

GENERALIZED GROWTH CURVE MODEL FOR COVID-19 IN BRAZILIAN STATES

Magali Teresopolis Reis AMARAL¹
Katiane Silva CONCEIÇÃO²
Marinho Gomes de ANDRADE²
Carlos Roberto PADOVANI³

- **ABSTRACT:** The present paper consists of using the Chapman-Richard generalized growth model to functionally relate the number of people infected by COVID-19 with the number of days. The objective of this work is to estimate the instant that the number of infected people stops growing using the dataset of the accumulated amount of infected. For this propose, one conducted a comparative study of the performances of three models of Richard in eight Brazilian States. In the methodological context, the Gauss Newton procedure was used to estimate the parameters. In addition, selection criteria of the models were used to select the one that best fits the dataset. The methodology used allowed consistent estimates of the number of people infected by COVID-19 as a function of time and, consequently, it was possible to conclude that the projections provided by the growth curves point to a scenario of general contamination acceleration. Besides, the models predict that the epidemic is close to reaching its peak in Amazonas, Ceará, Maranhão, Pernambuco, and São Paulo States.
- **KEYWORDS:** Corona virus; generalized Richard model; growth curves; Gauss Newton method.

¹Universidade Estadual de Feira de Santana - UEFS, Departamento de Ciências Exatas, CEP: 44036-900, Feira de Santana, BA, Brasil. E-mail: *mteresopolis@uefs.br*

²Universidade de São Paulo - USP, Instituto de Ciências Matemáticas e Computação, Departamento de Matemática Aplicada e Estatística, Caixa Postal 668, CEP: 13560-970, São Carlos, SP, Brasil. E-mail: *katiane@icmc.usp.br*; *marinho@icmc.usp.br*

³Universidade Estadual Paulista - UNESP, Instituto de Biociências, Departamento de Bioestatística, CEP: 18618-970, Botucatu, SP, Brasil. E-mail: *cr.padovani@unesp.br*

1 Introduction

In Wuhan, Hubei province, in December 2019, an outbreak of pneumonia caused by a new coronavirus occurred and it spread rapidly around the world. It has since been identified as a zoonotic coronavirus, similar to SARS and MERS coronaviruses and it was named COVID-19 (LIU *et al.*, 2020). Due to the high epidemic spread power of the virus in a short period of time, on the 30th of January, 2020 the World Health Organization (WHO) declared the outbreak of SARS-CoV-2, a public health emergency of international concern (ZHENG *et al.*, 2020).

In general, COVID-19 is an acute disease with the symptoms of severe infection and possibly curable, but it can also be deadly, with a 2% case fatality rate (XU *et al.*, 2020). In Brazil, the mortality rate from the virus is reported at 4.2%.

In this sense, a new approach is proposed in this epidemiological study to model the evolution of the number of people infected, and the maximum number of people infected with COVID-19, through the Chapman-Richard non-linear model, considering eight Brazilian States. We used as outcome variable, the cumulative number of infected people up to 88 days after the appearance of the first case of the disease.

When we consider the accumulated processes, the objective is looking for sigmoidal functions that best fit the curves. Then, it is appropriate to consider a function $P(t)$ with a sigmoid form, ideally originating in $P(0)$, with a point of inflection occurring early in the one third of the stage before approaching a maximum value, with an asymptote that represents the end of the growth process. For more severe pandemics, this change in curvature can occur closer to the peak point.

Therefore, due to the heterogeneity of the Brazilian population, one decided to build the models by States. However, it is known that there is a high number of underreporting cases in Brazil, due to the unavailability of tests, which can cause problems of estimation of the parameters and prediction of these models.

In this paper, the procedure the Gauss Newton estimation was used to obtain the estimates and inference of the parameters of the Chapman-Richard growth model, in which the value of form parameter m was fixed at 0.5, 1.0, 1.5 and 2.0. In addition, selection criteria of the models were used to select the one that best fits dataset.

It is important to note that, in these models, it was not possible to incorporate some factors compatible with reality, such as isolation rate, disease incubation, among others. It is important to note that, in these models, it was not possible to incorporate some factors compatible with reality, such as isolation rate, disease incubation, among others. The presentation and discussion of the methodology of the statistical analysis study were developed with data available at <https://covid.saude.gov.br/>, until May 23rd, 2020. All analyzes were performed using the computational programming environment R CORE TEAM (2009), whose free version can be found at www.r-project.org/.

This paper is organized as follows. Section 2 defines the methods of estimation

and criteria for selecting models. Real data applications considering the number of cases accumulated by COVID-19 in Brazilian States are illustrated in Section 3. Finally, Section 4 draws some conclusions on the relevant results achieved.

2 Estimation methods

In order to model the phenomenon of the growth curve of the number of people infected by COVID-19, one opted for the Chapman-Richard non-linear model in the genuine and modified forms. The model considered is a well-known generalization of the logistic and Gompertz models. Its main advantage is that this model generally provides a more realistic description of various phenomena, however, it is little explored because it presents several difficulties in the procedure of non-linear estimation (AMARAL, 2009).

The Chapman-Richard nonlinear model is described by:

$$P(t) = a(1 - ce^{-kt})^m, \quad (1)$$

where $P(t)$ represents the average number of infected individuals; t is the number of days; a is the parameter that represents the asymptotic value of the model, i.e. the point of growth of cases free of seasonal variations; c is a constant without biological interpretation that is related to the initial cases of COVID-19; k is the growth rate of the variable of interest that determines the efficiency of the disease's growth; and m the parameter which defines the shape of the model's curve and, consequently, the point that it starts to grow less efficiently.

As a consequence, for the case of COVID-19, the parameters k and c are determined by performing several fittings on the growth rate-trends of infection capacity of the viruses that mainly affect the respiratory system (SONNINO, 2020). Figure 1 shows the influence on the growth of the curve for the different values of the parameters a and k respectively. In order to construct these curves, one assumes that the shape parameter takes on the following values $a = 350, 370, 390, 410, 430$ and 450 and the growth rate $k = 0.030, 0.035, 0.040, 0.045$ and 0.050 .

It is worth mentioning that Von Bertalanffy and Brody models are special cases of Chapman-Richard model for the inflection point $m = 3$ and $m = 1$ respectively. Several authors have reported difficulties in fitting for the function of Richards (BATES & WATTS, 1988; BROWN *et al.*, 1976). Due to the difficulty of convergence of the parameters of the models, SARMENTO *et al.* (2006) reported that in applications using this model the convergence in the iterative process was not reached in approximately 50% of the fit attempts. Consequently, this difficulty was attributed to the need to estimate one more parameter with this model, and mainly due to the negative correlation between c and m . In this context, one will fit this model by fixing the values of the parameter m at 0.5, 1.0, 1.5 and 2.0; for the other parameters, adequate initial values were found, through which the convergence of the iterative procedure was achieved.

