

Article

Ensemble Algorithms to Improve COVID-19 Growth Curve Estimates

Raydonal Ospina ^{1,2,*} , Jaciele Oliveira ², Cristiano Ferraz ² , André Leite ² and João Gondim ³ ¹ Statistics Department, LInCa, Federal University of Bahia, Salvador 40170-110, Brazil² Statistics Department, CASTLab, Federal University of Pernambuco, Recife 50670-901, Brazil; jaciele.jesus@ufpe.br (J.O.); cferraz@castlab.org (C.F.); leite@castlab.org (A.L.)³ Mathematics Department, Federal University of Pernambuco, Recife 50670-901, Brazil; joao.gondim@ufpe.br

* Correspondence: raydonalmartinez@ufba.br

Abstract: In January 2020, the world was taken by surprise as a novel disease, COVID-19, emerged, attributed to the new SARS-CoV-2 virus. Initial cases were reported in China, and the virus rapidly disseminated globally, leading the World Health Organization (WHO) to declare it a pandemic on 11 March 2020. Given the novelty of this pathogen, limited information was available regarding its infection rate and symptoms. Consequently, the necessity of employing mathematical models to enable researchers to describe the progression of the epidemic and make accurate forecasts became evident. This study focuses on the analysis of several dynamic growth models, including the logistics, Gompertz, and Richards growth models, which are commonly employed to depict the spread of infectious diseases. These models are integrated to harness their predictive capabilities, utilizing an ensemble modeling approach. The resulting ensemble algorithm was trained using COVID-19 data from the Brazilian state of Paraíba. The proposed ensemble model approach effectively reduced forecasting errors, showcasing itself as a promising methodology for estimating COVID-19 growth curves, improving data forecasting accuracy, and providing rapid responses in the early stages of the pandemic.

Keywords: infectious diseases; growth rate; forecasting; ensemble modeling



Citation: Ospina, R.; Oliveira, J.; Ferraz, C.; Leite, A.; Gondim, J. Ensemble Algorithms to Improve COVID-19 Growth Curve Estimates. *Stats* **2023**, *6*, 990–1007. <https://doi.org/10.3390/stats6040062>

Academic Editor: Wei Zhu

Received: 28 August 2023

Revised: 22 September 2023

Accepted: 26 September 2023

Published: 29 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The World Health Organization (WHO) defines a pandemic as an epidemic that occurs worldwide, or over a very large area, affecting many people across different countries [1]. Throughout history, humanity has been confronted with numerous pandemics. The so-called “Black Death” afflicted Europe from 1348 to 1351, resulting in the mortality of approximately one-third of the continent’s population [2]. In the 19th century, tuberculosis, an infection caused by the bacterium *Mycobacterium tuberculosis* (MTB), accounted for nearly 25% of all deaths [3]. Another significant example was the Spanish Flu pandemic, also known as the 1918 Flu, which caused the deaths of approximately 100 million people and infected 3% to 5% of the world population [4,5].

In January 2020, the world was caught off guard by a novel and highly transmissible disease, subsequently named COVID-19, attributable to the SARS-CoV-2 virus. This epidemic originated in China, specifically in the city of Wuhan, toward the end of 2019, and swiftly disseminated to other nations, formally attaining pandemic status within a few months [6].

Scientists have determined that the novel virus has a zoonotic origin in bats, which is recognized as a significant viral reservoir. The genetic sequence of SARS-CoV-2 exhibits a 96% similarity to the genetic sequences of other coronaviruses found in bats in China [7]. Notably, this virus exhibits higher lethality and transmissibility compared to other respiratory infections. It can be transmitted through airborne particles, such as respiratory

secretions (cough and saliva), close contact with infected individuals, and the contamination of personal items [8]. Furthermore, it possesses the capability to persist on surfaces for extended periods [9].

In Brazil, the first case of the new coronavirus was officially reported by the Ministry of Health on 26 February 2020, in the state of São Paulo. The patient was a 61-year-old individual with a history of recent travel to Italy [10]. On 12 March, which was fifteen days after the confirmation of the first case in the country, the first fatality attributed to the disease occurred. The deceased was a 57-year-old woman who had been hospitalized with symptoms of COVID-19 one day prior to her passing [11]. Since that time, the virus has rapidly disseminated throughout the country, leading to a significant increase in fatal cases.

In response to this situation, governors expressed significant concern with the objective of preventing a substantial portion of the population from becoming infected simultaneously. This concern stemmed from the potential to overwhelm the public health system, which could lead to a potential increase in the mortality rate due to the infection. To mitigate the spread of SARS-CoV-2, various measures were implemented, including stringent social distancing restrictions and, in many countries worldwide, the imposition of strict lockdowns.

Since SARS-CoV-2 was a novel pathogen, there was a lack of knowledge regarding its behavior, causing fear and concern among citizens worldwide when the pandemic was declared. Consequently, there was an urgent need to employ tools capable of describing the trajectory of the epidemic, assessing the impact of restrictive measures, forecasting potential virus spread scenarios, and ultimately assisting governments in formulating effective policies to combat COVID-19.

In this context, numerous studies utilizing epidemiological models have been undertaken to comprehend and depict the spread of the virus. These studies involve the estimation of critical epidemiological parameters, including disease transmission rates and the basic reproductive number. For instance, in the study conducted by Ospina et al. (2022) [12], data-driven analytical tools were employed to discern shifts in the trends of COVID-19 cases and calculate the effective reproductive numbers. Furthermore, several other research efforts have primarily focused on growth models. These models center on the examination of the accumulation of infected cases over a defined time frame and seek to estimate the associated growth rates.

When investigating epidemics, it is important to make predictions to better assist authorities in decision-making, but these predictions are subject to errors. For example, in [13], the authors examine the accuracy of autoregressive integrated moving average (ARIMA) models, emphasizing their potential for short-term forecasting, even though they are not best suited for long-term predictions. Hence, it would be ideal to find a model that controls this uncertainty the best possible way. Ensemble models are pointed out in the literature as an efficient approach in this regard and, according to [14–16], these models allow for an easier determination of a curve that best fits the observed data.

In this study, we initially applied the logistic, Gompertz, and Richards growth models to the data. Nevertheless, in pursuit of enhancing forecast accuracy, we employed ensemble models with a bootstrap approach. This method involves the combination of individual models, thereby integrating predictive precision among them, ultimately providing better control over forecast errors.

The novelty of this research lies in several aspects. Firstly, it employs a comprehensive analysis of COVID-19 cumulative deaths in the State of Paraíba, Brazil, during a critical period, offering insights into the pandemic's dynamics in a specific regional context. Secondly, the study introduces an ensemble modeling approach, which combines multiple growth models to enhance prediction accuracy, providing a novel solution to the challenges of forecasting the pandemic. This ensemble method's application in epidemiological modeling is innovative and can be adapted to different infectious diseases.

2. Background

The origins of the utilization of mathematical models to describe the spread of infectious diseases can be traced back to the 18th century when mathematician Daniel Bernoulli employed differential equation models to investigate the smallpox epidemic that afflicted Europe during that era [17]. Since then, numerous models for various infectious diseases have been researched and put into practice.

In 1906, with the objective of comprehending the recurring patterns in the measles epidemics, Hamer formulated the first model that took into consideration factors such as the numbers of susceptible and infected individuals, as well as the contact rate between them, in relation to the incidence rate [18]. In 1915, while investigating the incidence of malaria, Ronald Ross highlighted the existence of a limiting value in mosquito density below which malaria would naturally extinguish itself [19]. Ross's hypothesis may have foreshadowed the threshold theorem developed by Kermack and McKendrick in 1926 [20], which denotes a critical density of individuals below which the entry of newly infectious individuals is insufficient to sustain an epidemic [21,22].

In 1927, Kermack and McKendrick formulated the initial compartmental epidemiological model, categorizing the population into distinct classes and employing differential equations, known as the SIR model (susceptible-infected-removed/recovered). From the SIR model, additional models emerged, including the SEIR model (susceptible-exposed-infected-removed/recovered) and SIRD model (susceptible-infected-removed/recovered-deceased) [23].

