A PIECEWISE GROWTH MODEL FOR MODELING THE ACCUMULATED NUMBER OF COVID-19 CASES IN THE CITY OF CAMPO GRANDE

Erlandson Ferreira SARAIVA[1]
Leandro SAUER[2]
Basílio de Bragança PEREIRA[3]
Carlos Alberto de Bragança PEREIRA[4]

▪ ABSTRACT: In December of 2019, a new coronavirus was discovered in the city of Wuhan, China. The World Health Organization officially named this coronavirus as COVID-19. Since its discovery, the virus has spread rapidly around the world and is currently one of the main health problems, causing an enormous social and economic burden. Due to this, there is a great interest in mathematical models capable of projecting the evolution of the disease in countries, states and/or cities. This interest is mainly due to the fact that the projections may help the government agents in making decisions in relation to the prevention of the disease. By using this argument, the health department of the city (HDC) of Campo Grande asked the UFMS for the development of a mathematical study to project the evolution of the disease in the city. In this paper, we describe a modeling procedure used to fit a piecewise growth model for the accumulated number of cases recorded in the city. From the fitted model, we estimate the date in which the pandemic peak is reached and project the number of patients who will need treatment in intensive care units. Weekly, was sent to HDC a technical report describing the main results.

▪ KEYWORDS: COVID-19, Growth models, Estimation, Non-linear least squares, Model selection.

[1]Universidade Federal de Mato Grosso do Sul - UFMS, Instituto de Matemática, CEP: 79070-900, Campo Grande, MS, Brasil. E-mail: *erlandson.saraiva@ufms.br*
[2]Universidade Federal de Mato Grosso do Sul - UFMS, Escola de Administração e Negócios, CEP: 79070-900, Campo Grande, MS, Brasil. E-mail: *leandro.sauer@ufms.br*
[3]Universidade Federal do Rio de Janeiro, Faculdade de Medicina e COPPE, CEP: 21941-972, Rio de Janeiro, RJ, Brasil. E-mail: *basilio@hucff.ufrj.br*
[4]Universidade de São Paulo - USP, Departamento de Estatística, CEP: 05508-010, São Paulo, SP, Brasil. E-mail: *cadebp@gmail.com*

1 Introduction

On February 26th of 2020, the first case of COVID-19 was confirmed in Brazil. A man 61-years-old resident in the city of São Paulo, SP, Brazil. And on March 14th of 2020, the first case was confirmed in the city of Campo Grande, MS, Brazil. From the first case, the health department of the city (HDC) of Campo Grande asked the Federal University of Mato Grosso do Sul (UFMS) for the development of a mathematical model to monitor and project the evolution of the disease in the city. The HDC argued that the modeling results would serve to support some decisions in relation to disease prevention, such as, proposing the intensification of social isolation, the purchase of hospital equipment, an increase in the number of intensive care units in public hospitals, among others.

From the request of the HDC, the authors of this paper began a study with the aim of fitting a growth model for the accumulated number of the COVID-19 cases recorded in the city of Campo Grande. The modeling was started by calculating the main descriptive statistics, followed by the fit of the most known growth model, the exponential model. However, as described by several epidemiological studies the number of people becoming ill in an epidemic grows until a limit (MEYER, 1999); and not indefinitely as described by the exponential model. Due to this, it is usual to consider that the growth process of the number of cases of a disease is given by a model in which the graphic is a sigmoidal curve (S-shaped curve).

Thus, besides the exponential model, we also consider the fit of the Logistic and Gompertz growth models. Both models are characterized by an S-shape curve defined by two distinct phases. The first phase is characterized by growth at an increasing rate (positive slope) and the second phase is characterized by growth at a decreasing rate (negative slope). The point in which the curve changes the slope (positive to negative) is called the inflection point. In the context of an epidemiological study, this point indicates when the pandemic peak is reached. The choice for these two growth models is based on studies described in the literature that indicate that both models are excellent for use in quantitative longitudinal data, see for example Budimulyati et al., (2012) and its references.

However, in the course of the study, we noticed that the fit of a single growth model would not be suitable, since the recorded values indicated a change in the pandemic's growth behavior. Due to this, we adopted the fit of a piecewise growth model. In order to fit a piecewise model, we separated the dataset into four sub-datasets. This separation was done in the course of the analysis using the model with the lowest mean square error as a criterion for choosing the separation point.

For each one of the sub-datasets, we fit the three growth models; where the estimates for the parameters of the models were obtained using the non-linear least square method, as decribed by Vieira and Hoffman (1977) and Hsieh (2017). To choose the best model for each sub-dataset, we consider as a criterion the mean square error (MSE) and the model selection criteria Akaike Information criterion (AKAIKE, 1974; BOZDOGAN, 1987), denoted by AIC, and the Bayesian Information criterion (SCHWARZ, 1978), denoted by BIC. This procedure allowed

us to get a fitted model with the smallest mean square error among the tested models. The resulting piecewise growth model has the following configuration: Gompertz, Gompertz, Exponential and Gompertz model.

Based on the fitted model, we get the coordinates of the inflection point and consequently the estimated date for the peak of the pandemic. In addition, using a moving-sums procedure, we project the number of patients who will need treatment in the clinical units and in intensive care units. These pieces of information were sent weekly to the HDC in a technical report format. Until 11/08/20 were sent twenty one technical reports to HDC. However, in this paper, we focus on describing the modeling procedure adopted to obtain the piecewise model. That is, we focus on the four periods in which happened the change in the pandemic's growth behavior. If it is of interest to the reader, the technical reports sent to HDC can be obtained upon request by email to the authors.

The remainder of the paper is organized as follows. In section 2, we present the data and the three growth models considered. Section 3, describes the modeling procedure adopted to get a piecewise growth model. Section 4, presents the projections for the number of patients who will need care in clinical or intensive care units. Section 5 concludes the paper with the final remarks. Additional details are provided in the supplementary material, denoted by the prefix "SM" when referred to in this paper.

2   Dataset and growth models

Let X_t be the number of recorded cases of COVID-19 in the city of Campo Grande on the t-th day, for $t = 0, \ldots, T = 239$, where $t = 0$ represents the day that the first case was recorded (03/14/20) and T is the last day considered in the study (11/08/20). This dataset is publicly available on the website www.sesau.gov.br. In this period of 240 days, $37,429$ cases of COVID-19 were recorded and 696 people died due to this disease. Besides, of the $37,429$ recorded cases, $36,017$ are considered recovered and 601 people are in home isolation.

Our first analysis consists of visualizing the recorded values by using some plots and by calculating the main descriptive measures. Figure 1 shows the number of confirmed cases at the day t, for $t = 0, \ldots, T$. The highest number of registered cases at a day was $1,434$ cases on 09/11/20.

Figure 2 shows the barplot of the frequency distribution of the number of cases confirmed by month. Out of the total cases, 0.80% (299 cases) were registered in the months of March (38 cases), April (90 cases) and May (171 cases), 5.19% (1,942 cases) in June, 22.12% (8,285 cases) in July, 29.21% (10,940 cases) in August, 24.78% (5,347 cases) in September, 14.61% (4,515 cases) in the October and 3.28% (1,230 cases) in the first eight days of November.