In many practical applications, the linearization procedure is adopted; this can be done by applying the logarithmic transformation in the model, $Y(t) = \ln(P(t))$.

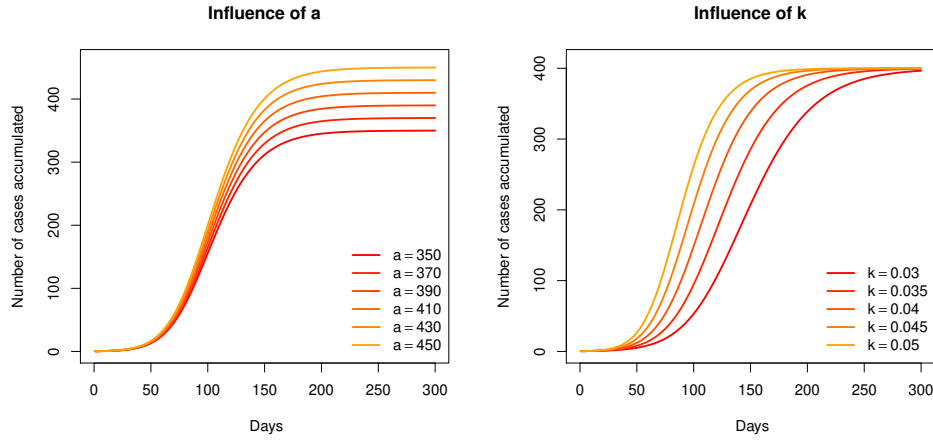


Figure 1: Growth curves fitted through Chapman-Richard model for different values of parameters a and k , with $m = 2$ and $c = 35$.

In this way, the general model can be written as:

$$Y(t) = h(t; \beta) + e_t, \quad (2)$$

where β can be defined as a vector of parameters, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T = (\ln(a), b, c, m)^T$, and e_t is an independent random error, normally distributed $N(0, \tau^{-1})$, with variance $\sigma^2 > 0$ and $\tau = 1/\sigma^2$ (FEKEDULEGN & MAC SIÚRTÁIN, 1999).

The transformation of a non-linear model into a linear model in the parameters facilitates the fit process; however, it implies unrealistic assumptions such as normality and homoscedasticity of the errors. For the general model described in (1), using the natural logarithm takes us to the model given by:

$$Y(t) = \ln(a) + m \ln(1 - ce^{-kt}) + e_{t_i}. \quad (3)$$

The use of transformations for linearization of non-linear models becomes more critical when the estimates obtained using the linearized models are made for the transformed parameters and not for the original parameters of the model. In these cases, information about the standard errors of the original parameters is lost and this results in great difficulties in inferring the original parameters and, in general, makes it impossible to calculate confidence intervals and test hypotheses about the original parameters of the model (a broader approach of this model). For more details, see MYERS (1990).

2.1 Gauss-Newton method

The Gauss-Newton method is a particular case of the weighted least squares method, also known as the linearization method. This method uses a Taylor series

expansion to approximate the nonlinear regression model with linear terms and then applies ordinary least squares to estimate the parameters. Iterations of these steps generally lead to a solution to the problem of nonlinear regression.

The starting point of the Gauss-Newton method is to find initial values for the parameters $\beta_0, \beta_1, \dots, \beta_p$, denoted by $\beta_0^{(0)}, \beta_1^{(0)}, \dots, \beta_p^{(0)}$. These initial values can be obtained by means of previous studies known in the literature (AMARAL, 2009). With the initial values of the parameters, the expected function $h(x, \beta)$ is approached to the desired power by linear terms of the Taylor series expansion, around the initial values $\beta^{(0)}$. Thus, the general solution for the i -th case is given by:

$$h(t_i; \beta) \cong h(t_i; \beta^{(0)}) + \sum_{j=1}^p \frac{\partial h(t_i; \beta^{(0)})}{\partial \beta_j} (\beta_j - \beta_j^{(0)}), \quad \forall j = 1, \dots, p. \quad (4)$$

Note that the derivatives of $h(t_i, \beta)$ are valued at $\beta_j = \beta_j^{(0)}$, and $\beta_j - \beta_j^{(0)}$ represents the difference between the true parameters of the regression and their initial estimates. Thus, the regression coefficients represent a correction that must be made in the initial regression coefficients.

In matrix notation, $h(t; \beta^{(0)}) = \{h(t_i; \beta^{(0)}), i = 1, \dots, n\}$ is a vector of order $(n \times 1)$, $\gamma = \{(\beta_j - \beta_j^{(0)}), j = 1, \dots, p\}$ has dimension $(p \times 1)$ and the derivatives $\frac{\partial h(t_i; \beta^{(0)})}{\partial \beta_j}$, with $i = 1, \dots, n$ and $j = 1, \dots, p$, form a matrix W of dimension $(n \times p)$.

Passing $h(t_i; \beta^{(0)})$ to the left side and denoting $\gamma_j = \beta_j - \beta_j^{(0)}$, we have the equation (4) rewritten as:

$$\mathbf{y} - h(\mathbf{t}; \beta^{(0)}) = \gamma W, \quad (5)$$

where $\mathbf{y}$ is a vector with $y_i = h(t_i; \beta)$.

Therefore, it is possible to estimate the parameters β by the method of ordinary least squares, considering as initial value of $\gamma = (\gamma_1, \dots, \gamma_p)^T$ obtained from the equation (5):

$$\hat{\gamma}^{(0)} = (W^T W)^{-1} W^T (\mathbf{y} - h(\mathbf{t}; \beta^{(0)})),$$

providing a first estimate for the parameter vector $\beta = (\beta_1, \dots, \beta_p)^T$, given by:

$$\hat{\beta}_j^{(1)} = \hat{\beta}_j^{(0)} + \hat{\gamma}_j^{(0)}, \quad \forall j = 1, \dots, p.$$

One can formally affirm that the vector of the estimates of s -th iteration, defining by $\beta^{(s)} = (\beta_1^{(s)}, \dots, \beta_p^{(s)})^T$, it is given by $\hat{\beta}^{(s)} = \hat{\beta}^{(s-1)} + \hat{\gamma}^{(s-1)}$, that is:

$$\hat{\beta}^{(s)} = \hat{\beta}^{(s-1)} + (W^{(s-1)T} W^{(s-1)})^{-1} W^{(s-1)T} (\mathbf{y} - h(\mathbf{t}; \beta^{(s-1)})). \quad (6)$$

At this point, verifying the corrected regression coefficients represents an improvement in the appropriate direction. Any inference about the parameter

estimates from the Gauss-Newton algorithm is based on the covariance matrix asymptotic regression, given by $\sigma^2(W^TW)^{-1}$. Thus, we have:

$$Var(\hat{\beta}^{(s)}) = diag\left\{\hat{\sigma}^2(W^{(s)T}W^{(s)})^{-1}\right\}, \quad (7)$$

where $\hat{\sigma}^2$ is the mean quadratic residue given by

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_i - h(t_i; \hat{\beta}))^2. \quad (8)$$

We denote by $SS_R^{(s)}$ the sum of the squares of the residuals in the iteration s . This sum of squares in the s -th iteration is given by:

$$SS_R^{(s)} = \sum_{i=1}^n (y_i - h(t_i; \hat{\beta}^{(s)}))^2.$$

and it is updated to the last iteration.