Growth and Ensemble Models in the Context of Infectious Diseases

During the 18th century, the world underwent profound transformations brought about by the Industrial Revolution, including the rapid population growth and urbanization. It was in this context that the Malthusian theory emerged [24]. Formulated by Malthus in 1798, this theory proposed that the population was growing more rapidly than food production, leading to concerns about a global famine. Growth models, also known as population dynamics models, analyze the rates of change in the quantities of individuals within a specific population over time [25].

Growth models are applied in various studies to model growth curves, such as the dynamics of dengue fever and tuberculosis [26], the description of the growth of citrus black spot disease [27], the characterization of prostate tumor growth [28,29], and the investigation of growth curves in animals [30–32].

Numerous recent examples of these models in the literature have been applied to assess the impact of COVID-19. In [33], a predictive analysis of the number of confirmed COVID-19 cases in Brazil and eight other countries was conducted using the Gompertz growth model. Similarly, in [34], the authors utilized the Gompertz growth model to forecast the maximum numbers of COVID-19 cases and deaths. A bi-logistic model was employed in [35] to depict the temporal trends of COVID-19 among indigenous populations in the Brazilian states of Amapá and Pará. This model demonstrated statistical significance and identified 12 May and 22 July as the dates when the disease decelerated in this population. Lastly, References [36,37] employed a Richards generalized growth model to analyze the COVID-19 epidemic curves in the cities of Recife and Teresina, Brazil.

Ensemble models are commonly found in the literature across various research fields for data analysis. For instance, in [38], ensemble models were utilized to predict wind power production, effectively addressing issues of overestimation that were present in individual models, and achieving favorable results. In [39], the authors proposed an ensemble model for predicting electrical demand across the four Brazilian sub-systems. A Bayesian ensemble of models was employed in [40] to generate predictions for death rates and life expectancy.

In the realm of medical research, Reference [41] employed five machine learning models and an ensemble of these models to evaluate the performance of traditional scores in the European System for Cardiac Operative Risk Evaluation. The study concluded that the ensemble model exhibited improved accuracy, enhancing decision curve analysis by 1–6%.

In epidemiological studies, Reference [42] adopted a Bayesian ensemble approach to forecast epidemiological curves. Similarly, in [43], three distinct prediction systems for dengue fever outbreaks in San Juan, Puerto Rico, were developed, and an ensemble of these predictions was created using Bayesian averaging methods. This research demonstrated that the combined predictions yielded greater precision compared to those generated by individual approaches.

3. Growth Models and Ensemble Algorithm

Nonlinear growth models are employed to estimate growth rates and have broad applications in various fields, including economics, animal nutrition, the study of infectious diseases, among others. Unlike the SIR compartmental model, growth models rely on the cumulative number of infected cases, which encompasses the sum of the infected and recovered compartments of the SIR model. These models are applied to analyze population growth, specifically to investigate the behavior of S-shaped cumulative curves.

The ensemble models of [44,45] have excelled due to their robustness in prediction and forecasting processes [46,47]. This approach combines the advantages of many models instead of choosing the best model according to some selection criterion [44]. One of the advantages is the reduction in prediction and forecasting errors [48]. In [16], the authors presented an ensemble model based on bootstrapping that aims to improve precision performance by systematically integrating the predictive precision of each model. This methodology is employed to forecast the evolution of a dynamical growth process defined by a system of nonlinear differential equations, producing more accurate solutions.

The analysis of nonlinear models depends on an iterative process to find solutions to equations because, unlike the linear case, it is generally not possible to find them analytically. The iterative process begins with initial parameter values and calculates the residual sum of squares (RSS) based on these values. The parameters are continuously adjusted until the RSS is minimized.

3.1. Gompertz Model

To describe the growth of solid tumors, mathematician Benjamin Gompertz developed an equation in 1938, now known as the Gompertz equation [49,50]. Gompertz observed that, in his model, the growth rate is higher in the earlier stages of the process and rapidly transitions to slower growth. This model is widely applied to describe the general growth of cells, including plants, bacteria, and tumors [51]. The Gompertz equation is expressed as follows:

$$\frac{dC}{dt} = \gamma \ln\left(\frac{K}{C}\right), \quad (1)$$

Here, C represents the cumulative total of cases, K denotes the maximum number of cumulative cases or the final size of the epidemic, and γ is the intrinsic per capita growth rate of the infected population.

After solving this ordinary differential equation (ODE), one obtains

$$C(t) = Ke^{-e^{-\gamma t} \ln \frac{K}{C_0}}, \quad (2)$$

where $C(t)$ represents the quantity of cumulative cases at time t . In this model, the growth is typically smaller in the early and later stages of the outbreak [52].

3.2. Exponential Model

The exponential model, developed by Thomas Robert Malthus in 1798 [53], assumes that the rate of change of a quantity C at time t is directly proportional to C . The exponential model is described by

$$\frac{dC}{dt} = \frac{\gamma}{N}C,$$

where γ represents the exponential growth rate, N is the population size, and $C(t)$ represents the cumulative number of cases at time t . Equation (3) provides the solution to this ordinary differential equation.

$$C(t) = C_0 e^{\frac{\gamma}{N}t}, \quad (3)$$

where $C_0 = C(0)$ is the initial number of cases.

3.3. Logistic Model

Mathematician Pierre F. Verhulst proposed a model in 1837 that presumed a population could grow until it reaches its maximum limit, at which point it stabilizes. In this model, the population's effective growth rate varies with time [54]. This model serves as an alternative to the exponential growth model, where the growth rate remains constant, and there are no constraints on population growth [55]. The logistic model is described by the following differential equation:

$$\frac{dC}{dt} = \gamma C \left(1 - \frac{C}{K}\right), \quad (4)$$

where γ , C , and K have the same interpretations presented in the Gompertz model. The solution of this ODE is given by Equation (5) below

$$C(t) = \frac{K}{1 + \left(\frac{K}{C_0} - 1\right)e^{-\gamma t}}. \quad (5)$$

3.4. Richards Model

The Richards model [56] extends the logistic model by introducing a third parameter, α , which quantifies the deviation from the growth curve. Proposed by Richards in 1959, this model was initially developed to describe the growth of fish populations and represents a generalization of the von Bertalanffy model [57]. The Richards equation is expressed as the following differential equation:

$$\frac{dC}{dt} = \gamma C \left(1 - \frac{C}{K}\right)^{\frac{1}{\alpha}}. \quad (6)$$

After solving this ODE, we obtain the Richards model, as shown in Equation (7):

$$C(t) = K \left(1 - e^{\alpha \gamma t} \left(1 - \left(\frac{C_0}{K}\right)^{-\alpha}\right)\right)^{-\frac{1}{\alpha}}. \quad (7)$$

Here, K represents the final size of the epidemic, γ is the growth rate, C_0 is the number of cases at the onset of the epidemic, and α is the shape parameter that governs the curvature of the curve. When $\alpha = 1$, Equation (6) reduces to Verhulst's logistic growth model [54] given in (4).

The introduction of the shape parameter in this model provides greater flexibility in selecting the curve's shape. The model assumes that the daily incidence curve exhibits a unique peak of high incidence, which corresponds to the inflection point of the epidemic, marking the transition from increasing to decreasing accumulation rates or vice versa. These inflection points can be determined by observing when the epidemic curve begins to decline [58]. The inflection point N_{inf} for this model is a function of α and K and is given by

$$N_{\text{inf}} = \left(\frac{1}{1 + \alpha}\right)^{\frac{1}{\alpha}} K.$$

This quantity holds significant relevance in epidemiology, as it indicates the beginning or end of a phase, representing the moment of acceleration after deceleration or vice versa [58].

3.5. Performance Metrics

The performance of a particular model can be evaluated using various metrics, including the adjusted coefficient of determination R^2 , the mean square error (MSE), and the absolute square error (ASE). These performance criteria all share the common characteristic of considering the model's residuals, which indicate how closely the fitted results align with the data.

