	97%
AIC	5017.95	4839.9
RMSE	1540.2	1124.4

TABLE 1. Parameters used for the Gompertz and Hill models. These parameters were estimated by non–linear least squares from the observed data of cumulative COVID–19 cases in Guayas province (Ecuador) from February 29 to December 9 (2020) and used for future predictions. Parameter values are shown $\pm$ standard errors.

4. Conclusions

In this study, we used two empirical functions to assess the predictive accuracy of COVID–19 diagnosed cases in Guayas province (Ecuador) since February 29, 2020. By comparing the models outcomes with the diagnosed cases, it has been observed that our estimated values have good correspondence with the diagnosed cases. If the current pattern is going on then according to our estimate, following the Hill function there will be a maximum of 28,000 positive cases by the end of 2021, and the stabilization of the cases will be achieved approximately on February (2021). Following the Gompertz function there will be a maximum of 30,000 cases by the end of 2021, and the stabilization will be achieved approximately on December (2020). Taking into account the previous scenario, it is recommended to the government and health authorities of Guayas province, to maintain isolation and quarantine measures, at least until the stabilization of the epidemic curve has been achieved, that is, at least until February, 2021.

Although these models are simple, they are very helpful in designing control measures; however, the results of this work are conditioned to the quality and truthfulness of the epidemic data from the province. The initial conditions and the parameter regions may also affect the results significantly.
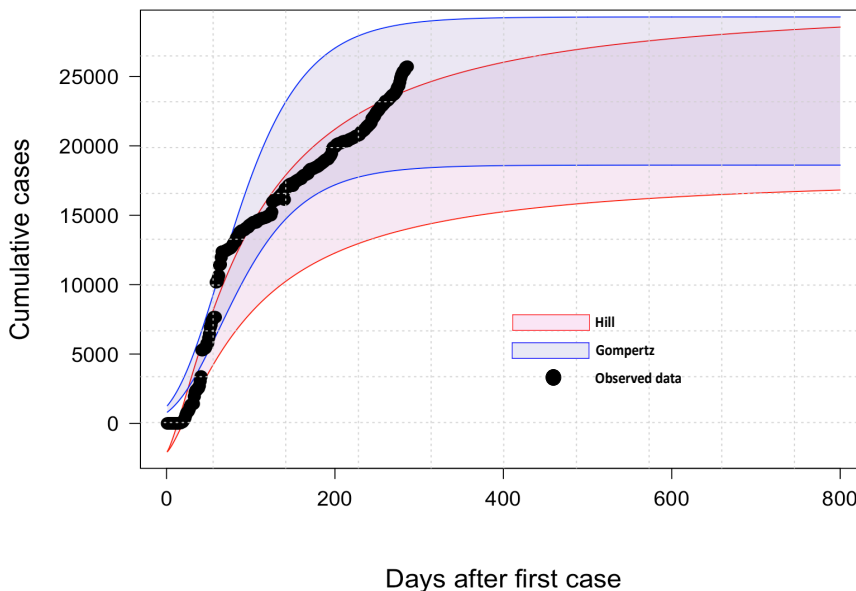
FIGURE 2. Results of the Gompertz and Hill models for predicting COVID–19 cases in Guayas province (Ecuador). Shaded areas represent the uncertainty in predictions according to Gamma error terms. The vertical black dotted lines represent the stabilization for each model.
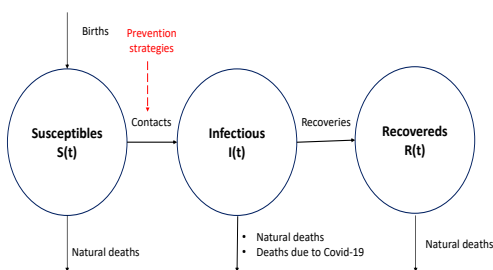

FIGURE 3. Compartmental SIR–type model for COVID–19 disease.