If the Gauss-Newton algorithm is in the right direction, $SS_R^{(s)}$ must be less than $SS_R^{(s-1)}$. In this case, the process is repeated until the desired convergence is verified. It is worth mentioning that the choice of initial estimates in the Gauss-Newton method is very important, as a bad choice can result in a very large number of iterations, often not converging.

2.1.1 The use of fractional increments

A drawback of the Gauss-Newton procedure is the fact that, in some practical problems, the increment in the γ can be very small in some cases causing very slow convergence. In more severe situations, the algorithm may follow in the wrong direction and not converge. To overcome these difficulties, an increment strategy presented in MYERS (1990) is given by:

Gauss-Newton algorithm with fractional increments

1. Use $(W^{(s-1)T}W^{(s-1)})^{-1}W^{(s-1)T}(\mathbf{y} - h(t; \hat{\beta}^{(s-1)})) = \hat{\gamma}^{(s-1)}$ to calculate the standard Gauss-Newton increment for iterations $s = 1, 2, \dots$;
2. Calculate $\hat{\beta}^{(s)} = \hat{\beta}^{(s-1)} + \hat{\gamma}^{(s-1)}$, as standard procedure suggests;
3. Se $SS_R^{(s)} < SS_R^{(s-1)}$ continue on to the next iteration using $\hat{\beta}^{(s)}$;
4. Se $SS_R^{(s)} > SS_R^{(s-1)}$ go to step 2; use $\hat{\gamma}^{(s-1)}/2$ as an increment vector;
5. Make $s \leftarrow s + 1$ and repeat the steps (2)-(4).

2.2 Practical aspects

To implement the Gauss-Newton algorithm, we consider the model and the Taylor series expansion given in equations (3) and (4) to calculate the elements w_{ij} ($i = 1, 2, \dots, n$ and $j = 1, 2, 3, 4$) of the matrix W . Table 1 presents the elements w_{ij} of the matrix W of the Chapman-Richards model.

Table 1: Derivatives of the linearized Chapman-Richards model used in the Gauss-Newton algorithm

Parameters	Derived
a	$\frac{\partial h(t_i; \beta)}{\partial a} = \frac{1}{a}$
c	$\frac{\partial h(t_i; \beta)}{\partial c} = -\frac{me^{-kt_i}}{1 - ce^{-kt_i}}$
k	$\frac{\partial h(t_i; \beta)}{\partial k} = \frac{mct_i e^{-kt_i}}{1 - ce^{-kt_i}}$
m	$\frac{\partial h(t_i; \beta)}{\partial m} = \ln(1 - ce^{-kt_i})$

2.3 Model selection

The choice of criteria to select the growth model that best fits the dataset is an important step in the proposed methodology. Those criteria become necessary when several models are fitted to the same dataset, because the model adopted must be the one that best predicts the dependent variable according to the studied biological reality (AMARAL & PADOVANI, 2020). Given the above, for illustrating the phenomenon studied, we adopt two statistical indicators, the Akaike information criterion (AIC) (AKAIKE, 1974) and the Schwarz Bayesian criterion (SBC) (SCHWARZ, 1978), that under the hypothesis of normality and independence residual, are given respectively by

$$AIC = n[\ln(2\pi SS_R^{(s)}) + 1] + 2(p + 1)$$

and

$$BIC = n \ln \left(\frac{SS_R^{(s)}}{n} \right) + \ln(n)(p + 1),$$

where n is the sample size and p is the number of free parameters.

3 Real data analysis

From the data available at <https://covid.saude.gov.br> until May 23rd, 2020, we presented the results of the model fitting for the data accumulated of the

number of cases by COVID-19 in the Amazonas, Bahia, Ceará, Maranhão, Pará, Pernambuco, Rio de Janeiro, and São Paulo States. The purpose of this section is to determine the parameter estimates and compare the efficiency of the Chapman-Richard model in estimating the number of contagions considering different values for the parameter m . In addition, we provide long-term predictions, such as the estimated peak dates for each State, as well as predict when the pandemic will reach contagion stability to contribute with the conduct of intervention policies relevant to the control of the epidemic.

For the knowledge of the growth of the pandemic phenomenon in the eight Brazilian States, we show in Table 2 an overview of the number of infections of the new coronavirus until May 23rd, 2020. And through it, it can be seen that the first case of the new coronavirus in Brazil occurred in São Paulo State on February 26th, and since then, the number of cases has been expanding in an overwhelming way across the country. According to the survey, most of the cases are concentrated in São Paulo State, with 80,558 infected, followed by the Ceará State, which despite presenting the first case of the disease in the twelfth epidemiological week, is already the second State with the highest number of infections (35,122), and Rio de Janeiro State, which registers 34,533 cases. The Bahia State, although it started to spread in the tenth epidemiological week, is among the States with the lowest incidence of infection of the disease. Under this focus, the description of the data presented, corroborate with the indication made in the methodology to use the accumulated trajectory of the pandemic in the analysis of the non-linear growth model to predict the data.

Table 2: Description of the population, date of first infection and number of infected people until May 23rd

State	Population	Date of first infection	Epidemiological week	Infected people
Amazonas	4,144,597	03/15/2020	12	28,802
Bahia	14,873,064	03/06/2020	10	13,000
Ceará	9,132,078	03/17/2020	12	35,122
Maranhão	7,075,181	03/21/2020	12	18,767
Pará	8,602,865	03/19/2020	12	22,697
Pernambuco	9,557,071	03/12/2020	11	26,786
Rio de Janeiro	17,264,943	03/05/2020	10	34,533
São Paulo	45,919,049	02/26/2020	09	80,558

For the fit of the model, we considered the population of each State, presented in Table 2 to calculate the number of daily cases of infected people accumulated $P(t)$ per hundred thousand inhabitants.

Giving the analysis segment, the estimates and confidence intervals (95%) of the parameters of the Chapman-Richard model for the eight Brazilian States are presented in Table 3. In the process of fitting the models, we fixed the shape

parameter m , $m = 0.5, 1.0, 1.5$ and 2.0 ; for the other parameters, we provided initial values, through which the convergence was reached.