Table 1 shows the descriptive statistics for the number of cases recorded per day. The median value is 109.50 confirmed cases, with an average of 155.95 recorded cases by day and a standard deviation (S.D.) of 179.95 cases.
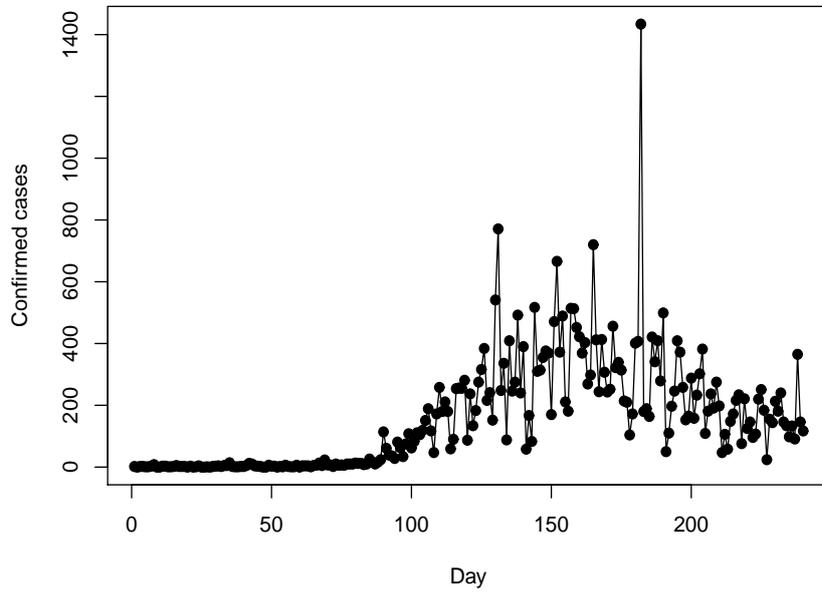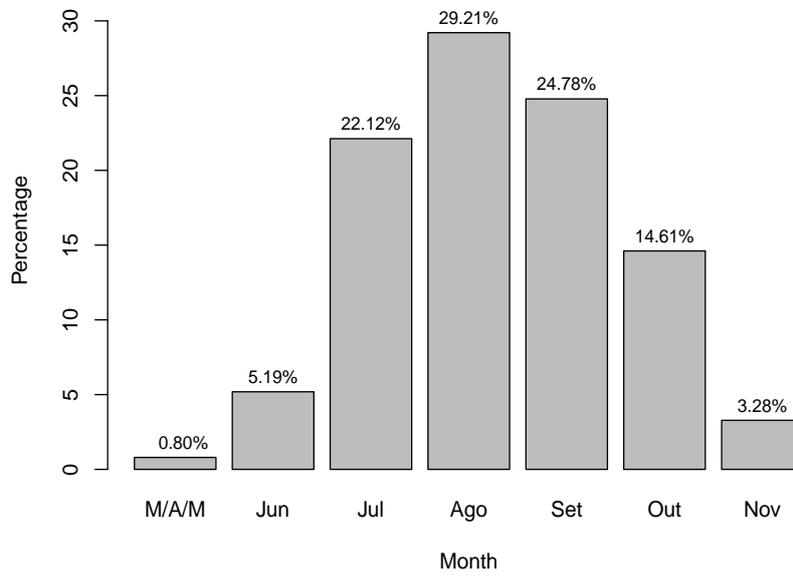
Figure 1 - Confirmed number of cases by day.

Figure 2 - Percentage of cases by month.

Table 1 - Desciptive Statistics

Minimum	1° Quartile	Median	Average	S. D.	3° Quartile	Maximum
0	5.75	109.50	155.95	179.95	246.50	1,434.00

Figure 3 shows the graphic of the moving-average of seven days for the number of recorded cases. In the last 21 days, the moving average value ramained inside the range of ±10% of the average of the moving-average values of this period (dotted lines). Due to this, we consider the confirmed number of cases was stable in this period. The moving-average value on the 139th day (11/08/20) was of 154.86, meaning that in the period from 11/02/20 to 11/08/20 was confirmed, in average, 154.86 case per day. Compared to the moving-average value of seven days ago (157.57 on 11/01/20), there was a reduction of 1.72%. Compared to the moving-average value of fourteen days ago (161.29 on 10/25/20), there was a reduction of 3.99%.
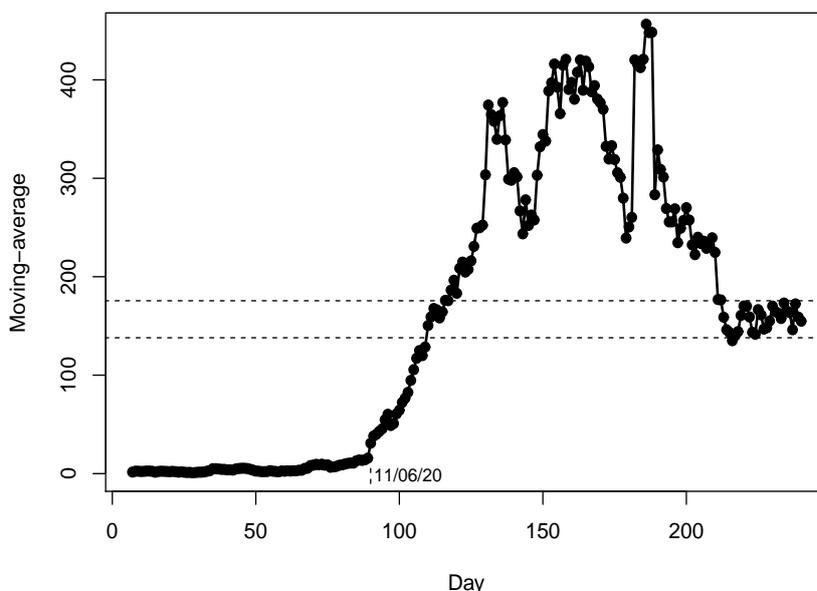

Figure 3 - Moving-average of seven days.

Figure 14 in Appendix 1 of the SM shows the graphic of the moving-average for the number of death. In the last four days, there was a reduction in the moving-average value. On 11/08/20 the moving-average value was of 2 death. Compared to the value of seven days ago (2.14 on 11/01/20), there was a reduction of 7%. Compared to fourteen days ago (3.86 on 10/25/20) the reduction was of 58.38%.

Figure 4 show the graphic of the accumulated number of cases in the original scale and in the log-scale. Note the fast growth of the accumulated number of cases after the 89th day (06/11/20). Our interest is to model the accumulated number of cases using a nonlinear growth model. To facilitate the modeling procedure, we opt to model the accumulated number of cases in the log-scale.
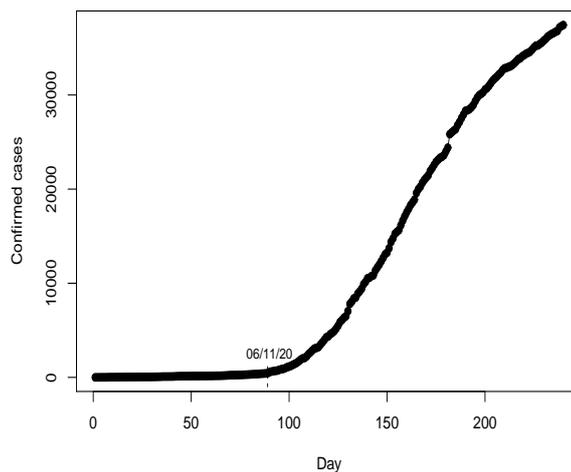
(a) Original scale.
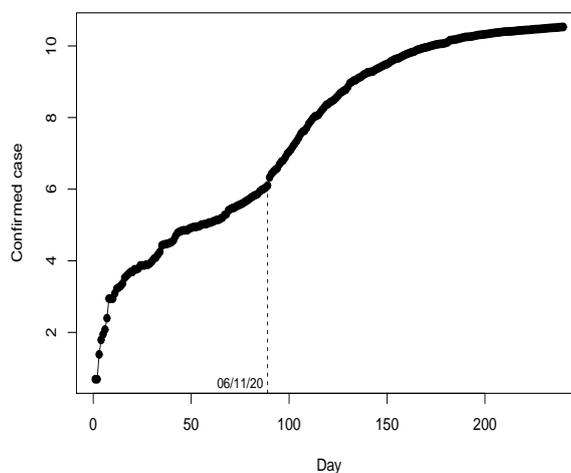
(b) Log-scale.

Figure 4 - Accumulated number of cases.

2.1 Growth models

In order to model the data described in the earlier section, consider N_t be the accumulated number of COVID-19 cases at time t and $f(t|\theta)$ a nonlinear function indexed by parameter θ (scalar or vector), for $t > 0$. Consider that $N_t = f(t|\theta)$ and $Y_t = log(N_t) = g(t|\theta)$, where $g(t|\theta) = log(f(t|\theta))$, for $t > 0$.

2.1.1 Exponential growth model

One of the most knows nonlinear models to describe the growth of an epidemiology disease is the exponential model. Its Equation is given by

$$N_t = f(t|\theta) = \alpha_1 \exp\{\alpha_2 t\}, \tag{1}$$

for $\theta = (\alpha_1, \alpha_2)$, where α_1 is the number of cases in the initial time, $t = 0$, and α_2 is the growth rate, for $t \geq 0$. For more details on the exponential model, please see Abramowitz (1965), Thomson (2005) and their references.

