UFSM

CIÊNCIA e NATURA

ISSN 2179-460X
OPEN ACCESS

Ci.e Nat., Santa Maria, v. 45, spe. n. 3,p. 01-23, e73812, 2023 • https://doi.org/10.5902/2179460X73812
Submitted: 01/18/2023 • Approved: 03/23/2023 • Published: 12/01/2023

Special Edition

Selection of models and parameter estimation for monitoring the COVID-19 epidemic in Brazil via bayesian inference

Seleção de modelos e estimação de parâmetros para o monitoramento da epidemia de COVID-19 no Brasil via inferência bayesiana

Lucas Martins Inez[I], Carlos Eduardo Rambalducci Dalla[I], Wellington Betencurte da Silva[II], Julio Cesar Sampaio Dutra[I], José Mir Justino da Costa[III]

[I] Universidade Federal do Espírito Santo, Vitória, ES, Brasil
[II] Universidade Federal Fluminense, Niterói, RJ, Brazil
[III] Universidade Federal do Amazonas, Manaus, AM, Brazil

ABSTRACT

In 2019, a new strain of coronavirus led to an outbreak of disease cases named COVID-19, evolving rapidly into a pandemic. In Brazil, delayed decision-making and a lack of knowledge have resulted in an alarming increase in daily transmissions and deaths.In this context, researchers used mathematical models to assist in determining the parameters that act in the spread of diseases, revealing containment measures. However, numerous mathematical models exist in the literature, each with specific parameters to be specified, leading to an important question about which model best represents the pandemic behavior. In this regard, this work aims to apply the Approximate Bayesian Computation method to select the best model and simultaneously estimate the parameters to resolve the abovementioned issue. The models adopted were susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR), susceptible-infected-recovered-susceptible (SIRS), and susceptible-exposed-infected-recovered-susceptible (SEIRS). Approximate Bayesian Computation Monte Carlo Sequencing (ABC-SMC) was used to select among four competing models to represent the number of infected individuals and to estimate the model parameters based on three periods of Brazil COVID-19 data. A forecasting test was performed to test the ABC-SMC algorithm and the selected models for two months. The result was compared with the actual number of infected that were reported. Among the test models, it was found that the ABC-SMC algorithm had a promising performance since the data were noisy and the models could not predict all parameters.

Keywords: Approximate Bayesian Computation; Epidemiological models; COVID-19

Article published by Ciência e Natura under a CC BY-NC 4.0 license.

RESUMO

Em 2019, uma nova variante do coronavírus conduziu a um surto de casos de doença denominada COVID-19, evoluindo rapidamente para uma pandemia. No Brasil, o atraso na tomada de decisões e a falta de conhecimento resultaram num aumento alarmante da transmissão diária e das mortes. Nesse contexto, pesquisadores utilizaram modelos matemáticos para auxiliar na determinação dos parâmetros que atuam na propagação das doenças, revelando medidas de contenção. Entretanto, inúmeros modelos matemáticos existem na literatura, cada um com parâmetros específicos a serem determinados, levando a uma importante questão sobre qual modelo melhor representa o comportamento da pandemia. Nesse sentido, este trabalho tem como objetivo aplicar o método de Computação Bayesiana Aproximada para selecionar o melhor modelo e simultaneamente estimar os parâmetros para resolver a questão acima mencionada. Os modelos adotados foram: suscetível-infectado-recuperado (SIR), suscetível-exposto-infectado-recuperado (SEIR), suscetível-infectado-recuperado-suscetível (SIRS) e suscetível-exposto-infectado-recuperado-suscetível (SEIRS). A Computação Bayesiana Aproximada Monte Carlos Sequencial (ABC-SMC) foi utilizada para selecionar entre quatro modelos concorrentes para representar o número de indivíduos infectados e para estimar os parâmetros dos modelos com base em três períodos de dados da COVID-19 do Brasil. Foi realizado um teste de previsão para testar o algoritmo ABC-SMC e os modelos selecionados para dois meses. O resultado foi comparado com o número real de infectados que foram notificados. Entre os modelos testados, verificou-se que o algoritmo ABC-SMC tem um desempenho promissor, uma vez que os dados eram ruidosos e os modelos não conseguiam prever todos os parâmetros.

Palavras-chave: Computação Bayesiana Aproximada; Modelos epidemiológicos; COVID-19

1 INTRODUCTION

At the end of 2019, there was the discovery of COVID-19 caused by a mutated virus in a wild animal. With globalization, it spreads easily and reaches the global scale, which is the pandemic level. However, this is not uncommon since H_1N_1, EBOLA, and the Black Plague are well known to have inflicted on the world population in the past. Thus, history shows that COVID-19 is one of the epidemics facing humanity, being neither the first nor the last, but posing the question of “when” will be the next one. (Ujvari, 2020).