For the Bahia State, it is worth mentioning the results presented in Table 3, from where it was possible to notice that the amplitude of the confidence interval for parameter a (asymptote of the curve of infected) is large, and classical hypothesis tests would reject the significance this parameter. From a more detailed analysis of the data for this State, it was observed that between the May 18th and May 22nd, the cumulative number of cases presented a variation above 50%, leaving a total of 8,581 cases for 13,000 cases. This variation in a short period caused a high variance for the estimator of the parameter a . The fitted model considering the data until May 18th presents a significant estimate for this parameter ($\hat{a} = 305.297$ and $s.e. = 47.681$); however, we chose to consider the fitting of the model with the data until May 23rd for reflecting the current situation.

The selection criteria for the fitted models are presented in Table 4. The results indicate that the AIC and BIC criteria were decisive in favor of the fitted Chapman-Richard model considering the form parameter $m = 2.0$ for almost all States, except Amazonas, in which the selected model was with $m = 1.0$. Accordingly, based on these results, the analyzes were performed with the selected models. Next, we provide an interpretation with the fitted model for the data for each state.

Table 3: Estimates and standard error of parameters in the growth curve model

Parameters	$m = 0.5$		$m = 1.0$		$m = 1.5$		$m = 2.0$	
	MLE	(CI 95%)	MLE	(CI 95%)	MLE	(CI 95%)	MLE	(CI 95%)
Amazonas								
a	920.563	(870.955; 970.170)	2200.040	(1871.916; 2568.164)	1770.871	(1551.240; 1990.503)	1329.065	(1211.626; 1446.504)
c	202071.481	(162718.785; 241424.178)	958.306	(860.487; 1056.126)	140.280	(129.641; 150.919)	49.833	(46.705; 52.962)
k	0.177	(0.172; 0.182)	0.099	(0.095; 0.102)	0.072	(0.069; 0.076)	0.059	(0.056; 0.062)
Bahia								
a	291.273	(0; 493.028)	768.596	(0; 1547.400)	1678.829	(0; 3852.122)	3258.040	(0; 8348.075)
c	684319.095	(0; 1441716.422)	2397.504	(471.965; 4323.044)	316.148	(97.318; 534.978)	106.708	(38.491; 174.925)
k	0.141	(0.135; 0.147)	0.073	(0.068; 0.077)	0.050	(0.046; 0.054)	0.039	(0.035; 0.042)
Ceará								
a	578.397	(525.704; 631.091)	887.134	(762.107; 1012.161)	1240.142	(1008.909; 1471.374)	1617.381	(1250.195; 1984.567)
c	52737.539	(43506.846; 61968.232)	458.957	(420.858; 497.056)	82.802	(77.992; 87.612)	32.688	(31.120; 34.256)
k	0.156	(0.151; 0.162)	0.086	(0.083; 0.090)	0.063	(0.063; 0.062)	0.051	(0.048; 0.054)
Maranhão								
a	410.695	(375.882; 445.507)	665.988	(578.771; 753.205)	974.326	(807.164; 1141.488)	1320.508	(1046.643; 1594.373)
c	134286.463	(113643.484; 154929.442)	770.634	(716.294; 824.974)	121.339	(115.369; 127.309)	44.778	(42.938; 46.618)
k	0.178	(0.173; 0.183)	0.097	(0.094; 0.100)	0.070	(0.067; 0.072)	0.056	(0.053; 0.059)
Pará								
a	484.285	(431.864; 536.706)	898.538	(734.290; 1062.786)	1479.384	(1110.649; 1848.119)	2226.919	(1538.284; 2915.554)
c	1325870.264	(1115126.111; 1536614.416)	2739.764	(2513.086; 2966.442)	307.949	(287.592; 328.306)	96.098	(90.206; 101.989)
k	0.201	(0.196; 0.206)	0.107	(0.103; 0.110)	0.075	(0.072; 0.078)	0.060	(0.057; 0.062)
Pernambuco								
a	374.324	(335.601; 413.048)	503.049	(434.708; 571.390)	624.852	(524.882; 724.822)	735.829	(603.929; 867.728)
c	78966.081	(54536.488; 103395.675)	587.514	(495.033; 679.995)	99.927	(88.729; 111.125)	38.189	(34.711; 41.667)
k	0.155	(0.147; 0.163)	0.089	(0.084; 0.094)	0.067	(0.063; 0.072)	0.056	(0.052; 0.060)
Rio de Janeiro								
a	585.538	(193.418; 977.659)	1075.918	(303.629; 1848.206)	1799.012	(260.262; 3337.761)	2761.045	(12.862; 5509.228)
c	143007.869	(-8129.819; 294145.557)	813.081	(396.622; 1229.539)	131.399	(78.200; 184.599)	49.353	(31.515; 67.191)
k	0.123	(0.117; 0.130)	0.065	(0.060; 0.070)	0.046	(0.042; 0.050)	0.036	(0.032; 0.040)
São Paulo								
a	225.019	(209.178; 240.861)	305.699	(275.988; 335.411)	383.588	(338.342; 428.834)	455.628	(394.172; 517.084)
c	60904.184	(46582.532; 75225.836)	500.587	(442.616; 558.559)	88.053	(80.914; 95.191)	34.247	(32.018; 36.477)
k	0.128	(0.123; 0.133)	0.073	(0.070; 0.076)	0.055	(0.052; 0.057)	0.051	(0.049; 0.052)

Table 4: Summary of model selection criteria

States	$m = 0.5$		$m = 1.0$		$m = 1.5$		$m = 2.0$	
	AIC	SBC	AIC	SBC	AIC	SBC	AIC	SBC
Amazonas	475.145	484.139	463.303	472.297	465.199	474.193	467.961	476.955
Bahia	277.410	286.888	275.799	285.277	274.670	284.150	273.928	283.405
Ceará	415.814	424.692	400.928	409.8062	396.227	405.105	394.316	403.194
Maranhão	305.997	314.63	288.672	297.307	281.721	290.356	278.124	286.759
Pará	292.343	301.101	283.194	291.953	279.307	288.066	277.230	285.989
Pernambuco	458.839	468.001	433.184	442.346	419.512	428.674	411.251	420.413
Rio de Janeiro	454.263	463.791	449.732	459.260	446.877	456.405	445.079	454.607
São Paulo	438.010	447.919	406.178	416.087	390.974	400.883	382.358	392.267

3.1 Results for Amazonas State

In Figure 2, we presented the number of new daily cases and the number of accumulated cases. In graph 2a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 2b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 6.081$) and the coefficient of determination $r^2 = 0.998$.