The determination coefficient R^2 , also known as the square of Pearson's correlation coefficient, is a widely used performance metric in the literature for assessing the quality of a model's fit to the data. This coefficient ranges from 0 to 1, and the closer it is to 1, the better the fit. This implies that the model can effectively explain most of the response variables [59,60]. The calculation of R^2 requires the residual sum of squares (RSS) and the total sum of squares (TSS) as inputs, which are defined as

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

and

$$\text{TSS} = \sum_{i=1}^n (y_i - \bar{y})^2,$$

respectively. Here, n is the number of observations, y_i represents the i -th observed value, $\hat{y}_i$ is the i -th fitted value, and $\bar{y}$ is the mean of all the observations. The adjusted coefficient of determination is then calculated as

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}.$$

The mean absolute error (MAE) is calculated as the average of the absolute differences between the actual parameters and their estimated values. Similarly, the mean square error (MSE) is determined as the average of the squared differences between these values, as expressed in Equations (8) and (9), respectively.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

The IFMS (interval forecast mean score) assesses the width of the forecast interval by taking into account forecast uncertainty. This is different from metrics such as MAE, MSE, and R^2 , which primarily focus on the discrepancies between the model and the data [61]. The IFMS is calculated as follows:

$$\text{IFMS} = \frac{1}{h} \sum_{i=1}^h \left[(U_{t_i} - L_{t_i}) + \frac{2}{0,05} (L_{t_i} - y_{t_i}) I\{y_{t_i} < L_{t_i}\} \frac{2}{0,05} (y_{t_i} - U_{t_i}) I\{y_{t_i} > U_{t_i}\} \right],$$

where L_{t_i} and U_{t_i} are, respectively, the lower and upper bounds of the forecast interval at time t at 95% confidence and $I\{\cdot\}$ is an indicator function.

3.6. Ensemble Method

The ensemble approach combines the strengths of multiple models through a weighted average, essentially creating a linear combination of nonlinear models. Numerous methods for constructing ensemble models exist in the literature, including neural networks, Bayesian averaging, among others [62]. However, the method described here is based on the weighted combination of individual models, as proposed in [16].

Indeed, we can consider a set of I parametric models, such as

$$I = \{\text{Gompertz, Exponential, Logistic, Richards}\} = \{f_1(t, \theta_1), f_2(t, \theta_2), f_3(t, \theta_3), f_4(t, \theta_4)\}, \quad (10)$$

where θ_i represents the parameters that describe the i -th model. Using the training dataset, the parameter set, and the average ensemble incidence curve for each model i , estimated for $i = 1, \dots, I$, we calculate the weight w_i of each model based on the quality of its fit. The quality is assessed using metrics like the mean square error (MSE) or other criteria such as the AIC. In this work, we use MSE to evaluate the quality of the fit. Therefore, the weight for each model is computed as follows:

$$w_i = \frac{\frac{1}{\text{MSE}_i}}{\frac{1}{\text{MSE}_1} + \frac{1}{\text{MSE}_2} + \dots + \frac{1}{\text{MSE}_I}}, \quad i = 1, \dots, I,$$

where

$$\text{MSE}_i = \frac{1}{n} \sum_{j=1}^n (f_i(t_j, \hat{\theta}_i) - y_{t_j})^2,$$

with the constraint that $\sum_{i=1}^I w_i = 1$, ensuring a convex linear combination of models. If $f_i(t_j, \hat{\theta}_i)$ represents the fitted curve by the i -th model, the *average incidence curve* of the ensemble model is given by

$$f_{\text{ens}}(t) = \sum_{i=1}^I w_i f_i(t, \hat{\theta}_i).$$

In the context of this work, it can be assumed that the observed data (cases) follow a probabilistic structure, which adheres to a Poisson distribution [16] with a mean of $f_{\text{ens}}(t)$. To obtain a 95% confidence interval (or forecast interval) for the incidence curve at time t , the parametric bootstrap method can be employed. To do this, consider that the training sample consists of n data points: $t_1, t_2, \dots, t_n$. A bootstrap sample is created by generating a random variable y_i from the Poisson distribution with a mean of $f_{\text{ens}}(t_j)$ for each data point t_j , where $j = 1, 2, \dots, n$:

$$y_j \sim \text{Poisson}(f_{\text{ens}}(t_j)), \quad j = 1, 2, \dots, n.$$

Therefore, $\{y_1, y_2, \dots, y_n\}$ forms a bootstrap sample. This sample is then used to refit each of the I models, calculate weights for each refitted model, estimate parameters, and generate forecasts for the ensemble model. By repeating this process B times, it becomes possible to construct a 95% confidence interval (or forecast interval) based on the 2.5th and 97.5th percentiles.

As an example, consider four individual models given in (10) from which we will build the ensemble model. Assume $n = 100$, which corresponds to 100 time points. Here is the step-by-step process:

1. Fit each of the four models to the original series and estimate the parameters;
2. Calculate the MSE of each model and find the corresponding weight w_i based on the MSE;
3. Find the ensemble average incidence curve

$$f_{\text{ens}}(t) = \sum_{i=1}^4 w_i f_i(t, \hat{\theta}_i);$$

4. Assume that the data follow a Poisson distribution with mean $f_{\text{ens}}(t)$, to build a 95% confidence interval (or forecast interval) for the incidence curve at time t using the parametric bootstrap method;
5. Generate a random variable y_j for the incidence at each point t_j , $j = 1, \dots, 100$ via the Poisson distribution with mean $f_{\text{ens}}(t_j)$, i.e.,

$$y_j \sim \text{Poisson}(f_{\text{ens}}(t_j)), \quad j = 1, \dots, 100;$$

6. Repeat the process described in the previous step B times to generate B bootstrap replicas and construct the confidence interval;
7. Refit the I growth models for each replica, calculate the respective MSEs and weights of the refit models, and construct the prediction and forecast intervals for each one;
8. Obtain B ensemble mean incidence curves using the process described in the previous item, calculate the MSE, and build confidence intervals for each of these mean curves.

4. Results

The exponential, Gompertz, logistic, and Richards models were fitted to the COVID-19 data from the State of Paraíba, Brazil, to study the disease's growth rates in the State. Subsequently, an ensemble model was constructed using the results from these individual models to produce forecasts ranging from 15 to 30 days ahead. Confidence intervals were also established for each forecasting approach.

The first confirmed COVID-19 case in the State of Paraíba was reported on 18 March 2020. This case involved a 60-year-old man from João Pessoa, who had returned from a trip to Europe on 29 February. Following this initial case, the virus began to spread throughout the State. In response, the Government of Paraíba declared a state of emergency to prevent and combat the pandemic.

Given that the under-reporting of deaths due to COVID-19 is typically less severe than the under-reporting of cases (as case reporting often depends on testing availability), this study utilizes the cumulative death curve attributed to the disease in the State of Paraíba. The first COVID-19-related death in Paraíba was recorded on 31 March 2020, exactly 14 days after the first confirmed case in the State. The deceased individual was a 36-year-old man with diabetes residing in the city of Patos, located in the Sertão region of the State. This man exhibited initial symptoms on 25 March, just six days prior to his passing.

The scope of the pandemic period analyzed in this study encompasses the year 2020, as 2021 was marked by a second wave of the outbreak. Figure 1 illustrates the daily number of COVID-19 deaths in Paraíba, commencing from the first recorded death until 31 December 2020. Notably, the highest death counts occurred on 25 May and 5 June.

Following this, four growth models were fitted, and an ensemble model was constructed to analyze the death curve in Paraíba and the associated growth rates. The model fitting process considered data spanning from 31 March to 16 December, encompassing a total of 261 days, with the last 15 days of the year reserved for forecasting 15 days ahead. The selected individual models included the exponential, logistic, Gompertz, and Richards growth models.