Since the epidemic process is compartmental (see e.g., Figure 3), in addition to our two functions, other epidemic models such as differential equations and time–delayed equations should be considered in future researches to cope with problems such like control strategies, the low testing rates which may cause underestimation

of cases and the delays from infection to diagnose (10–15 days in average). For instance, it would be interesting to contrast the predictions of this research with the predictions given by a ODEs system with quarantine (ν) as a control strategy such as

$$
\begin{cases}
\frac{dS(t)}{dt} = \Lambda - \beta \frac{I(t)S(t)}{a+\nu I(t)} - \mu S(t) \\[2mm]
\frac{dI(t)}{dt} = \beta \frac{I(t)S(t)}{a+\nu I(t)} - (\alpha + \mu + \delta)I(t) \\[2mm]
\frac{dR(t)}{dt} = \alpha I(t) - \mu R(t) \\[2mm]
(S(0), I(0), R(0)) = (S_0, I_0, R_0),
\end{cases}
$$

where $N(t)$ represents the total population of Guayas at time t, which is partitioned into susceptible (S), infectious (I) and recovered (R) individuals at time t, and $(S(0), I(0), R(0)) = (S_0, I_0, R_0)$ is the initial state of population just as the first infected individual enters to the country (February 29, 2020), and other parameters involved in the model (Λ , μ, α, β, δ, a and ν) represent rates as born, death, recovery, transmission among others.

References

1. *Johns Hopkins Coronavirus Resource Center*, https://coronavirus.jhu.edu/, 2020, Acces 05-10-2020.
2. *Worldometer Coronavirus Updates*, https://www.worldometers.info/, 2020, Acces 05-10-2020.
3. Tian-Mu Chen, Jia Rui, Qiu-Peng Wang, Ze-Yu Zhao, Jing-An Cui, and Ling Yin, *A mathematical model for simulating the phase–based transmissibility of a novel coronavirus*, Infectious Diseases of Poverty **9** (2020), no. 1, 1–8.
4. Sunhwa Choi and Moran Ki, *Estimating the reproductive number and the outbreak size of novel coronavirus disease (COVID–19) using mathematical model in republic of korea*, Epidemiology and Health (2020), 123–145.
5. John L Haybittle, *The use of the gompertz function to relate changes in life expectancy to the standardized mortality ratio*, International journal of epidemiology **27** (1998), no. 5, 885–889.
6. Laura Díez Izquierdo et al., *Informe técnico nuevo coronavirus 2019–ncov*, Ph.D. thesis, Instituto de Salud Carlos III, 2020.
7. Catharine I Paules, Hilary D Marston, and Anthony S Fauci, *Coronavirus infections more than just the common cold*, Jama **323** (2020), no. 8, 707–708.
8. LJ Saif, *Animal coronavirus vaccines: lessons for SARS.*, Developments in biologicals **119** (2004), 129–140.
9. Pedro V Savi, Marcelo A Savi, and Beatriz Borges, *A mathematical description of the dynamics of coronavirus disease (COVID–10): A case study of Brazil*, arXiv preprint arXiv:2004.03495 (2020).
10. Amjad S Shaikh, Iqbal N Shaikh, and Kottakkaran Sooppy Nisar, *A mathematical model of COVID–19 using fractional derivative: Outbreak in india with dynamics of transmission and control*, Preprints (2020).

FELIPE BENAVIDES AND JHOANA P. ROMERO–LEITON*

11. Kathleen MC Tjørve and Even Tjørve, *The use of gompertz models in growth analyses, and new gompertz–model approach: An addition to the unified–richards family*, PloS one **12** (2017), no. 6.
12. Chayu Yang and Jin Wang, *A mathematical model for the novel coronavirus epidemic in wuhan, china*, Mathematical Biosciences and Engineering **17** (2020), no. 3, 2708–2724.
13. Wuyue Yang, Dongyan Zhang, Liangrong Peng, Changjing Zhuge, and Liu Hong, *Rational evaluation of various epidemic models based on the covid–19 data of china*, arXiv preprint arXiv:2003.05666 (2020).

Instituto de Estudios del Pacífico, Universidad Nacional de Colombia, Colombia
Email address: pipeben@gmail.com

Facultad de Ingeniería, Universidad Cesmag, Colombia
Email address: jpatirom3@gmail.com

150