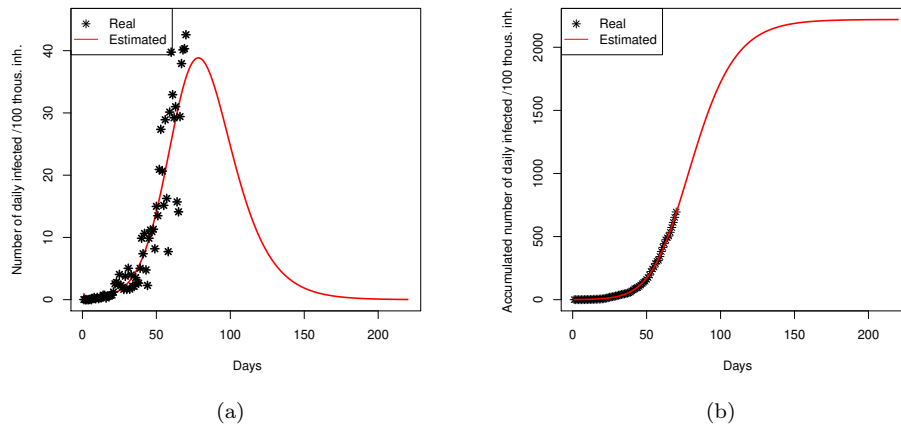


Figure 2: Growth curves fitted through Chapman-Richard model to Amazonas State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

We note in this Figure that Amazonas State is close to the peak of pandemic, and that the cumulative number of cases has already exceeded the first inflection point of the curve for new daily cases, indicating a decrease in the growth in the

number of new cases. This behavior must continue until the peak is reached. The fitted model for this State resulted in an estimate of the growth rate of $\hat{k} = 0.099$, which is high, indicating a fast growth. Besides, the estimates asymptote was $\hat{a} = 2,200$, giving an idea of the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 91,181 total cases).

In addition, Figure 2 also shows the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data).

3.2 Results for Bahia State

Figure 3 presents the number of new daily cases and the number of accumulated cases in Bahia State. The red line in graph 3a represents the expected number of new daily cases obtained with the derivative of the fitted model. The red line in graph 3b represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 1.332$) and the coefficient of determination $r^2 = 0.996$.

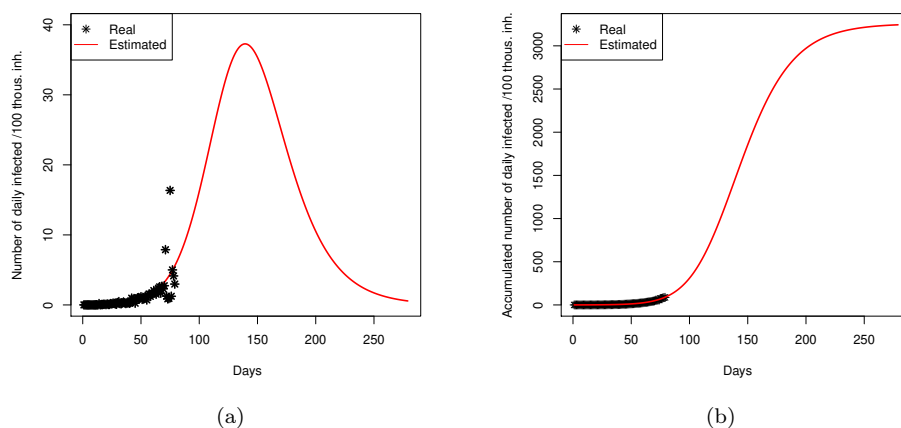


Figure 3: Growth curves fitted through Chapman-Richard model to Bahia State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

When comparing the graphs of Figure 3 with those of the other States, it was seen that the progression in the number of cases is less. The growth is currently upward in the capital; however, in the interior of Bahia, this growth is more controlled, which reduces the rate of the number of cases in the entire State. Note that, in Table 3, the value of the parameter determines the growth efficiency of the disease k is median ($\hat{k} = 0.039$). This current scenario presumably occurred due to

public policies imposed by the government to contain the dissemination, which is reflected in the curves and in the distance from the peak of contamination.

Besides, Figure 3 also shows the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data). From the fitted model for the data from Bahia, we have $\hat{a} = 3,258$ which corresponds to the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 484,564 total cases).

3.3 Results for Ceará State

The number of new daily cases and the number of accumulated cases in Ceará State are shown in the graphs in Figure 4. In graph 4a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 4b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 4.238$) and the coefficient of determination $r^2 = 0.998$.

The epidemic curve in Ceará State continues to rise. Note that the cumulative number of cases exceeded the first inflection point of the daily new cases curve. Its accelerated growth is evident in graph 4b and in the value of parameter k that determines the growth rate of the disease, which is above 0.05 (Table 3).

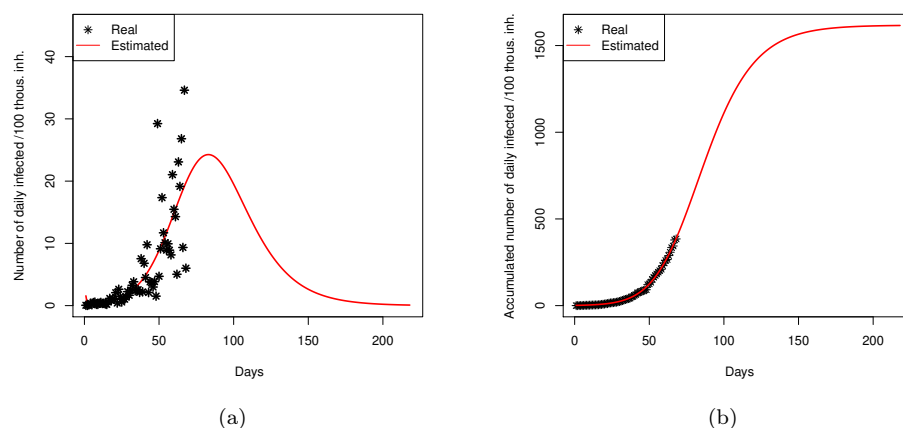


Figure 4: Growth curves fitted through Chapman-Richard model to Ceará State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

We also presented in Figure 4 the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data). The fitted model for the data of Ceará State

provides us $\hat{a} = 1,617$, which corresponds to the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 147,666 total cases).

3.4 Results for Maranhão State

In Figure 5, we presented the number of new daily cases and the number of accumulated cases in Maranhão State. In graph 5a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 5b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 2.045$) and the coefficient of determination $r^2 = 0.999$.