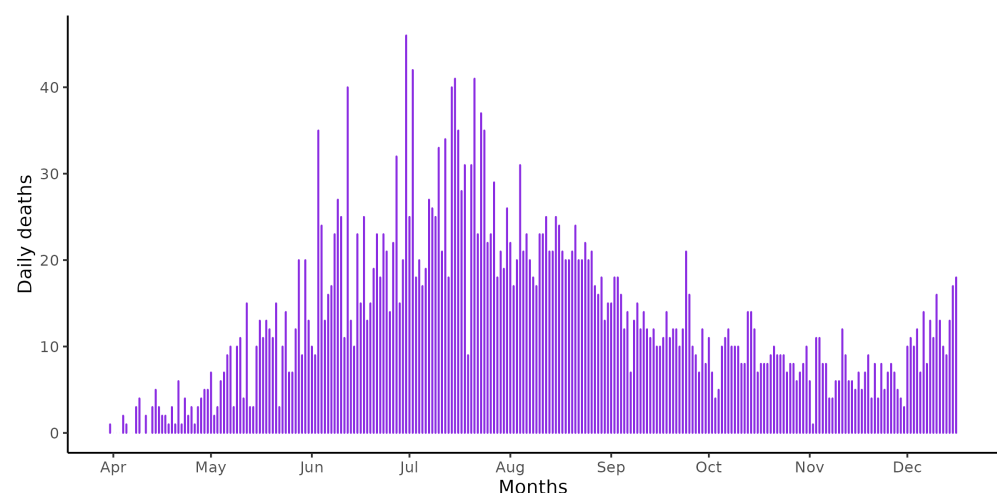


Figure 1. Epidemic curve for Paraíba, Brazil, with temporal evolution of recorded daily COVID-19 deaths during epidemic up to 31 to December 2020 (First wave).

Figure 2 visually compares the observed cumulative data curve (in orange) with the fitted curve generated by the exponential model (in red). It is evident that the exponential model does not align well with the data, as its curve exhibits a notably different behavior. This discrepancy is likely due to the fact that genuine exponential growth is unattainable in reality, as it would result in unbounded growth, while the total population is inherently limited.

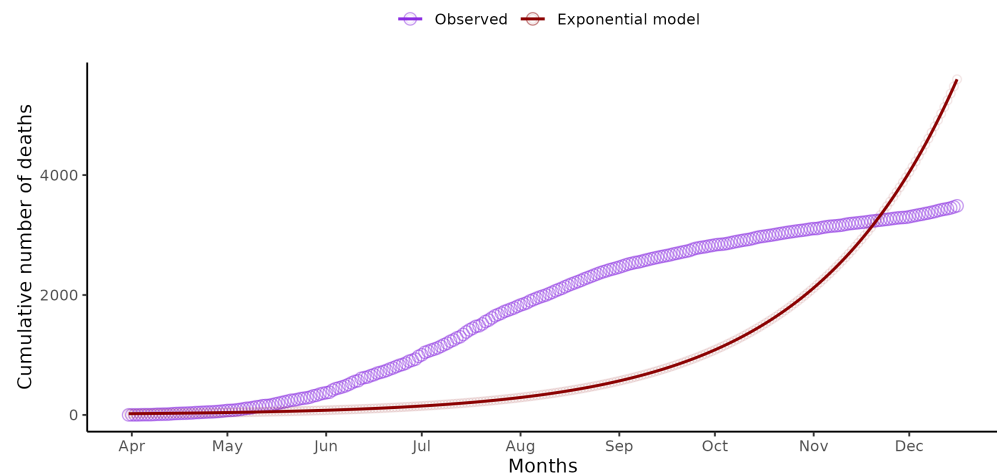


Figure 2. Daily cumulative incidence of death counts and exponential model fit.

Figure 3 presents the fitted curves generated by the logistic, Gompertz, and Richards models, alongside the cumulative death count reported by the Health Ministry. Additionally, 15-day forecasts were produced for each of these models, with a vertical dashed line indicating the point from which the forecasting starts, relative to 17 December 2020. Upon observing the behavior of these curves, it becomes evident that both the Gompertz and Richards models offer a better fit to the data compared to the logistic model. However, when it comes to forecasting, all three models consistently underestimate the observed death curve. These plots highlight a significant change in the death curve's behavior just prior to 17 December, making accurate forecasting challenging.

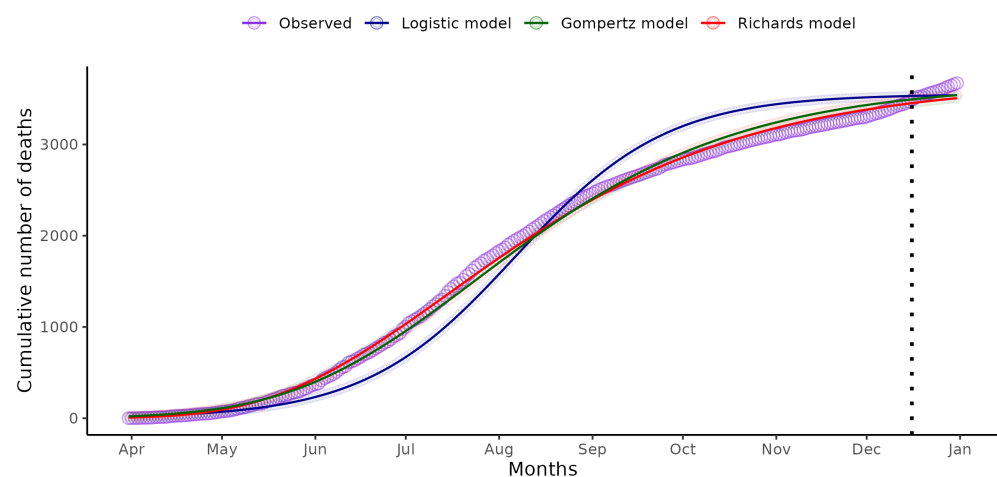


Figure 3. Daily cumulative incidence of deaths counts and logistics, Gompertz, and Richards model fit.

From the fitted growth models, the ensemble model was constructed using the logistic, Gompertz, and Richards models (excluding the exponential model due to its drastically different behavior compared to the data). To create the ensemble model, we calculated the weighted average curve for the growth models based on the mean square error. Next, we generated one thousand (1000) bootstrap replicates using the ensemble model, assuming a

Poisson distribution as the counts structure for the weighted average. For each of these replicates, we performed the following steps:

1. Reconstructed the 95% confidence intervals.
2. Refitted the logistic, Gompertz, and Richards models.
3. Built an ensemble model for each replica.
4. Calculated a new ensemble average curve using the models refitted to the replicas.

This comprehensive process allowed us to generate a robust ensemble model and assess its performance under various conditions and uncertainties.

Table 1 provides the estimated parameters for each growth model and the ensemble model. These parameters include the final size of the pandemic (K), the growth rate (γ), the shape parameter (α), and the corresponding standard errors. Additionally, the table displays the weights assigned to each model based on the mean square error, as explained in Section 3. The Gompertz model carries the highest weight in the ensemble model due to its lower mean square error (MSQ) compared to the logistic and Richards models. However, the logistic model receives a relatively small weight. Regarding the growth rates, the logistic model estimates a 3.99% growth rate, while the Gompertz model estimates a 1.73% growth rate. In contrast, the Richards model estimates a high growth rate of 8%. The ensemble model's estimated growth rate falls in between, at 3.54%.

Table 1. Estimated parameters by growth model and ensemble model.