Taking the logarithmic transformation on both sides of the Equation (1), one gets the following linearized form of the model, called log-exponential model,

$$Y_t = log(N_t) = \alpha_1 + \alpha_2 t, \tag{2}$$

for $t \geq 0$.

Figure 15 in Appendix 2 of the SM shows the graphics of the exponential and log-exponential models for an initial value $\alpha_1 = 2$ and growth rate $\alpha_2 = \{0.10, 0.20, 0.30\}$. Increasing the value of α_2 more inclined is the curve, meaning a fast growth in the number of disease cases. In addition, the curve of this model grows indefinitely regardless of the population size. This can be viewed as a practical problem since the number of accumulated cases is restricted, for instance, to the population size.

2.1.2 Logistic growth model

Consider now the Logistic growth model (BLUMBERG, 1968). Its Equation is given by

$$N_t = f(t|\theta) = \alpha_1 \left(1 + \alpha_2 \exp\{-\alpha_3 t\}\right)^{-1} \tag{3}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3)$ are the model parameters, for $t \geq 0$.

In the opposite of the exponential model, this model has an S-shape curve and consequently a growth limit. The parameter α_1 is the upper asymptote. In the context of the COVID-19 the value of α_1 is an estimate for the maximum number of cases. The parameter α_2 is related to the coordinates (t_m, N_{t_m}) of the inflection point, where $t_m = \frac{log(\alpha_2)}{\alpha_3}$ and $N_{t_m} = \frac{\alpha_1}{2}$. The parameter α_3 is the intrinsic growth rate at the inflection point.

Taking the logarithmic transformation on both sides of the Equation (3), we get the log-logistic model,

$$Y_t = log(N_t) = log(\alpha_1) - log\left(1 + \alpha_2 \exp\{-\alpha_3 t\}\right), \quad \text{for} \ \ t \geq 0. \tag{4}$$

Figure 16 in Appendix 2 of the SM shows the graphics of the logistic and log-logistic models for a upper asymptote $\alpha_1 = 10,000$, $\alpha_2 = 5,000$ and $\alpha_3 = \{0.15, 0.25, 0.55\}$. Similar to the exponential model, by increasing the value of the parameter α_3 more inclined is the curve. In that Figure, the symbols $\bullet$ represents the inflection point.

2.1.3 Gompertz growth model

As the third growth model, consider the Gompertz model (GOMPERTZ, 1825; WINSOR, 1932). Its Equation is given by

$$N_t = f(t|\theta) = \alpha_1 \exp\left\{-\alpha_2 \exp\{-\alpha_3 t\}\right\}, \tag{5}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3)$ are the model parameters, for $t > 0$. The interpretation of the parameters is similar to the described for the logistic model.

The graphic of the Gompertz model is also a curve with an S-shape. The main difference in relation to the Logistic model is that the curve of the Gompertz model is not symmetric in relation to the inflection point. While the ordinate of the inflection point of the Logistic model is $\frac{\alpha_1}{2}$, for the Gompertz model this ordinate is approximately 37% of the α_1 value.

Taking the logarithmic transformation on both sides of the Equation (5), we get the log-Gompetz model,

$$Y_t = log(N_t) = log(\alpha_1) - \alpha_2 \exp\{-\alpha_3 t\}, \tag{6}$$

for $t \geq 0$.

Figure 17 in Appendix 2 of the SM shows the graphics of the Gompertz and log-Gompertz models for a upper asymptote $\alpha_1 = 10,000$, $\alpha_2 = 8$ and $\alpha_3 = \{0.10, 0.20, 0.40\}$. Similar to the exponential and logistic models, by increasing the value of the parameter α_3 more inclined is the curve. In that Figure, the symbols $\bullet$ represents the inflection point.

3 Piecewise model

Consider now the fitting of a growth model for the accumulated number of COVID-19 cases recorded in the city of Campo Grande. For this, assume that the log-transformed measures, Y_t, are generated according to the following model

$$Y_t = g(t|\theta) + \varepsilon_t,$$

where $g(t|\theta)$ is given by one of the Equations in (2), (4) or (6) and ε_t is a random error assumed from a normal distribution with mean 0, variance σ^2 and $Cov(\varepsilon_t, \varepsilon_{t'}) = 0$, for $t, t' = 1, \ldots, n$ and $t \neq t'$. For more details on the normality assumption, please see (VIEIRA and HOFFMANN, 1977; HSIEH, 2017; ZHAO, 2019).

In order to get the parameter estimates, we adopt the nonlinear least square method. For this, we use the software R (R CORE TEAM, 2020) and the command *nls* of the package *nlstools* (FLORENT *et al.*, 2015; PINHEIRO *et al.*, 2020). Besides, we compare the models using as a criterion the mean square error (MSE) and the model selection criteria AIC and BIC, which are calculated according to the following expressions

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\left(\hat{Y}_t - Y_t\right)^2, \;\; AIC = -2l(\hat{\theta}|\mathbf{y}) + 2k \;\; \text{and} \;\; BIC = -2l(\hat{\theta}|\mathbf{y}) + klog(n),$$

where $\hat{Y}_t$ is the estimated value by the model, $l(\hat{\theta}|\mathbf{y}) = log(L(\hat{\theta}|\mathbf{y})$, in which, $L(\hat{\theta}|\mathbf{y})$ is the maximum value of the likelihood function for the model and k is the number of estimated parameters in the model. The best model is the one that has the smallest MSE, AIC and BIC values.

3.1 Model fitting 1

Our first analysis was done thirty days after the registration of the first case (03/14/20). Thus, let $D_1 = \{y_0, \ldots, y_{29}\}$ be the registered number of cases from day 0 (03/14/20) to 29th day (04/12/20) in an accumulated way and log-transformed.

Defined the dataset, we fit the three growth models described in Section 2. Table 2 shows the MSE, AIC and BIC values for the three fitted growth models. The smallest values are highlighted in bold. As one can note, the log-Gompertz model is the best model.

Table 2 - MSE, AIC and BIC values

Model	Criterion		
	MSE	AIC	BIC
Log-Exponential	0.1771	39.2053	43.4089
Log-Logistic	0.0176	-28.0450	-22.4402
Log-Gompertz	**0.0092**	**-47.2643**	**-41.6595**

Table 3 shows the estimates and the standard errors for parameters of the log-Gompertz model. The fitted model is given by

$$\hat{Y}_t = 3.9909 - 3.9493\exp\{-0.1346t\}, \;\; \text{for} \;\; t > 0.$$

Table 3 - Parameter estimates

Values	Parameter		
	$log(\alpha_1)$	α_2	α_3
Estimates	3.9909	3.9493	0.1346
Standard error	0.0460	0.0862	0.0070

Figure 5 shows the graphic of the values of the dataset D_1 and the fitted model (black line) for a period of 50 days (from day 0 to day 49), being 30 of fitting and 20 of projection. The 49th day represents the date 05/02/20. The projection of the fitted model for the 49th day is $\hat{y}_{49} = 3.9862$. By making the back-transformation, this projection indicates that until the 49th day Campo Grande would register 54 cases of the COVID-19 disease.
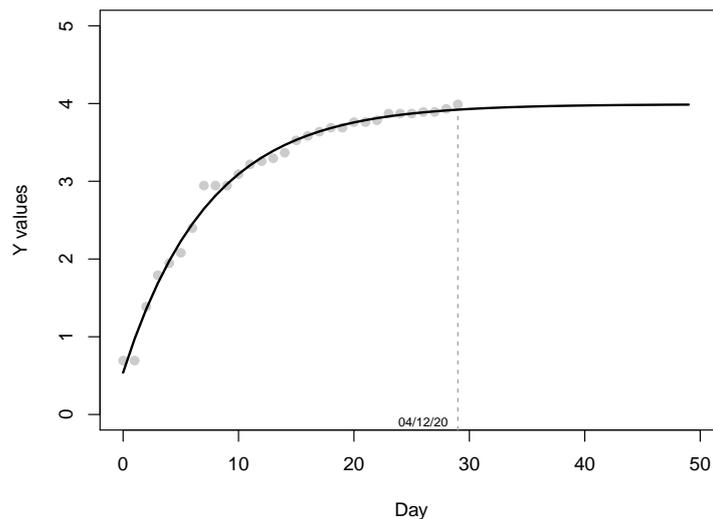

Figure 5 - Observed data and fitted model for D_1.