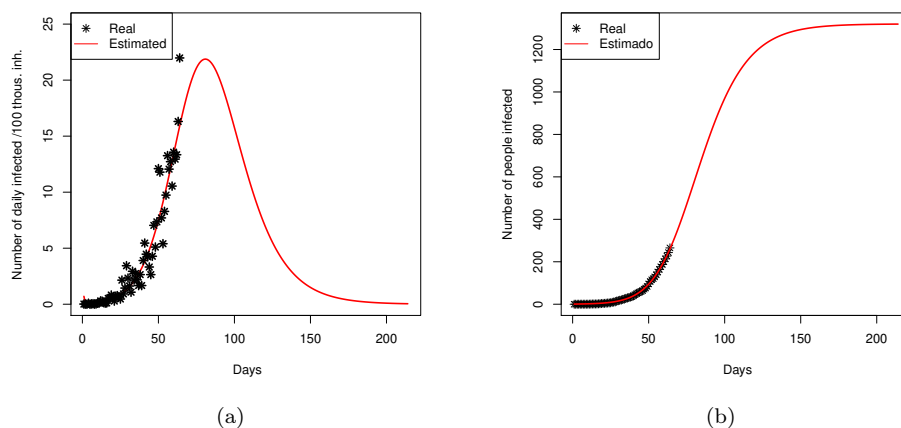


Figure 5: Growth curves fitted through Chapman-Richard model to Maranhão State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

The projection of the evolution of the epidemic in Maranhão State predicts in relation to the beginning – in which growth in this State was accelerated – a decrease in the growth rate of the numbers of accumulated cases and, consequently, of the number of new daily cases (Figure 5). In this State, the peak is predicted for the second half of May. This delay was possibly due to the express measures of restriction of movement of people across the State.

Similarly, the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data) are also provided in the graphs of Figure 5. Based on the fitted model for the data of Maranhão State, we have $\hat{a} = 1,329$, which corresponds to

the total member expected of cases per hundred thousand inhabitants until the end of the pandemic (or 55,082 total cases).

3.5 Results for Pará State

In Figure 6, we presented the number of new daily cases and the number of accumulated cases in Pará State. In graph 6a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 6b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 1.903$) and the coefficient of determination $r^2 = 0.999$.

With the analysis of the graphs of this Figure, it is possible to verify that Pará State is far from the epidemic peak. Temporary preventive measures with the lockdown implemented in ten cities in the State – whose average number of infected people was 50% higher than the average of the State (AGÊNCIA BRASIL, 2020) – probably slowed the growth of the epidemic curve of State in the recent days. However, the model demonstrates that the growth rate k is still high ($\hat{k} = 0.060$) until the present date of data collection.

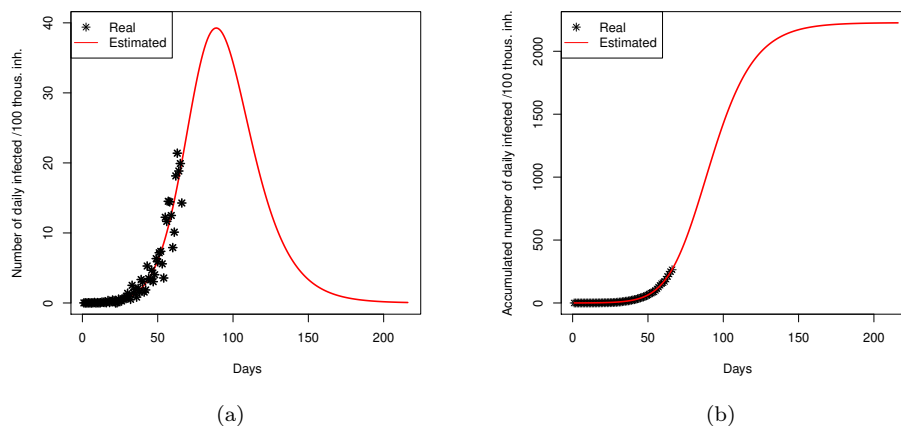


Figure 6: Growth curves fitted through Chapman-Richard model to Pará State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

Besides, Figure 6 also shows the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data). From the fitted model with the data of State Pará, we have $\hat{a} = 2,226$, which corresponds to the total number expected of cases

per hundred thousand inhabitants until the end of the pandemic (or 191,500 total cases).

3.6 Results for Pernambuco State

The number of new daily cases and the number of accumulated cases in Pernambuco State are shown in the graphs in Figure 7. In graph 7a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 7b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 3.911$) and the coefficient of determination $r^2 = 0.997$.

For this State, we noticed a high acceleration in the growth of the number of cases, and the peak is projected to occur soon and that, soon after, there will be a significant reduction in the number of new cases. This fact is confirmed by the estimate of asymptotic parameter a , which is approximately $\hat{a} = 735$, corresponding to the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 70,244 total cases).

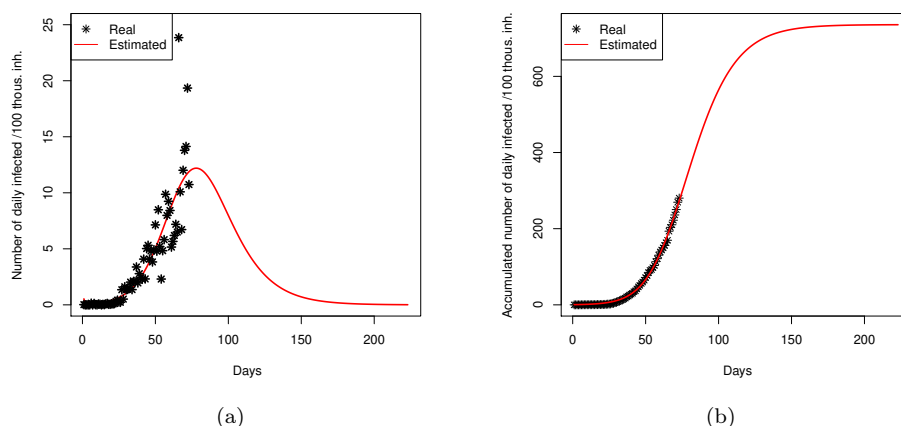


Figure 7: Growth curves fitted through Chapman-Richard model to Pernambuco State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

Similarly, the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data) are also provided in the graphs of Figure 7.

3.7 Results for Rio de Janeiro State

Figure 8 presents the number of new daily cases and the number of accumulated cases in Rio de Janeiro State. The red line in graph 8a represents the expected number of new daily cases obtained with the derivative of the fitted model. The red line in graph 8b represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 3.788$) and the coefficient of determination $r^2 = 0.995$.

From the analysis of the graphs presented in Figure 8, an expressive acceleration in the number of contagions is expected. Note that, in Table 3, the asymptotic parameter a of the growth curve was estimated at approximately 2,761, which corresponds to the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 476,685 total cases). As the contagion rate is considered to be median ($k = 0.036$), it is projected that the time to end the pandemic will be long, as shown in Figure 8.