Model	K	S.E.(K)	γ	S.E.(γ)	α	S.E.(α)	Weight
Logistic	3550	0.1056	0.0399	4.3107×10^{-6}	—	—	0.0329
Gompertz	3700	0.2182	0.0173	2.3139×10^{-6}	—	—	0.6902
Richards	3672	0.2937	0.0800	5.7843×10^{-4}	0.2585	0.0006	0.2769
Ensemble	3687	0.2354	0.0354	1.6191×10^{-4}	—	—	1

For comparative purposes, we also generated 1000 replicas of the Gompertz, logistic, and Richards models to build the respective confidence intervals and calculate the interval forecast mean score (IFMS). Table 2 presents forecast performance metrics for each model: the determination coefficient R^2 , the mean absolute error (MAE), the mean square error (MSE), and the IFMS with a 95% confidence level for the cumulative number of deaths. The results indicate that the logistic model had a smaller determination coefficient, with larger MAE and MSE compared to the other models, consistent with the observations in Figure 3 and the weights assigned to this model (Table 1). The Gompertz, Richards, and ensemble models showed high determination coefficients (above 0.99), indicating a good fit to the data. Notably, the Gompertz model had the smallest MAE and MSE, outperforming even the ensemble model. Additionally, the confidence intervals constructed using the Gompertz and ensemble models exhibited the smallest IFMS, indicating superior performance in these intervals.

Table 2. Determination coefficient, mean absolute error, mean square error, and interval forecast mean score to assess forecast performance.

Model	R^2	MAE	MSE	IFMS
Logistic	0.9614	205.8528	59,393.680	5640
Gompertz	0.9982	46.1195	2831.737	205.9703
Richards	0.9955	71.0488	7058.065	608.7089
Ensemble	0.9963	52.0208	3914.654	234.8977

Table 3 displays forecast performance metrics and IFMS for each model. The results indicate that the logistic and Richards models exhibited better fitting performance as measured by MAE and MSE. Additionally, these models had the smallest IFMS, indicating superior interval estimation performance. In Figure 4, you can see a comparison between the curve fitted by the ensemble model using the original data and the average ensemble curve for the 1000 replicas. It is evident that the ensemble curve and the average ensemble curve for the replicas are very close and fit the data well. However, there is a noticeable deviation in the predictions after the month of October and in the forecast. This difficulty in forecasting deaths may be attributed to the sudden change in the curve toward the end of 2020.

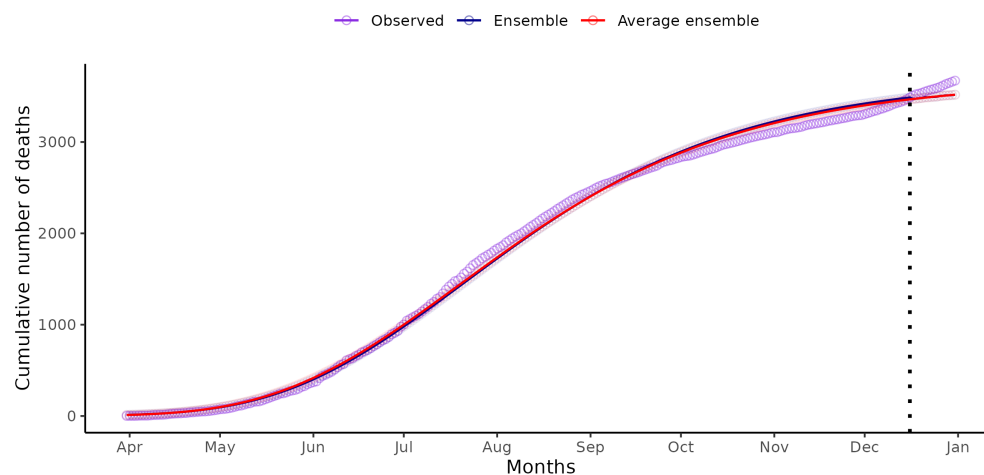


Figure 4. Daily cumulative incidence of death counts, ensemble model, and average of the ensemble models based on replicas.

Table 3. Mean absolute error, mean square error, and interval forecast mean score to assess the forecast performance.

Model	MAE	MSE	IFMS
Logistic	55.6799	4860.5070	278.7717
Gompertz	106.8016	12,591.6100	642.5833
Richards	69.2534	6162.4490	270.2633
Ensemble	73.2892	6629.3380	293.6600

Figure 5 provides a visual representation of the 95% confidence interval constructed for the cumulative number of deaths using the ensemble model. The vertical line marks the starting point of the forecast. It is evident that the interval has a small width, indicating good precision in the interval estimation. While there are a few data points outside the interval boundaries, the distance between these points and the interval is not substantial. Overall, the interval appears to satisfactorily capture the observed death curve.

After generating the ensemble model replicas, the growth models were refitted to each of the 1000 replicas, resulting in 1000 fitted curves for each of the three models. Subsequently, the MSE, MAE, and model weights were recalculated for each replica, and an ensemble curve was constructed for each of them. Table 4 presents the means of the estimates obtained from these fittings, including the final size of the pandemic K , growth rate γ , shape parameter α , their respective standard errors, and the weights of each model. Notably, the estimates for the final size of the pandemic and the growth rate for the ensemble model were slightly larger than those in Table 1. This variation is due to changes in the weights assigned to the growth models. The weight for the Gompertz model decreased from 0.6902 (Table 1) to 0.5649, while the weight of the Richards model increased from

0.2669 to 0.4123. The estimated growth rate was 4.35%, slightly higher than that found for the model with a direct fit to the data.

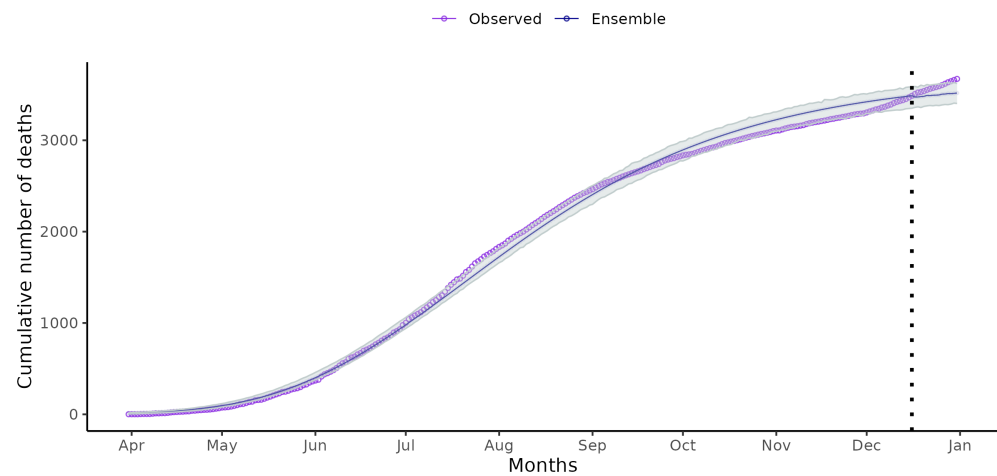


Figure 5. Daily cumulative incidence of death counts and ensemble model. The shaded region indicates the 95% uncertainty intervals.

Table 4. Replicates: Parameter estimates for each growth model and the ensemble model.

Model	K	S.E.(K)	γ	S.E.(γ)	α	S.E.(α)	Weight
Logistic	3550	0.1078	0.0274	2.7990×10^{-6}	—	—	0.0228
Gompertz	3730	0.2188	0.0175	2.4002×10^{-6}	—	—	0.5649
Richards	3672	0.2935	0.0801	1.5098×10^{-4}	0.2586	5.7834×10^{-4}	0.4123
Ensemble	3702	0.2471	0.0435	6.3664×10^{-5}	—	—	1

Table 5 presents the mean performance metrics obtained from the refitted models. These metrics include the determination coefficient R^2 , MAE, and MSE. To calculate the MSE for the ensemble model, each growth model's weight was multiplied by the respective MSE for the b -th replica, where $b = 1, 2, \dots, 1000$. The average of these MSE values was then computed. The same process was applied to calculate the determination coefficient and MAE. The high MSE value for the logistic model resulted in a significantly lower weight compared to the Gompertz and Richards models. Additionally, it is worth noting that the models refitted to the replicas exhibited a lower average MSE compared to the models fitted directly from the data, including the ensemble model.

Table 5. Ensemble replicas results: determination coefficient, mean absolute error, and mean square error to assess the prediction performance.