However, as can be viewed in Figure 6, the recorded number of cases on the ten days after the date 04/12/20 (from 04/13/20 to 04/22/20) has led to an accumulated number of cases (log-transformed) that did not follow the projections done by the fitted model. Due to this, we insert these ten accumulated values into the dataset D_1 and update the model.
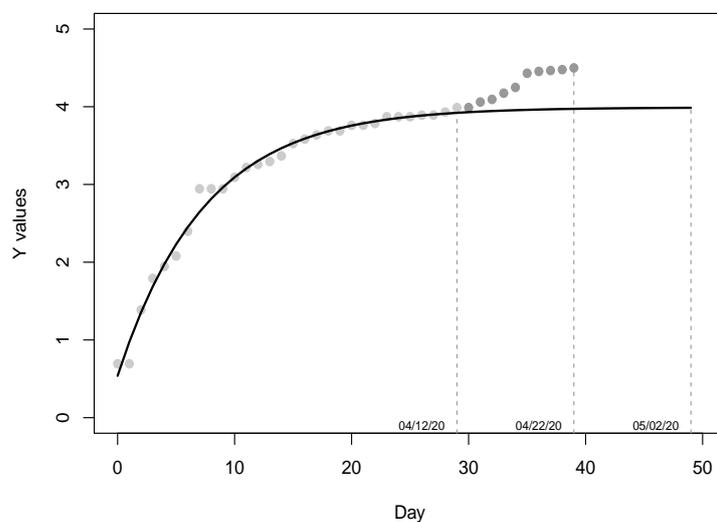
Figure 6 - Fitted model and data from 04/13/20 to 04/22/20.

Figure 7(a) shows the graphic of the accumulated values and the curve of the updated model. The MSE of the updated model is 0.0252. That is, an MSE value greater than the first fitted model. Besides, the accumulated number of cases on 04/22/20 (97 cases) is greater than the estimate of the updated model for the maximum number of cases (77 cases). This result made us to conjecture that a new period with a different growth rate from the first 30 days could be beginning.

In order to empirically verify our conjecture, we registered the accumulated values in the next 10 days after the date 04/22/20 (from 04/23/20 to 05/02/20) and plot these values in the same cartesian plan of the updated model, as shown by Figure 7(b). As one can note, our conjecture is very plausible. Due to this, hereafter we adopt the fit of a piecewise growth model. The information on the beginning of a new period of the pandemic's growth, with a more aggressive growth rate, was described in a technical report and repassed to HSC on 05/02/20.
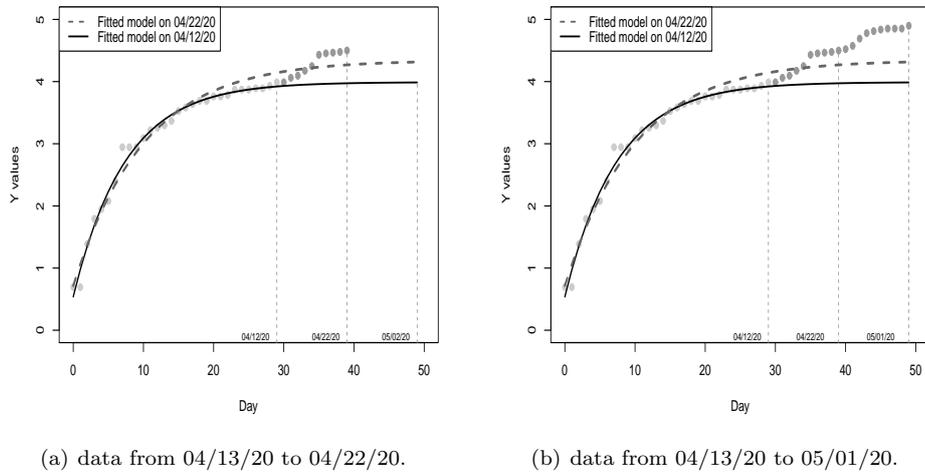
(a) data from 04/13/20 to 04/22/20.

(b) data from 04/13/20 to 05/01/20.

Figure 7 - Fitted models and data from 04/13/20 to 05/01/20.

3.2 Model fitting 2

Let now the recorded number of cases in the first 60 days since the first case (from 03/14/20 to 05/12/20) in an accumulated way and log-transformed. Consider the separation of the this dataset into two sub-datasets as follows: $D_1 = \{y_0, \ldots, y_{28}\}$ and $D_2 = \{y_{29}, \ldots, y_{59}\}$. The choice o the 28th day to separate the two sub-datasets was based on the criterion of smallest MSE among ten fitted models. Please, see Appendix 3 of the SM for more details.

Defined the two sub-datasets, we fit the three growth model for each one of the sub-datasets. Table 4 shows the MSE, AIC and BIC values for the three fitted growth models for D_1 and D_2. The smallest values are highlighted in bold. For both datasets, the Log-Gompertz is the best model.

Table 4 - MSE, AIC and BIC values

Model	Dataset D_1			Dataset D_2		
	MSE	AIC	BIC	MSE	AIC	BIC
Log-Exponential	0.1667	35.2897	39.2863	0.0100	-50.4207	-46.0235
Log-Logistic	0.0170	-26.6779	-21.3490	0.0013	-112.7321	-106.8692
Log-Gompertz	**0.0097**	**-42.2755**	**-36.9467**	**0.0012**	**-115.8826**	**-110.0197**

Table 5 shows the estimates and the standard errors for parameters of the fitted Log-Gompertz model for D_1 and D_2. The fitted model is given by the following

piecewise model

$$\hat{Y}_t = \begin{cases} 3.9738 - 3.4417 \exp\{-0.1365t\} & \text{, for } 0 \le t < 29; \\ 5.2288 - 1.3354 \exp\{-0.0686t\} & \text{, for } t \ge 29; \end{cases} \qquad (7)$$

Table 5 - Estimates for model parameters

Values	Model parameters for D_1			Model parameters for D_2		
	$log(\alpha_1)$	α_2	α_3	α_1	α_2	α_3
Estimates	3.9738	3.4417	0.1365	5.2288	1.3354	0.0686
Standard error	0.0525	0.0723	0.0077	0.0441	0.0376	0.0055

Figure 8(a) shows the graphic of the values of the sub-datasets D_1 and D_2 and the fitted model (black line), for a period of 80 days (from day 0 to day 79), being 60 of fitting and 20 of projection. The 79th day represents the date 06/01/20. The MSE of the fitted piecewise model is 0.0019.

However, as can be viewed in Figure 8(b), the registered values in the 20 days after 05/12/20 indicate the beginning of a third growth period. Again we informe the HSC on the beginning of a third growth period of the pandemic with a more aggressive growth rate. Then, we consider the following model fitting 3.
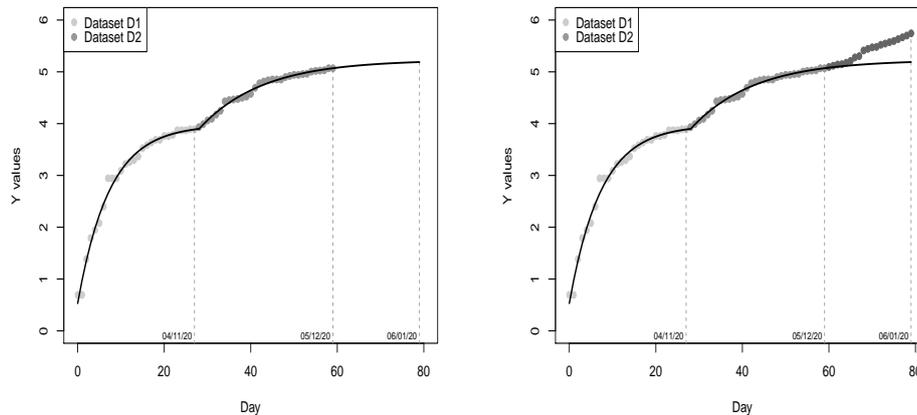
(a) data from 04/13/20 to 04/22/20.   (b) Data from 04/13/20 to 06/01/20.