Similarly, Figure 6 also shows the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data).

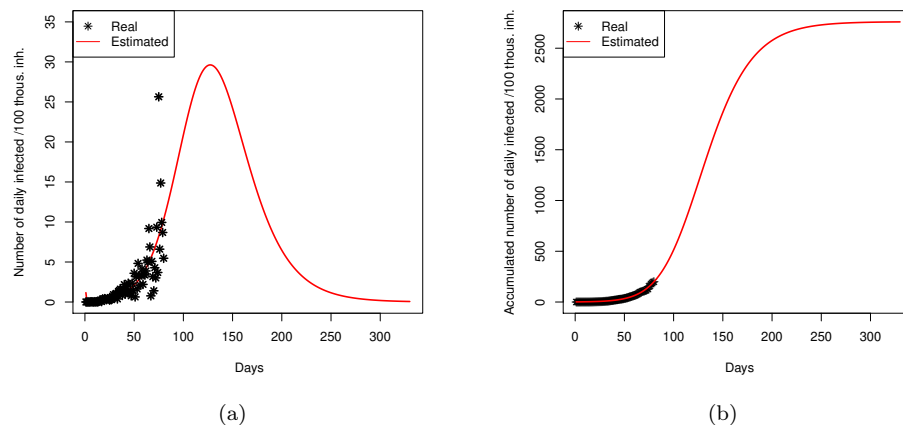


Figure 8: Growth curves fitted through Chapman-Richard model to Rio de Janeiro State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

3.8 Results for São Paulo State

São Paulo State is considered the epicenter of the pandemic in Brazil, where most of the cases are concentrated, leading to the highest infection rate. The

number of new daily cases and the number of accumulated cases in São Paulo State are shown in the graphs in Figure 9. In graph 9a, the red line represents the expected number of new daily cases obtained with the derivative of the fitted model. In graph 9b, the red line represents the expected number of accumulated cases. The quality of the fit was assessed with the square root of the mean square error ($RSME = 2.065$) and the coefficient of determination $r^2 = 0.998$.

The estimate of the growth rate k of the number of cases presented in Table 3 is higher than 0.05, which justifies the accelerated growth of the total number of cases.

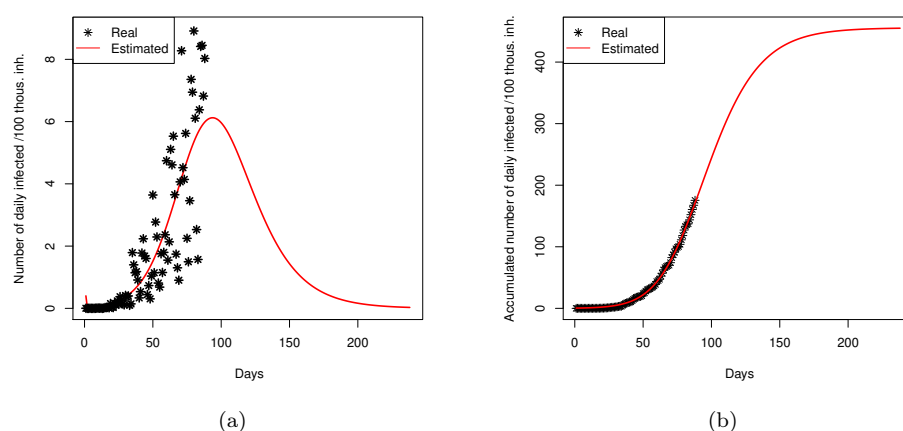


Figure 9: Growth curves fitted through Chapman-Richard model to São Paulo State, with the projection for next 150 days: (a) Number and expected value of new daily infected cases; (b) Number and expected value of the accumulated daily number of infected cases.

We also presented in Figure 9 the prediction for the next 150 days (from May 23rd) for the number of new daily cases and the cumulative number of cases (red line exceeding the observed data). The fitted model for the data of São Paulo State provides us $\hat{a} = 455$, which corresponds to the total number expected of cases per hundred thousand inhabitants until the end of the pandemic (or 208,932 total cases).

3.9 Peak and the end pandemic prediction

The peak of the pandemic is predicted by calculating the maximum point of the prediction curve for new cases daily (red line of the graphic to the left of Figures 2-9). The prediction of the end of the pandemic can be obtained by calculating the time in which the cumulative number of cases reaches the expected value of the total number of cases (given by the estimate of parameter a). Another alternative

is to calculate the time taken to account for 97% or 99% of the expected value of the total number of cases.

In Table 5, we present the prediction of the date (Date), number of days (N Days), the accumulated number of infected cases (N Infected) with the respective standard error (s.e.) for two moments of interest: the day peak and end date of the pandemic. The standard error, associated with the cumulative number of cases, was obtained considering the variance of the residual of each model (given by equation 8).

We can see in Table 5 that the São Paulo State will reach the peak around 94 days after the notification of the first case (that is, on May 29th, 2020), while the prediction of the accumulated number of cases will continue to grow more slowly if the rate of growth (parameter k) remains at the estimated value. A possible decrease in this rate can occur if social isolation control strategies are adopted. Other States worth mentioning are Amazonas and Pernambuco. It is projected that for these States the peak will occur in 78 days after the start of the infection. In addition, to the end of the pandemic, these States will have 91,336 and 68,654 cases, respectively. The peak for Ceará and Maranhão States are forecast for the month of June, while the Bahia and Rio de Janeiro States are forecast for July.

For Bahia and Rio de Janeiro States, the fitted models were strongly influenced by high values of the observations on the last days considered in the study. Therefore, it is expected that an increase in the number of observations may provide better predictions in the future.

Table 5: Forecast for date, number of days, number of infected and standard error in peak and until the end of the Pandemic

States	Peak				Predicted end pandemic			
	Date	N days	N Infected	s.e.	Date	N Days	N Infected	s.e.
Amazonas	05/31	78	41,054	64	08/22	161	91,181	43
Bahia	07/22	139	216,028	162	11/25	265	484,564	116
Ceará	06/07	83	66,624	86	09/02	170	147,665	57
Maranhão	06/18	87	51,622	63	08/02	158	93,392	41
Pará	06/15	98	87,317	125	09/05	171	191,500	88
Pernambuco	05/28	78	31,861	143	08/02	144	70,244	31
Rio de Janeiro	07/09	127	212,864	159	11/14	255	476,685	137
São Paulo	05/29	94	95,289	103	08/02	159	208,932	76

4 Conclusions and final considerations

In this paper was presented a study of the State of the pandemic that affects the Federative Republic of Brazil, the Gauss-Newton method of non-linear estimation was used, aiming to establish the models of growth dynamics considering the current situation of eight States in Brazil.