Model	R^2	MAE	MSE
Logistic	0.9707	206.7985	45,501.1777
Gompertz	0.9988	48.2287	1846.1060
Richards	0.9984	71.0308	2528.5940
Ensemble	0.9979	61.2420	3123.2370

Table 6 provides the average values of MAE and MSE for the forecasts produced by fitting the models to the 1000 replicas. The Gompertz model showed the highest MAE and MSE values, contrary to the results observed in the predictions.

Table 6. Ensemble replicas results: mean absolute error and mean square error to assess the forecast performance.

Model	MAE	MSE
Logistic	56.2020	4950.2090
Gompertz	74.5657	6830.1650
Richards	65.9720	5740.0720
Ensemble	70.6751	6349.1710

It is evident that the forecasts generated by the growth models and the ensemble model did not perform well. Despite the second wave of COVID-19 occurring in 2021, it is noticeable in Figure 1 that the number of deaths began to increase again around November and December 2020. Consequently, the ensemble model was fitted to the data of registered deaths until 30 September 2020, and forecasts of 15 and 30 days ahead were carried out.

Table 7 provides the results for each growth model, as well as the ensemble model. Among the three growth models, the Richards model had the largest weight in the construction of the ensemble model, which estimated that the growth rate in Paraíba is approximately 8.3%, with a final number of deaths in the state projected to be 3281. (Using the official data, the model estimated that the pandemic would end around 27 November 2020).

Table 7. Estimated parameters for each growth model and the ensemble model for the number of registered deaths due to COVID-19 in Paraíba until 30 September.

Model	K	S.E.(K)	γ	S.E.(γ)	α	S.E.(α)	Weight
Logistic	2775	0.2019	0.0458	8.1582×10^{-6}	—	—	0.0380
Gompertz	3444	0.5655	0.0203	5.1910×10^{-6}	—	—	0.2765
Richards	3244	0.8318	0.1102	3.9789×10^{-4}	0.2239	9.7707×10^{-4}	0.6855
Ensemble	3281	0.7342	0.0829	2.7450×10^{-4}	—	—	1

Tables 8 and 9 display the performance metrics for the predictions and the 15-day ahead forecasts. The determination coefficient suggests that all four models provide a satisfactory fit to the data. The Richards and ensemble models outperformed the others in terms of MAE and MSE for both predictions and 15-day ahead forecasts, which aligns with the higher weight assigned to the Richards model. It is worth noting that the ensemble model exhibited superior forecast performance compared to the individual models.

Table 8. Performance metrics of the predictions for each model fitted to the registered number of deaths up to 30 September 2020: determination coefficient, mean absolute error, and mean square error.

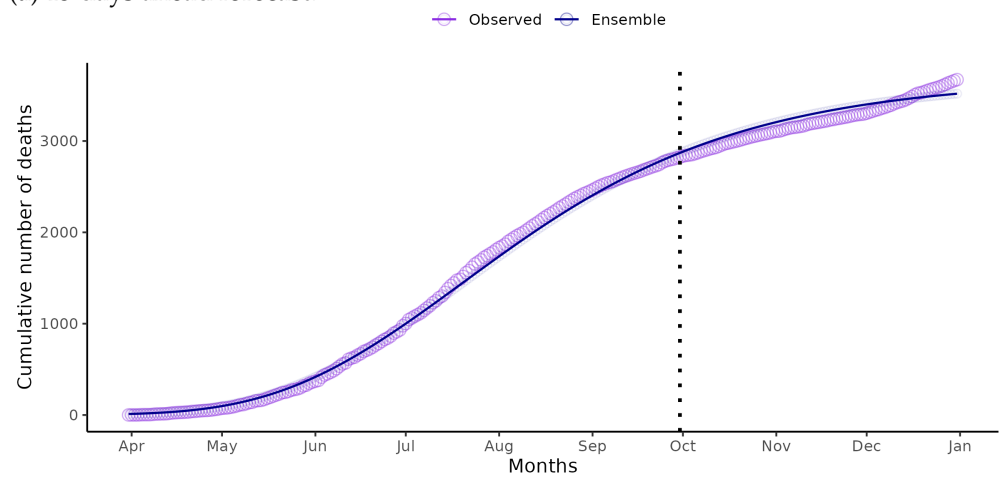
Model	R^2	MAE	MSE
Logistic	0.9970	43.7131	3003
Gompertz	0.9996	17.1538	412
Richards	0.9998	9.7566	166
Ensemble	0.9997	10.4711	183

Table 9. Performance metrics of the forecasts for each model fitted to the registered number of deaths up to 30 September 2020: determination coefficient, mean absolute error, and mean square error.

Model	MAE	MSE
Logistic	181.4285	33,874
Gompertz	25.9068	698
Richards	9.7615	160
Ensemble	7.0135	75

Figure 6 presents the death curve simulated by the ensemble model and the 15- and 30-day ahead forecasts. The model fits the data perfectly and generates excellent forecasts in both cases. Therefore, it is evident that using data up to the point where the real curve started to accelerate again resulted in better prediction and forecast performance.

(a) 15 days ahead forecast.



(b) 30-day ahead forecast.

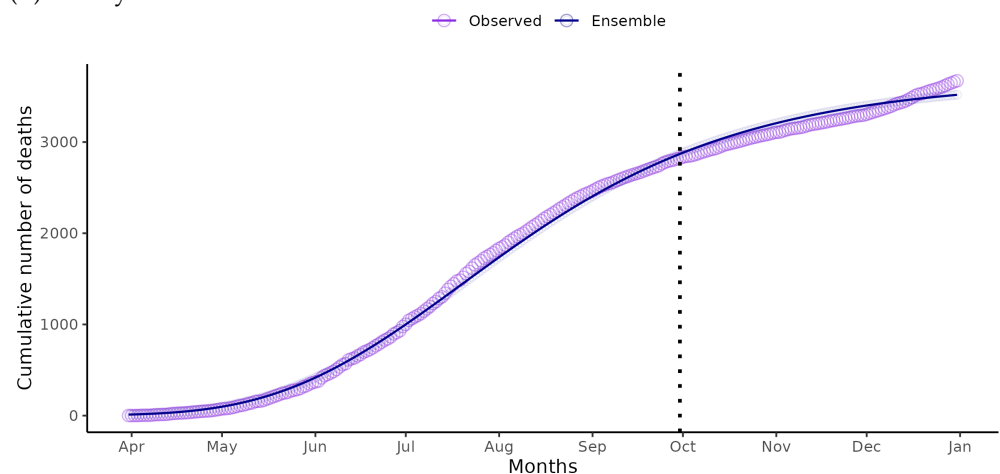


Figure 6. Daily cumulative incidence of death counts and ensemble model.

Figure 7 displays the confidence interval produced by the ensemble model for the number of registered deaths up to 30 September and the 30-day ahead forecast. The interval performs exceptionally well and closely aligns with the death curve provided by the Ministry of Health.

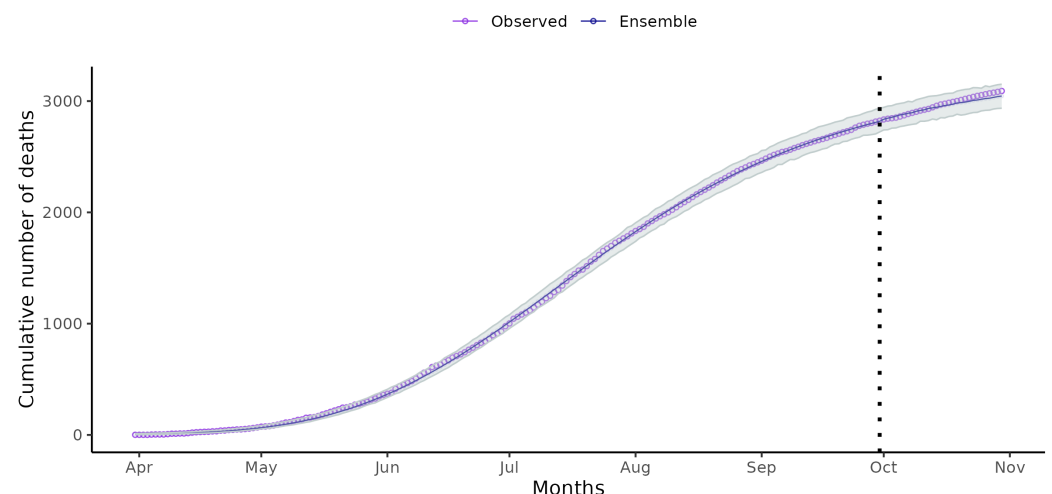


Figure 7. Daily cumulative incidence of death counts up to 30 September 2020 and ensemble model. The shaded regions indicate the 95% uncertainty intervals.