Figure 8 - Fitted model and data from 04/13/20 to 06/01/20.

3.3 Model fitting 3

Consider now the accumulated number of cases (log-transformed) recorded in the period from 03/14/20 (day 0) to 06/11/20 (day 89). Let $D_1 = \{y_0, \dots, y_{28}\}$

and set up $D_2 = \{y_{29}, \ldots, y_{60}\}$ and $D_3 = \{61, \ldots, 89\}$. The choice o the 60th day to separate the two sub-datasets was also based on the criterion of smallest MSE as done for separating the sub-datasets D_1 and D_2. Please, see Appendix 3 for more details.

The fitted model for D_1 remains as describe in Equation (7). For the sub-dataset D_2, we fit the three growth models and again the log-Gompertz model presents the smaller MSE, AIC and BIC values. For the sub-dataset D_3 was not possible to fit the log-logistic and the log-Gompertz models due to the non-convergence of the estimation algorithm. Thus, a log-exponential model was fitted to the sub-dataset D_3. Table 6 shows the estimates and the standard errors for parameters of the fitted log-Gompertz model for D_2 and log-exponential model to D_3.

Table 6 - Estimates for model parameters

Values	Model parameters for D_2			Model parameters for D_3	
	$log(\alpha_1)$	α_2	α_3	α_1	α_2
Estimates	5.2337	1.3394	0.0681	5.1039	0.0361
Standard error	0.0411	0.0351	0.0052	0.0090	0.0006

The fitted model is given by the following piecewise model

$$\hat{Y}_t = \begin{cases} 3.9738 - 3.4417\exp\{-0.1365t\} & \text{, for } 0 \leq t < 29; \\ 5.2337 - 1.3394\exp\{-0.0681t\} & \text{, for } 29 \leq t < 61; \\ 5.1039 + 0.0361t & \text{, for } t \geq 61; \end{cases} \quad . \qquad (8)$$

The MSE of this fitted piecewise model is 0.0093. Figure 9(a) shows the graphic of the recorded values and the fitted model for a period of 139 days (from 03/14/20 to 07/31/20). Figure 9(b) shows the graphic of the fitted model and the accumulated values registered in the period from 06/11/20 to 07/31/20. As one can note, the registered values in this period do not follow the projection of the fitted model. Following the same procedure, we inform the HSC the beginning of the fourth growth period of the pandemic with a more aggressive growth rate. Then, we consider the following model fitting 4.
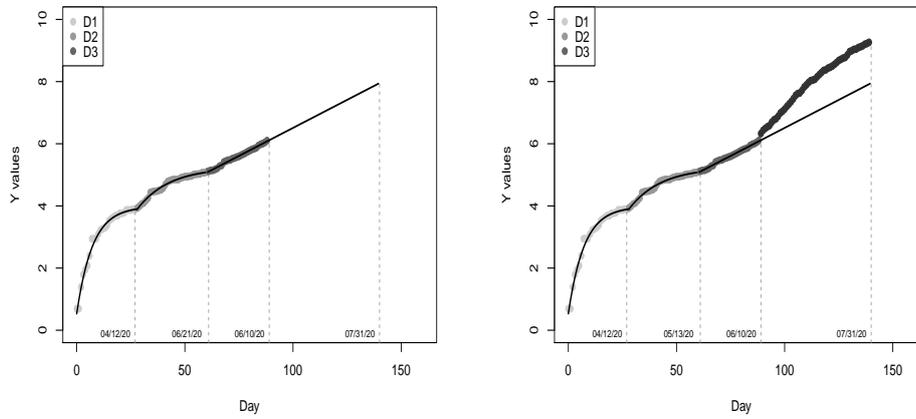
(a) data from 04/13/20 to 06/10/20.

(b) data from 04/13/20 to 04/22/20.

Figure 9 - Fitted model and data from 04/13/20 to 04/22/20.

3.4 Model fitting 4

For the fourth model fitting, consider the number of recorded cases in the period from 03/14/20 (day 0) to 07/31/20 (day 139). Let $D_1 = \{y_0, \dots, y_{28}\}$, $D_2 = \{y_{29}, \dots, y_{60}\}$, $D_3 = \{y_{61}, \dots, y_{88}\}$ and $D_4 = \{y_{89}, \dots, y_{139}\}$. Please, see Appendix 3 of the SM for details on the criterion used to separate the sub-datasets.

The fitted models for D_1, D_2 and D_3 remains as describe in Equation (8). We fit the three growth models for sub-dataset D_4. The Log-Gompertz model presents the smaller MSE, AIC and BIC values. Thus, the piecewise model has the following configuration: Log-Gompertz for D_1, Log-Gompertz for D_2, Log-exponential for D_3 and Log-Gompertz for D_4.

The fitted model is given by the following piecewise model

$$\hat{Y}_t = \begin{cases} 3.9738 - 3.4417 \exp\{-0.1365t\} & \text{, for } 0 \le t < 29; \\ 5.2337 - 1.3394 \exp\{-0.0681t\} & \text{, for } 29 \le t < 61; \\ 5.2647 + 0.0350t & \text{, for } 61 \le t < 89; \\ 11.5845 - 5.4009 \exp\{-0.0167t\} & \text{, for } t \ge 89; \end{cases}$$

The MSE of this fitted piecewise model is 0.0099. Figure 10 shows the graphic of the registered values and the fitted model for a period of 200 days (from 03/14/20 to 09/30/20). The inflection point coordinates are $(190, 11.5845)$. That is, the peak of the pandemic is projected for the 190th day (09/20/20) with $107,420$ cases.
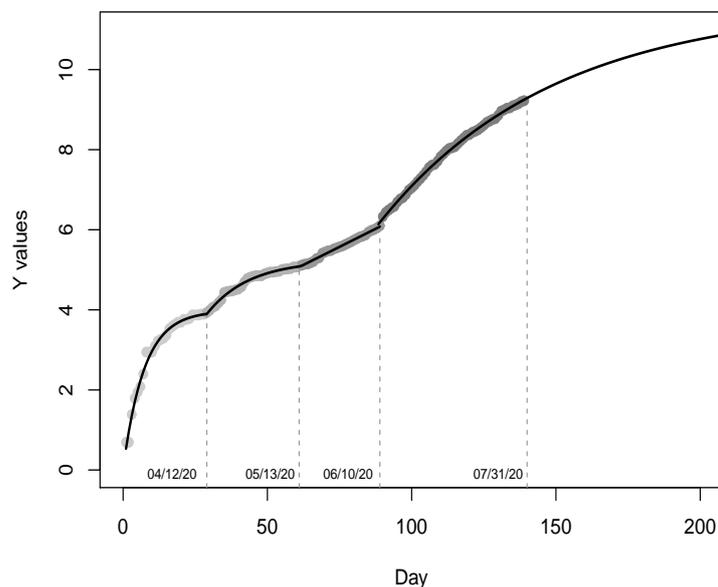
Figure 10 - Fitted model and data from 04/13/20 to 09/30/20.

Figure 11 shows the graphic of a model update run on the 154th day (08/15/20). As one can note, the registered values in the period from 08/01/20 to 08/15/20 have led to a flattening of the curve. The inflection point coordinates is $(168, 10.7469)$. That is, the peak of the pandemic is projected for the 168th day (08/25/20) with $46,486$ cases.