Due to a large number of deaths worldwide, COVID-19 reached the mark of 584 million infected individuals and 6.43 million deaths (on August 11, 2022), standing out as the worst crisis of today’s humanity. Contingency measures were the primary means of containment using masks, social distancing, and personal hygiene (Singh

and Arquam, 2022). In Brazil, the first case was confirmed on February 26, 2020. Since then, this country has reached over 34 million infected, with a value of 680.786 deaths (on August 11, 2022), an excessive growth from the first confirmed case, behind only the United States of America (Marinho et al., 2021).

Monitoring the spread of diseases is carried out with predictive analysis tools specific to each disease and period (Umar et al., 2020; Dhar & Bhattacharya, 2019; Rao & Upadhyay, 2013; Ng et al., 2003), helping in disease management. The countries, with the highest number of infections, are the main point of data collection, as in Cooper; Mondal; and Antonopoulos, (2020), during the beginning of the COVID-19 pandemic, applied the SIR model in China, South Korea, India, Australia, Italy, and USA specifically in the state of Texas, revealing information that only data could not provide future disease points. Even dealing with the simplest model, SIR, Postnikov (2020), Muñoz-Fernández; Seoane; and Seoane-Sepúlveda, (2021), and Marinov and Marinova (2022) revealed strong preventive measures, qualitatively and quantitatively, of their parameters.

As the models become more complex, new parameters bring perspectives to the system analyses, which are the key point in the trajectory of the disease, the path defined by the parameters. Paul and Kuddus (2022) used the SEIR model to analyze populous countries, Brazil and India, verifying the lack of control of the disease in the studied period and suggesting preventive measures such as curfews, containment zones, and reinforcement of social isolation. Wintachai and Prathom (2021) analyzed the efficiency of the integrated vaccination into the SEIR model, ensuring that an effective vaccine with a high vaccination rate would significantly reduce the number of infected. In Singapore, a study analyzed the COVID-19 disease, with the SEIR model based on area and displacement of individuals, showing promising results in containing policy measures. (Liu; Ong, & Pang, 2022).

Mathematical models were used to understand the spread of diseases, adapting parameters of each region to describe the observed cases and simulating

contexts to predict possible impacts caused by COVID-19 (Morens and Fauci, 2020). For instance, López and Rodó (2021) used the modified SEIR model to study Italy and Spain, obtaining accurate projection values and contributing to the understanding of other countries. Paul and Kuddus (2022) evaluated the application of two doses of the vaccine in Bangladesh, using the SEIR model, finding that a vaccination rate of 86.1% would control the transmission of the virus and reduce the number of severe cases. Apergis (2022) evaluated the volatility of 5 cryptocurrencies during the pandemic. They observed a significant role in investments due to the instability in market evaluation, resulting in low resources in the fight against the disease.

The spread of the disease was influenced by social, political, religious, climatic, and other aspects, requiring different strategies to be developed. However, the significant problem was that many health systems underestimated the effect of the coronavirus (Ujvari, 2020). Soon, other aspects can be incorporated into the models, according to the disease cycle or socioeconomic aspect of the country, such as hospitalization dynamic (Bekker et al., 2023; Kozyreff, 2021), under-reporting cases (Chen et al., 2022; Deo & Grover, 2021), containment strategies (Das et al., 2021), age of individuals (Franco, 2021), macroeconomic impacts (Chan, 2022), outbreak recurrence (Muñoz-Fernández; Seoane; & Seoane-Sepúlveda, 2021), and other aspects. It is noted that the models are limited, as they work with possible assumptions about the spread of diseases.

These models can be approached by means of machine learning methods and classical resolution of models based on differential equations. However, to address different aspects and understand the mechanisms of the problem, it is necessary to adapt the parameters to the reality of a specific region. A good model provides predictive data close to the real ones, mainly observing which parameters have the potential to reverse or attenuate the proliferation of viruses, simulating future conditions. Specific parameters are added in a attempt to approximate the real model, observing whether transmission occurs by a vector or personal contact, whether a bacterium or virus

causes the disease, whether it is controlled by medication or vaccines, climatic factors, social aspects, and social isolation, among others.