5. Conclusions

In the early stages of the pandemic, when our understanding of SARS-CoV-2 was limited, numerous scientific studies emerged to address the challenges posed by COVID-19. Among these challenges, the under-reporting of cases has been a significant concern, prompting researchers to explore alternative approaches. To enhance our data analysis, we focused on cumulative death numbers, which offer more stability than daily death counts and are less reliant on the notification of infected cases.

In our study, we employed the logistic, Richards, and Gompertz models to fit COVID-19 death data spanning from March to December 2020. These models were then used to construct an ensemble model. The logistic model exhibited poor performance in fitting the data, while the Richards and Gompertz models displayed better predictive capabilities. However, despite their strong fit to historical data, these models struggled to provide precise forecasts. This challenge became particularly evident in November 2020 when the death curve displayed an unexpected upward trend that the models could not anticipate.

In light of the challenges faced by the individual growth models in accurately forecasting the COVID-19 death data, we turned to the development of an ensemble model. This ensemble approach demonstrated remarkable prediction performance and generated forecasts that closely resembled those produced by the growth models.

It is worth noting that, although the second wave of COVID-19 in Brazil emerged in 2021, there was already an acceleration in the number of deaths during the later months of 2020. As a result, we trained the models using data up to 30 September and conducted forecasts for 15 and 30 days ahead. The ensemble model outperformed the individual growth models in both prediction and forecasting, proving its effectiveness in modeling a single wave of COVID-19 data.

Author Contributions: Conceptualization, R.O. and C.F.; methodology, J.O. and R.O.; software, J.O., A.L. and R.O.; validation, R.O., C.F. and A.L.; investigation, J.O., R.O., C.F. and A.L.; data curation, J.O., C.F. and R.O.; writing—original draft preparation, J.O., J.G. and R.O.; writing—review and editing, J.O., C.F., R.O., A.L. and J.G.; supervision, R.O. and C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by the National Council for Scientific and Technological Development (CNPq) through the grant 303192/2022-4 (RO) and Comissão de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) number 001 in Brazil. The authors thank the editor and anonymous referees for comments and suggestions.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and codes used in this study are available under request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization (WHO). Pandemic Definition. 2020. Available online: <https://www.publichealth.com.ng/world-health-organization-who-pandemic-definition/> (accessed on 12 July 2023).
2. Horrox, R. *The Black Death*; Manchester University Press: Manchester, UK, 2013.
3. Bloom, B.R.; Fine, P.E.M. The BCG Experience: Implications for Future Vaccines Against Tuberculosis. In *Tuberculosis: Pathogenesis, Protection, and Control*; ASM Press: Washington, DC, USA, 1994; pp. 531–558. [CrossRef]
4. Aassve, A.; Alfani, G.; Gandolfi, F.; Le Moglie, M. Epidemics and trust: The case of the Spanish Flu. *Health Econ.* **2021**, *30*, 840–857. [CrossRef] [PubMed]
5. Tsoucalas, G.; Kousoulis, A.; Sgantzios, M. The 1918 Spanish Flu Pandemic, the origins of the H1N1-virus strain, a glance in history. *Eur. J. Clin. Biomed. Sci.* **2016**, *2*, 23–28.
6. Suryasa, I.W.; Rodríguez-Gámez, M.; Koldoris, T. The COVID-19 pandemic. *Int. J. Health Sci.* **2021**, *5*, 6–9. [CrossRef]
7. Bulut, C.; Kato, Y. Epidemiology of COVID-19. *Turk. J. Med. Sci.* **2020**, *50*, 563–570. [CrossRef]
8. World Health Organization. Coronavirus Disease (COVID-19): How Is It Transmitted? 2021. Available online: <https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted> (accessed on 27 August 2023).
9. Bhardwaj, R.; Agrawal, A. How coronavirus survives for days on surfaces. *Phys. Fluids* **2020**, *32*, 111706-1–111706-7. [CrossRef]
10. Governo do Brasil, M. Brasil Confirma Primeiro Caso Do Novo Coronavírus. 2020. Available online: <https://www.gov.br/pt-br/noticias/saude-e-vigilancia-sanitaria/2020/02/brasil-confirma-primeiro-caso-do-novo-coronavirus> (accessed on 30 April 2020).
11. Ministério da Saúde, B.S.d.V.e.S. Boletim Epidemiológico Especial—COE Coronavírus—09 de abril de 2020. 2020. Available online: <https://www.saude.gov.br/images/pdf/2020/> (accessed on 30 April 2020).
12. Ospina, R.; Leite, A.; Ferraz, C.; Magalhaes, A.; Leiva, V. Data-driven tools for assessing and combating COVID-19 outbreaks in Brazil based on analytics and statistical methods. *Signa Vitae* **2022**, *18*, 18–32.
13. Ospina, R.; Gondim, J.A.; Leiva, V.; Castro, C. An Overview of Forecast Analysis with ARIMA Models during the COVID-19 Pandemic: Methodology and Case Study in Brazil. *Mathematics* **2023**, *11*, 3069. [CrossRef]
14. Ferreira, W.G.; Serpa, A.L. Ensemble of metamodels: The augmented least squares approach. *Struct. Multidiscip. Optim.* **2016**, *53*, 1019–1046. [CrossRef]
15. Leutbecher, M.; Palmer, T.N. Ensemble forecasting. *J. Comput. Phys.* **2008**, *227*, 3515–3539. [CrossRef]
16. Chowell, G.; Luo, R. Ensemble bootstrap methodology for forecasting dynamic growth processes using differential equations: Application to epidemic outbreaks. *BMC Med. Res. Methodol.* **2021**, *21*, 34. [CrossRef]
17. Hethcote, H.W. The Mathematics of Infectious Diseases. *SIAM Rev.* **2000**, *42*, 599–653. [CrossRef]
18. Hamer, W.H. *Milroy Lectures on Epidemic Disease in England*; Nabu Press: Charleston, SC, USA, 2010.
19. Ross, R. Some a priori pathometric equations. *Br. Med. J.* **1915**, *1*, 546. [CrossRef] [PubMed]
20. Bacaër, N.; Bacaër, N. McKendrick and Kermack on epidemic modelling (1926–1927). In *A Short History of Mathematical Population Dynamics*; Springer: London, UK, 2011; pp. 89–96.
21. Brauer, F.; Castillo-Chavez, C.; Feng, Z. *Mathematical Models in Epidemiology*; Springer: New York, NY, USA, 2019; Volume 32.
22. Busenberg, S. *Differential Equations and Applications in Ecology, Epidemics, and Population Problems*; Elsevier: Amsterdam, The Netherlands, 2012.
23. Chowell, G.; Sattenspiel, L.; Bansal, S.; Viboud, C. Mathematical models to characterize early epidemic growth: A review. *Phys. Life Rev.* **2016**, *18*, 66–97. [CrossRef] [PubMed]
24. Pingle, M. Introducing dynamic analysis using Malthus’s Principle of Population. *J. Econ. Educ.* **2003**, *34*, 3–20. [CrossRef]
25. Burghes, D.N. Population dynamics An introduction to differential equations. *Int. J. Math. Educ. Sci. Technol.* **1975**, *6*, 265–276. [CrossRef]
26. Espindola, A.L.; Girardi, D.; Penna, T.J.; Bauch, C.T.; Martinez, A.S.; Cabella, B.C. Exploration of the parameter space in an agent-based model of tuberculosis spread: Emergence of drug resistance in developing vs developed countries. *Int. J. Mod. Phys. C* **2012**, *23*, 1250046. [CrossRef]
27. Spósito, M.B.; Amorim, L.; Bassanezi, R.B.; Yamamoto, P.T.; Felipe, M.R.; Czeremainski, A.B. Relative importance of inoculum sources of *Guignardia citricarpa* on the citrus black spot epidemic in Brazil. *Crop Prot.* **2011**, *30*, 1546–1552. [CrossRef]
28. Hirata, Y.; Bruchovsky, N.; Aihara, K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *J. Theor. Biol.* **2010**, *264*, 517–527. [CrossRef]
29. Tanaka, G.; Hirata, Y.; Goldenberg, S.L.; Bruchovsky, N.; Aihara, K. Mathematical modelling of prostate cancer growth and its application to hormone therapy. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2010**, *368*, 5029–5044. [CrossRef]