After the date 08/15/31, we update the model more five times, on 09/15/20, 09/30/20, 10/15/20, 10/31/20 and on 11/08/20. Figure 12 shows the curves of Figure 11 and the curve of the fitted model on the 239th day (11/08/20). As one can note, there was a significant flattening between the fitted model on 07/31/20 and the fitted model on 08/15/20. However, models fitted after 08/15/20 presented smaller flattening. Due to this, in order to maintain a good visualization of graphs, we opt to plot only the curves of these three fitted models. This last updated model has the following piecewise Equations

$$\hat{Y}_t = \begin{cases} 3.9738 - 3.4417 \exp\{-0.1365t\} & \text{, for } 0 \le t < 29; \\ 5.2842 - 1.3905 \exp\{-0.0637t\} & \text{, for } 29 \le t < 61; \\ 5.2647 + 0.0350t & \text{, for } 61 \le t < 89; \\ 10.7074 - 4.6035 \exp\{-0.0222t\} & \text{, for } t \ge 89; \end{cases} \tag{9}$$

The peak is estimated to the 158th day (08/19/20). The projection for the maximum number of cases is $44,685$. Appendix 4 present a discussion on the residuals analysis of the fitted model. The information on the flattening of the curve was described in a technical report and repassed to HSC on 08/12/20.
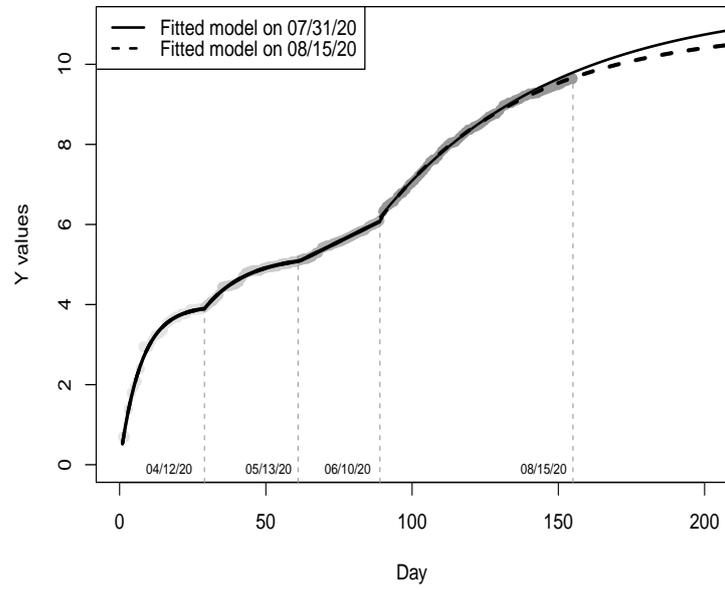
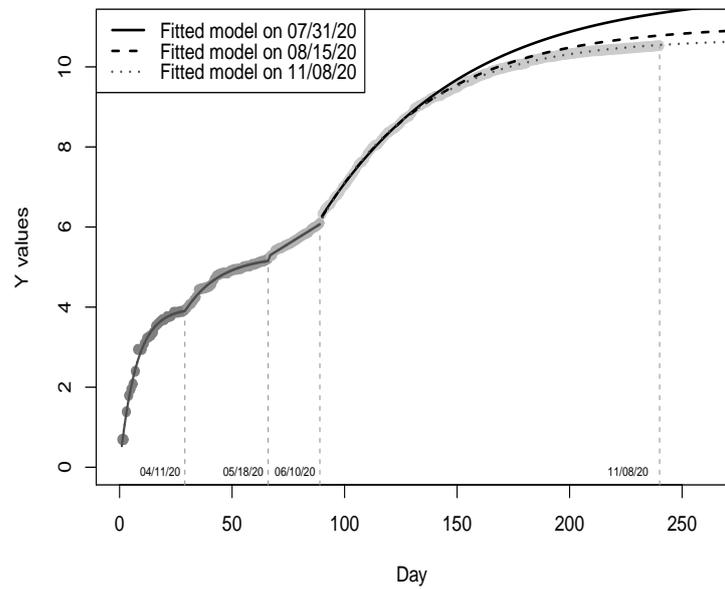Figure 11 - Fitted model and data from 04/13/20 to 08/15/20.

Figure 12 - Fitted model and data from 04/13/20 to 11/08/20.

4 Projections

In addition to estimates for the peak of the pandemic, the fitted model allows to project the number of cases that will be registered in next days. For example, Table 7 shows the projections for the seven days after the date 11/08/20. This Table also shows the registered values, the error (absolute) and the percent error (absolute) of the projections in relation to the real values. The biggest percentage error was 1.4507%.

Table 7 - Projections and percentage error for period from 11/09/20 to 11/16/20

Date	11/09	11/10	11/12	11/13	11/14	11/15	11/16
Projection	37,634	37,776	37,915	38,053	38,187	38,320	38,450
Real value	37,551	37,726	37,885	38,140	38,484	38,781	39,016
error	83	50	30	87	297	461	566
% error	0.2210	0.1325	0.0792	0.2281	0.7717	1.1887	1.4507

Another interest is to project the number of individuals who will need care in clinical care units and in intensive care units. In order to get these projections, we consider the projected values by the fitted model and a moving-sum procedure, in which, the number of clinical units occupied on a day d_1 is considered available after 5 days. For the intensive care units, a unit is considered available after 14 days. In addition, we consider that 10% of the projected values by the fitted model will needs care in clinical units and 3% in intensive units. Both percentual values were fixed according to the percentage observed in the data.

Figure 13 shows the graphic with the projections. According to the projections, on the 240th day (08/11/20) were expected 65 individuals in care in clinical units and 50 individuals in intensive care units. These values are highlighted in the graphics by the symbol •. The registered values were 63 in clinical care units and 52 in intensive care units.

In addition, the projections show that the peak for the number of patients who will need care in clinical and intensive care units has already happened. The peak for the number of patients who would need care in clinical units was estimated for the 169th day (08/30/20). And the peak for the number of patients who would need care in intensive care units was estimated for the 166th day (08/27/20). Since the health public system of Campo Grande city has available 341 clinical units and 157 intensive care units, these results indicated no evidence for the collapse of the public health system. These pieces of information were sent to HDC on 09/11/20.

(a) Clinical.
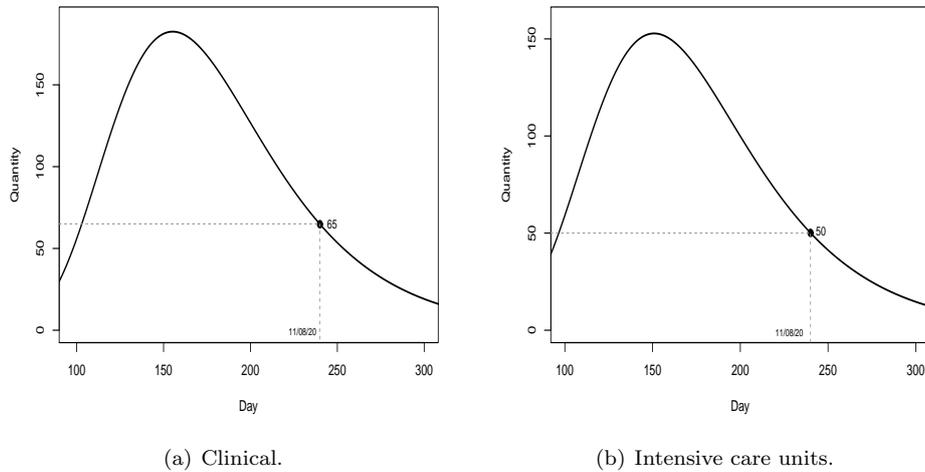
(b) Intensive care units.

Figure 13 - Projections.

5   Final Remarks

In this paper, we describe a study on the evolution of the COVID-19 pandemic in the city of Campo Grande, MS, Brazil. The aim of this study was to find out a non-linear growth model for the accumulated number of COVID-19 cases. However, in the course of the analysis, the data have presented temporal heterogeneity and due to this, we adopt the fitting of a piecewise growth model. The main advantage of this procedure is that it allowed us to obtain a model with the smallest MSE among the tested models.

The modeling procedure was developed by separating the accumulated number of cases registered in a period of 240 days into four sub-datasets. Then, we fitted the three non-linear growth models for each sub-dataset: Exponential, Logistic and Gompertz. In order to select the best model for each sub-dataset, we consider as criteria the MSE, the AIC and BIC. The fitted model has the configuration: Gompertz, Gompertz, Exponential and Gompertz. The resulting piecewise model project the peak of the pandemic for the 158th day (08/19/20). In addition, the projections do not show evidence for the collapse of the public health system of the city.

From a practical viewpoint, these results are very important since the prediction for the peak of the pandemic and projections that indicates the possibility or not for the collapse of the public health system may help government-agents to make a decision in order to try to reduce the transmission of the disease and to avoid the collapse of the health public system. Besides, the results may be used by public agents to show the population the importance to follow the recommendations of specialists from the health field and to maintain social isolation whenever possible.