The statistical methodology used to estimate the parameters of the Chapman-Richard model of pandemic growth, the fitted model was complemented with the

construction of the confidence intervals to describe all the dynamics of the curve of the officially infected and also provide elements so that peak day and, the end of pandemic providing indicative for effective health, social and economic actions.

In this line of study, one found that the projections provided by the growth curves point out a scenario of general acceleration of contamination, with nuances of heterogeneity in terms of speed of the growth in the considered States. It must be considered that this reality is already configured, as São Paulo has the largest State population and the largest number of infected people. Bahia State it has the lowest density of infected people (1.754/100,000 inhabitants). For all the States in this study, the peak day is projected soon. In this sense, it is very interesting to highlight that the growth models obtained are fundamental for a better understanding of this phenomenon and provide effective and consistent subsidies for decision-making and interventions in public health and planning of socio-economic activities for the return to normality.

As for the structures of the analyzed models, the results show that the model with the shape parameter with $m = 2$ was the most efficient in treating the evolution of the pandemic in seven States, as they provided statistically significant estimates and confidence intervals and their parameters were adequate for the biological interpretation of the data. It is worth noting that the methodology used in the present study is sensitive to intervention strategies and control of disease proliferation.

In summary, the models predict that the epidemic is close to reaching its peak in the Amazonas, Ceará, Maranhão, Pernambuco and São Paulo States. The indications for the forecasts presented were based on a social isolation of around 45% and that, if efforts to contain the disease continue according to the different presented scenarios, the heterogeneous behavior presented by the curves in the considered States must remain and, certainly, government actions should be carried out on a regional basis.

Finally, as a proposal for future work, other growth models can be adopted and the Bayesian approach can be used to make comparisons with the present study.

Acknowledgments

We are indebted to the Editorial Boarding and Referees for their valuable comments, criticisms and suggestions which have substantially improved the text of the manuscript. Katiane S. Conceição is supported by the Brazilian organization FAPESP (2019/22412-5); Marinho G. Andrade is supported by the Brazilian organization FAPESP (2019/21766-8).

AMARAL, M. T. R., CONCEIÇÃO, K. S.; ANDRADE, M. G.; PADOVANI, C. R. Modelo de curva de crescimento generalizado para Covid-19 nos estados brasileiros. *Rev. Bras. Biom.*, Lavras, v.38, n.2, p.125-146, 2020.

- **RESUMO:** O presente artigo consiste em usar o modelo de crescimento generalizado de Chapman-Richard para relacionar funcionalmente o número de pessoas infectadas pelo COVID-19 em função do número de dias. O objetivo deste trabalho é estimar o instante em que o número de pessoas infectadas estabiliza, usando o conjunto de dados da quantidade acumulada de infectados. Para essa proposta, realizou-se um estudo comparativo das performances de diferentes parametrizações do modelo de Chapman-Richard em oito Estados brasileiros. No contexto metodológico, o procedimento de Gauss Newton foi utilizado para estimar os parâmetros. Além disso, foram utilizados critérios de seleção de modelos para selecionar aquele que melhor se ajusta ao conjunto de dados. A metodologia utilizada permitiu estimativas consistentes do número de pessoas infectadas pelo COVID-19 em função do tempo e, conseqüentemente, foi possível concluir que as projeções fornecidas pelas curvas de crescimento apontam para um cenário de aceleração geral da contaminação. Além disso, os modelos preveem que a epidemia está perto de atingir seu pico nos Estados do Amazonas, Ceará, Maranhão, Pernambuco e São Paulo.
- **PALAVRAS-CHAVE:** Corona virus; modelo de Chapman-Richard generalizado; curvas de crescimento; método de Gauss Newton.

References

- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v.6, p.716-723, 1974.
- AMARAL, M. T. R. *Abordagem bayesiana para curvas de crescimento com restrições nos parâmetros*. Dissertação de Mestrado, Universidade Federal Rural de Pernambuco, Recife – PE, 2009.
- AMARAL, M. T. R.; PADOVANI, C. R. Growth curves with multivariate θ generalized normal distribution for cardiac dysfunction in rats. *Communications in Statistics: Case Studies, Data Analysis and Applications*, v.0,n.0, p.1-13, 2020.
- BATES, D. M.; WATTS, D. G. *Nonlinear regression analysis and its applications*, volume 2. Wiley New York, 1988.
- AGÊNCIA BRASIL. Notícias de Saúde - Estado do Pará. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-05/para-estende-lockdown-ate-o-proximo-dia-24> (Accessed: May 29th, 2020).
- BROWN, J. E.; FITZHUGH Jr, H. A.; CARTWRIGHT, T. C. A comparison of nonlinear models for describing weight-age relationships in cattle. *Journal of Animal Science*, v.42,n.4, p.810-818, 1976.

- FEKEDULEGN, D.; MAC SIÚRTÁIN, M. P.; COLBERT, J. J. Parameter estimation of nonlinear models in forestry. *Silva Fennica*, v.33 n.4, p.327-336, 1999.
- LIU, Y.; GAYLE, A. A.; WILDER-SMITH, A.; ROCKLÖV, J. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 2020.
- MYERS, R. H. *Classical and modern regression with applications*, volume 2. Duxbury press Belmont, CA, 1990.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2009.
- SARMENTO, J. L. R.; REGAZZI, A. J.; SOUZA, W. H.; TORRES, R. A.; BREDAS, F. C.; MENEZES, G. R. O. Estudo da curva de crescimento de ovinos santa inês. *Revista Brasileira de Zootecnia*, v.35, n.2, p.435-442, 2006.
- SCHWARZ, G. Estimating the dimension of a model. *Annals of Statistics*, v.6, n.2, p.461-464, 1978.
- SONNINO, G. Dynamics of the covid-19—comparison between the theoretical predictions and real data. *arXiv preprint arXiv:2003.13540*, 2020.
- XU, Z.; SHI, L.; WANG, Y.; ZHANG, J.; HUANG, L.; ZHANG, C.; LIU, S.; ZHAO, P.; LIU, H.; ZHU, L.; TAI, Y.; BAI, C.; GAO, T.; SONG, J.; XIA, P.; DONG, J.; ZHAO, J.; WANG, F. Pathological findings of covid-19 associated with acute respiratory distress syndrome. *The Lancet respiratory medicine*, v.8, n.4, p.420-422, 2020.
- ZHENG, Y. Y.; MA, Y. T.; ZHANG, J. Y.; XIE, X. Covid-19 and the cardiovascular system. *Nature Reviews Cardiology*, v.17, n.5, p.259-260, 2020.

Received on 04.06.2020.

Approved after revised on 15.06.2020.