30. Lôbo, R.N.B. Evaluation of body weight standardization methods for 205, 365 and 550 days of age. *R. Bras. Zootec.* **2002**, *31*, 1695–1706. [[CrossRef](#)]
31. Guedes, M.; Muniz, J.; Silva, F.; Aquino, L. Bayesian analysis of growth curve of Santa Inês lambs. *Arq. Bras. Med. Vet. Zootec.* **2005**, *57*, 415–417. [[CrossRef](#)]
32. Mello, F.d.; Oliveira, C.A.; Ribeiro, R.P.; Resende, E.K.; Povh, J.A.; Fornari, D.C.; Barreto, R.V.; McManus, C.; Streit, D., Jr. Growth curve by Gompertz nonlinear regression model in female and males in tambaqui (*Colossoma macropomum*). *An. Acad. Bras. Ciênc.* **2015**, *87*, 2309–2315. [[CrossRef](#)] [[PubMed](#)]
33. Valle, J.A.M. Predicting the number of total COVID-19 cases and deaths in Brazil by the Gompertz model. *Nonlinear Dyn.* **2020**, *102*, 2951–2957. [[CrossRef](#)] [[PubMed](#)]
34. Dutra, C.M.; Farias, F.M.; Madrid, M.G.; de Melo, C.A.R. Estimated number of deaths, confirmed cases and duration of the COVID-19 Pandemic in Brazil. *Braz. J. Health Rev.* **2020**, *3*, 10266–10284. [[CrossRef](#)]
35. da Silva, E.V.; da Silva Melo, J.; Leite, M.A. Modelo bi-logístico aplicado aos primeiros 1015 casos de COVID-19 em indígenas do Estado do Amapá e norte do Pará. *Sci. Knowl. Focus* **2021**, *3*, 77–88.
36. Vasconcelos, G.L.; Duarte-Filho, G.C.; Brum, A.A.; Ospina, R.; Almeida, F.A.; Macêdo, A.M. Analysis of COVID-19 epidemic curves via generalized growth models: Case study for the cities of Recife and Teresina. *SciELO Prepr.* **2020**. [[CrossRef](#)]
37. Vasconcelos, G.L.; Macêdo, A.M.; Ospina, R.; Almeida, F.A.; Duarte-Filho, G.C.; Brum, A.A.; Souza, I.C. Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ* **2020**, *8*, e9421. [[CrossRef](#)]
38. Ribeiro, M.H.D.M.; da Silva, R.G.; Moreno, S.R.; Mariani, V.C.; dos Santos Coelho, L. Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107712. [[CrossRef](#)]
39. Ribeiro, M.H.D.M.; Stefenon, S.F.; de Lima, J.D.; Nied, A.; Mariani, V.C.; Coelho, L.d.S. Electricity price forecasting based on self-adaptive decomposition and heterogeneous ensemble learning. *Energies* **2020**, *13*, 5190. [[CrossRef](#)]
40. Bravo, J.M.; Ayuso, M. Mortality and life expectancy forecasts using bayesian model combinations: An application to the portuguese population. *RISTI. Rev. Ibér. Sist. Tecnol. Inf. E* **2020**, *40*, 128–144.
41. Allyn, J.; Allow, N.; Augustin, P.; Filipi, I.; Martine, O.; Belghiti, M.; Provenchère, S. A Comparison of a Machine Learning Model with EuroSCORE II in Predicting Mortality after Elective Cardiac Surgery: A Decision Curve Analysis. *PLoS ONE* **2017**, *12*, e0169772. [[CrossRef](#)] [[PubMed](#)]
42. Lindstrom, T.; Tildesley, M.; Webb, C. A Bayesian Ensemble Approach for Epidemiological Projections. *PLoS Comput. Biol.* **2015**, *11*, e1004187. [[CrossRef](#)] [[PubMed](#)]
43. Yamana, T.K.; Kandula, S.; Shaman, J. Superensemble forecasts of dengue outbreaks. *J. R. Soc. Interface* **2016**, *13*, 20160410. [[CrossRef](#)] [[PubMed](#)]
44. Bühlmann, P. Bagging, boosting and ensemble methods. In *Handbook of Computational Statistics: Concepts and Methods*; Springer: Berlin, Germany, 2012; pp. 985–1022.
45. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
46. Dietterich, T.G. Ensemble methods in machine learning. In *International Workshop on Multiple Classifier Systems*; Springer: Berlin, Germany, 2000; pp. 1–15.
47. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; CRC Press: Boca Raton, FL, USA, 2012.
48. Kotu, V.; Deshpande, B. *Predictive Analytics and Data Mining*, 1st ed.; Morgan Kaufmann: Burlington, MA, USA, 2015; ISBN 9780128016503.
49. Gompertz, B. XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c. *Philos. Trans. R. Soc. Lond.* **1825**, *115*, 513–583.
50. Adam, J.A.; Bellomo, N. *A survey of Models for Tumor-Immune System Dynamics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1997.
51. Domingues, J.S. Gompertz model: Resolution and analysis for tumors. *J. Math. Model. Appl.* **2012**, *1*, 70–77.
52. Boyce, W.E. *Elementary Differential Equations and Boundary Value Problems*, 7th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2000; ISBN 978-0471319993.
53. Artzrouni, M.; Komlos, J. Population growth through history and the escape from the Malthusian trap: A homeostatic simulation model. *Genus* **1985**, *41*, 21–39.
54. Verhulst, P. Notice on the law that the population follows in its growth. *Corresp. Math. Phys.* **1838**, *10*, 113–26.
55. Bassanezi, R.C. *Ensino-Aprendizagem com Modelagem Matemática*, 3rd ed.; Editora Contexto: Madrid, Spain, 2011; ISBN 9788572442077.
56. Tsoularis, A.; Wallace, J. Analysis of Logistic Growth Models. *Math. Biosci.* **2002**, *179*, 21–55. [[CrossRef](#)]
57. Richards, F.J. A flexible growth function for empirical use. *J. Exp. Bot.* **1959**, *10*, 290–301. [[CrossRef](#)]
58. Hsieh, Y.H. Richards Model: A Simple Procedure for Real-time Prediction of Outbreak Severity. In *Modeling and Dynamics of Infectious Diseases Series in Contemporary Applied Mathematics (CAM)*; World Scientific: Singapore, 2009; Volume 11. . [[CrossRef](#)]
59. Nagelkerke, N.J. A note on a general definition of the coefficient of determination. *Biometrika* **1991**, *78*, 691–692. [[CrossRef](#)]
60. Isabona, J.; Imoize, A.L.; Kim, Y. Machine learning-based boosted regression ensemble combined with hyperparameter tuning for optimal adaptive learning. *Sensors* **2022**, *22*, 3776. [[CrossRef](#)] [[PubMed](#)]

-
61. Gneiting, T.; Raftery, A.E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [\[CrossRef\]](#)
 62. Araujo, L.N.; Belotti, J.T.; Alves, T.A.; de Souza Tadano, Y.; Siqueira, H. Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. *Environ. Model. Softw.* **2020**, *123*, 104567. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.