In the course of these 240 days, were sent to HSC 21 technical reports describing the main results and the projections for the peak and possibility of the collapse of the clinical and intensive care units. In addition, the results of this study were used by many media to inform the population of Campo Grande on the evolution of the pandemic and to emphasize the importance of the collaboration of the population to reduce the transmission.

Although the article does not present any innovative mathematical and/or statistical result, the results showed to be very important for government agents of the city of Campo Grande, especially to justify actions to combat the proliferation of the disease. In addition, due to the need for a quick response, the use of consolidated models and well described in the literature proved to be the best alternative for the moment; since the development and validation of new models, would require a more long period of research. The computational codes used for fitting the models are in the R language and can be obtained by e-mail to the authors.

Acknowledgments

▪ *RESUMO: Em dezembro de 2019, um novo coronavírus foi descoberto na cidade de Wuhan, na China. A Organização Mundial da Saúde nomeou oficialmente este coronavírus como COVID-19. Desde sua descoberta, o vírus se espalhou rapidamente pelo mundo e atualmente é um dos principais problemas de saúde, causando um enorme ônus social e econômico. Devido a isto, há um grande interesse em modelos matemáticos capazes de projetar a evolução da doença em países, estados e/ou municípios. Esse interesse se deve principalmente ao fato de que as projeções podem auxiliar os agentes governamentais na tomada de decisões em relação à prevenção da doença. Utilizando este argumento, a secretaria de saúde do município de Campo Grande solicitou à UFMS o desenvolvimento de um estudo matemático para projetar a evolução da doença no município. Neste artigo, descrevemos o procedimento de modelagem usado para ajustar um modelo de crescimento por partes para o número acumulado de casos registrados na cidade. A partir do modelo ajustado, estimamos a data em que o pico da pandemia é atingido e projetamos o número de pacientes que necessitarão de tratamento em unidades de terapia intensiva. Semanalmente, enviamos a secretária de saúde da cidade um relatório técnico descrevendo os principais resultados.*

▪ *PALAVRAS-CHAVE: Modelos de crescimento, Estimação, Mínimos quadrados não linear, Seleção de modelos.*

References

ABRAMOWITZ, M.; STEGUN, I. M. *Handbook of Mathematical Functions.* Dover, New York, 1965. 1046p.

AKAIKE, H. A. New look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.19, p.716-723, 1974.

BLUMBERG, A. A. Logistic Growth Rate Functions. *Journal of Theoretical Biology*, v.21, n.1, p.42-44, 1968.

BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrica*, v.52, p.345-370, 1987.

BUDIMULYATI, L. S.; NORR, R. R.; SAEFUDDIN, A.; TALIB, C. Comparison on accuracy of logistic, gompertz and von bertalanffy models in predicting growth of new born calf until first mating of holstein friesian heifers. *Journal of the Indonesian Tropical Animal Agriculture*, v.37, n.3, p.151-160, 2012.

CRAN R (2019). R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://R-project.org.

FLORENT, B.; RITZ, C.; CHARLES, M.; BRUSTSCHE, J. P.; FLANDROIS, P.; DELIGNETTE-MULLER, M. L. A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software*, v.66, 1-21, 2015.

GOMPERTZ, B. 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. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, v.182, p.513-585, 1825.

HSIEH, Y. H. Temporal patterns and geographic heterogeneity of zika virus (zikv) outbreaks in french polynesia and central america. *PeerJ*, v.5, e3015, 2017.

MEYER, P. S. A primer on logistic growth and substitution: The mathematics of the loglet lab software. *Technological Forecasting and Social Change*, v.61, 247-271, 1999.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SAKAR, D. R Core Team, Linear and Nonlinear Mixed. *R package version 3.1-147*, 2020. URL https://CRAN.R-project.org/package-nlme.

SCHWARZ, G. E. Estimating the dimension of a model. *The Annals of Statistics*, v.6, p.461-464, 1978).

THOMSON, D. G. *Blueprint to a Billion: 7 Essentials to Achieve Exponential Growth.* John Wiley & Sons, 2005. 288p.

VIEIRA, S.; HOFFMANN, R. Comparison of the logistic and the gompertz growth functions considering additive and multiplicative error terms. *Applied Statistica*, v.26, 143-148, 1977.

WINSOR, C. P. THE GOMPERTZ CURVE AS A GROWTH CURVE. *Proceedings of the national academy of sciences*, v.18, p.1-17, 1932.

ZHAO, S.; MUSA, S. S.; FU H.; HE, D.; QIN, J. Simple framework for real-time forecast in a data-limited situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. *Parasites Vectors*, v.12, n.1, p.12-344, 2019.

Supplementary Material of the Paper

"A piecewise growth model for modeling the accumulated number of COVID-19 cases in the city of Campo Grande"

Appendix 1: Moving-average for number of death

Until 11/08/20, 698 people have died due to COVID-19 in Campo Grande city. Figure 14 shows the moving-average of seven days for the number of death. In the last four days, there was a reduction in the moving-average value. On 11/08/20 the moving-average value was 2 death. Compared to the value of seven days ago (2.14 on 11/01/20), there was a reduction of 7%. Compared to fourteen days ago (3.86 on 10/25/20) the reduction was 58.38%.
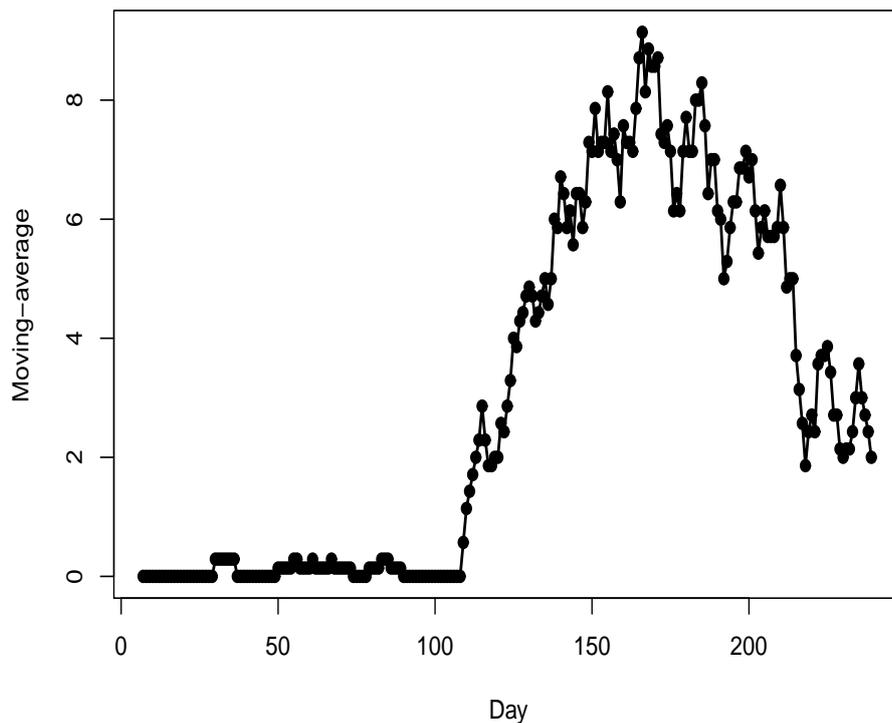

Figure 14 - Moving-average for number of death.

Appendix 2: Growth models

In this appendix, we present the graphics of the growth models described in Section 2 of the paper. Figure 15 shows the graphics of the exponential and log-exponential models. Increasing the value of the α_2 more inclined is the curve. As a characteristic, the graphic of this model shows unlimited growth.
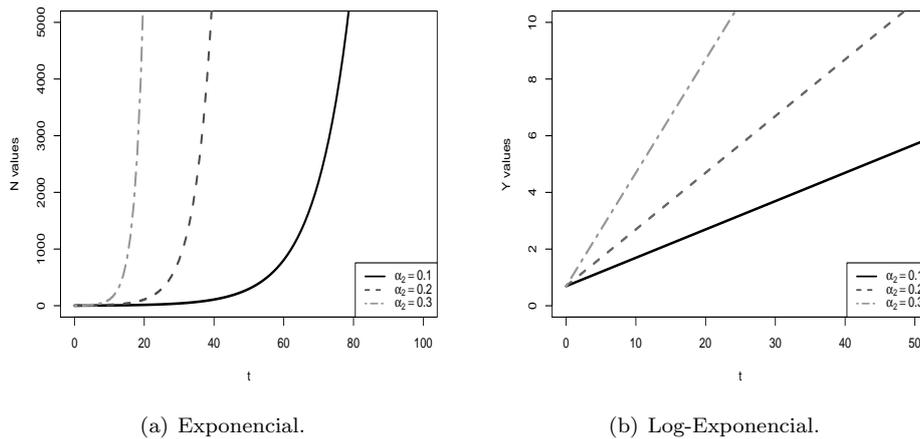

(a) Exponencial.

(b) Log-Exponencial.

Figure 15 - Exponencial and Log-Exponencial models.

Figure 16 shows the graphics of the Logistic and log-Logistic models. As a characteristic, the graphic of this model has an S-shape. Similar to the exponential model, by increasing the value of the parameter α_3 more inclined is the curve. In that Figure, the symbols ● represents the inflection point.
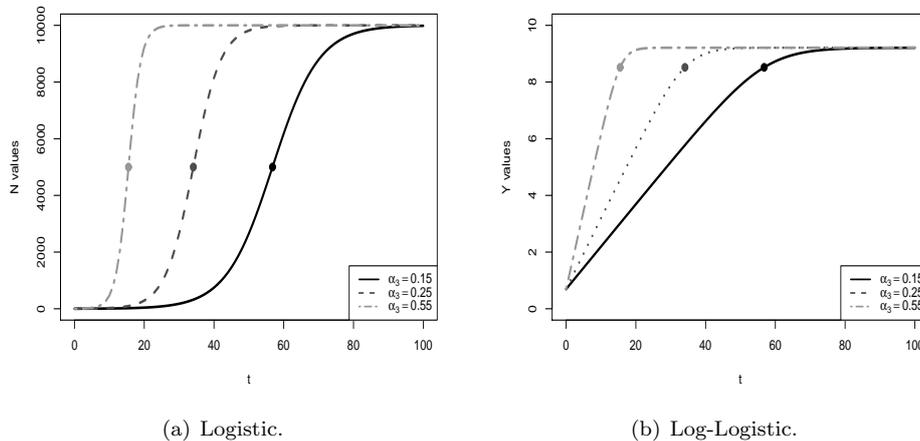

(a) Logistic.

(b) Log-Logistic.

Figure 16 - Logístic and Log-Logistic models.

Figure 17 shows the graphics of the Gompertz and log-Gompertz models. As a characteristic, the graphic of this model has an S-shape. Increasing the value of the parameter α_3 more inclined is the curve. In that Figure, the symbols $\bullet$ represents the inflection point.
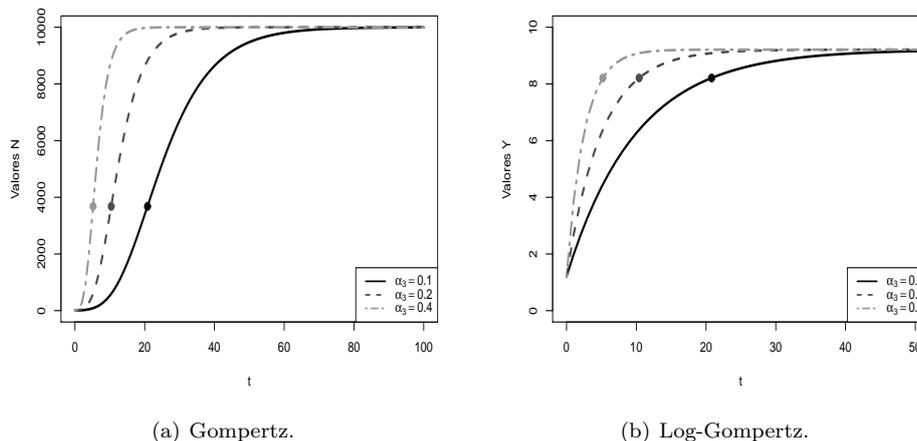

(a) Gompertz.

(b) Log-Gompertz.

Figure 17 - Gompertz and Log-Gompertz models.

Appendix 3: Sub-datasets

In order to get a point d that best separates the sub-datasets D_1 and D_2, we adopt the following procedure. Let $\mathcal{G} = \{25, \ldots, 34\}$ be a grid from 25 to 34 with increments of size 1. Then, we define the following 10 scenery: $D_1 = \{y_0, \ldots, y_d\}$ and $D_2\{y_{d+1}, \ldots, y_{59}\}$, for $d \in \mathcal{G}$.

For each one of the 10 scenery, we fit the three growth models to sub-datasets D_1 and D_2. We select the best model by using as a criterion the AIC and BIC. For all 10 scenery considered, the fitted model has the following configuration: Gompertz for D_1 and Gompertz for D_2.

Table 8 shows the MSE values from ten fitted models. The point $d = 28$ has lead to a fitted model with the smallest MSE value. Due to this, we set up $d = 28$ as the separation point for D_1 and D_2.

Table 8 - Mean square error for 10 models fitted for D_1 and D_2

Point d	25	26	27	28	29	30	31	32	33	34
MSE	0.0058	0.0056	0.0054	**0.0052**	0.0055	0.0056	0.0060	0.0066	0.0068	0.0084

To separate sub-datasets D_2 and D_3, we follow the same procedure described above for $d \in \mathcal{G} = \{55, \ldots, 64\}$. As can be seen, the point $d = 60$ has lead to a fitted model with the smallest MSE value. As shown by Figure 9(b) there is a clear separation on the day 88. Due to this, we set $d = 88$ for separating D_3 and D_4.

Table 9 - Mean square error for 10 models fitted for D_2 and D_3

Point d	55	56	57	58	59	60	61	62	63	64
MSE	0.0040	0.0040	0.0039	0.0039	0.0038	0.0039	0.0039	0.0041	0.0041	0.0042

Appendix 4: Residuals Analysis

After fitting the model, it is necessary to verify that the assumptions made are satisfied. In order to verify the normality assumption, we adopt the graphical method based on the qq-plot. This method consists of the plot the percentiles of the sample against the percentiles expected by the adjustment of a normal distribution. Figure 18(a) shows the qq-plot graphic for the residuals of the fitted model (see Equation 9). We also apply the Shapiro normality test, getting a p-value equal to 0.0959. That is, for a significance level $\{0.01, 0.05\}$ the normality assumption is not rejected. Figure 18(b) shows the graphic of the residuals against order of the data. As one can note, there is no reason to doubt the independence and homoscedasticity assumptions.
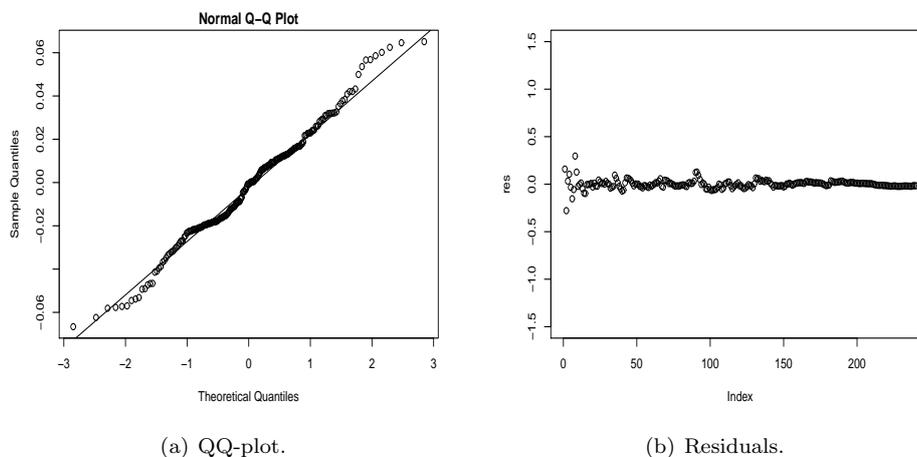

(a) QQ-plot.

(b) Residuals.

Figure 18 - QQ-plot and Residuals versus order.