In this regard, such models must be adapted to the particular reality of the population under study. Thus, implementing the approximate Bayesian Computation (ABC) algorithm contributes to the selection of an ideal model. ABC is a mathematical method implemented in algorithms capable of selecting parameters and concurrent models and inferring values when confronted with real data. In addition, ABC allows calculation without the need for a likelihood approach, which could be cumbersome mainly in biological systems, providing prediction of future points following less expensive distance functions. (Martin, Kumar, & Lao, 2021)

This work considers the application of the ABC algorithm to select the best model to forecast the number of COVID-19 infections. Besides that, the parameters of the selected model are calibrated, and factors influencing infection in Brazil are identified.

2 MATHEMATICAL FORMULATION

The infectious models chosen were susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR), susceptible-infected-recovered-susceptible (SIRS) and susceptible-exposed-infected-recovered-susceptible (SEIRS) models. Each model followed the considerations of previously established articles: the SIR model is treated as established by Qiu et al. (2022); the SIER as proposed by Weinstein et al. (2020); SIRS followed what was determined in Morando et al. (2022) and SIERS what was proposed by Lima (2021). The simplest model used to describe infectious processes is the SIR model, which includes three groups to represent dissemination. Susceptible individuals can be infected, and the infected become recovered or resistant. The total population number is assumed to be N(t), as observed in (1). The variation of groups with time is represented by (2), (3) and (4).

$$N_{(t)} = S_{(t)} + I_{(t)} + R_{(t)} \tag{1}$$

$$\frac{dS_{(t)}}{dt} = -\alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} \tag{2}$$

$$\frac{dI_{(t)}}{dt} = \alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} - gI_{(t)} \tag{3}$$

$$\frac{dR_{(t)}}{dt} = gI_{(t)} \tag{4}$$

In such equations, *N, S, I,* and *R* are the numbers of individuals in the total population, susceptible, infectious, and recovered successively. The parameter α represents the infection rate and g is the recovery rate. The reproducibility rate ($R_0 = \alpha/g$) is essential to determine the pandemic emergence state. It can be inferred that $R_0 > 1$ shows that the disease is out of control and the number of cases increases. On the other hand, when $R_0 < 1$, the pandemic is controlled.

The SEIR model is an extension of the SIR model, adding the condition of incubation time. It reflects a greater reality in predicting situations lacking hospital beds and vaccine efficacy. This model describes the parameters of the total population N(t), divided into four groups: Susceptible (S), Exposed (E), Infected (I), and Recovered (R), which has its mathematical foundation described by (5),(6),(7),(8),(9), where β represents the incubation rate.

$$N_{(t)} = S_{(t)} + E_{(t)} + I_{(t)} + R_{(t)} \tag{5}$$

$$\frac{dS_{(t)}}{dt} = -\alpha \frac{S_{(t)} I_{(t)} ana}{N_{(t)}} \tag{6}$$

$$\frac{dE_{(t)}}{dt} = \alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} - \beta E_{(t)} \tag{7}$$

$$\frac{dI_{(t)}}{dt} = \beta E_{(t)} - gI_{(t)} \tag{8}$$

$$\frac{dR_{(t)}}{dt} = gI_{(t)} \tag{9}$$

The SIRS model, also derived from the SIR, adds the condition of reinfection so that some infected individuals become susceptible again, being mathematically represented by (1), (10), (11), and (12). The parameter μ represents the reinfection rate, the only parameter distinct from the SIR model.

$$\frac{dS_{(t)}}{dt} = -\alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} + \mu R \tag{10}$$

$$\frac{dI_{(t)}}{dt} = \alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} - g I_{(t)} \tag{11}$$

$$\frac{dR_{(t)}}{dt} = g I_{(t)} - \mu R \tag{12}$$

Lima (2021) proposed the study of SEIRS, considering all the factors mentioned in the previous models, leading representation given by (5), (13), (14), (15), (16).

$$\frac{dS_{(t)}}{dt} = -\alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} + \mu R \tag{13}$$

$$\frac{dE_{(t)}}{dt} = \alpha \frac{S_{(t)} I_{(t)}}{N_{(t)}} - \beta E_{(t)} \tag{14}$$

$$\frac{dI_{(t)}}{dt} = \beta E_{(t)} - g I_{(t)} \tag{15}$$

$$\frac{dR_{(t)}}{dt} = g I_{(t)} - \mu R \tag{16}$$

This model comprises the parameters of μ, g, α, and β from the previous models. In the proposed models, the conditions define that the sum of individuals, resulting from (1) and (5), obtains the value of N(t) and that this value is constant. The variables $S(t)$, $E(t)$, $I(t)$, and $R(t)$ were normalized by the value of N(t) to obtain fair values for simulation. Furthermore, the Ministry of Health of Brazil assumed that exposed

individuals take about 2 to 5 days to develop the symptoms of the disease, with the value of 2 days being used in the simulations.

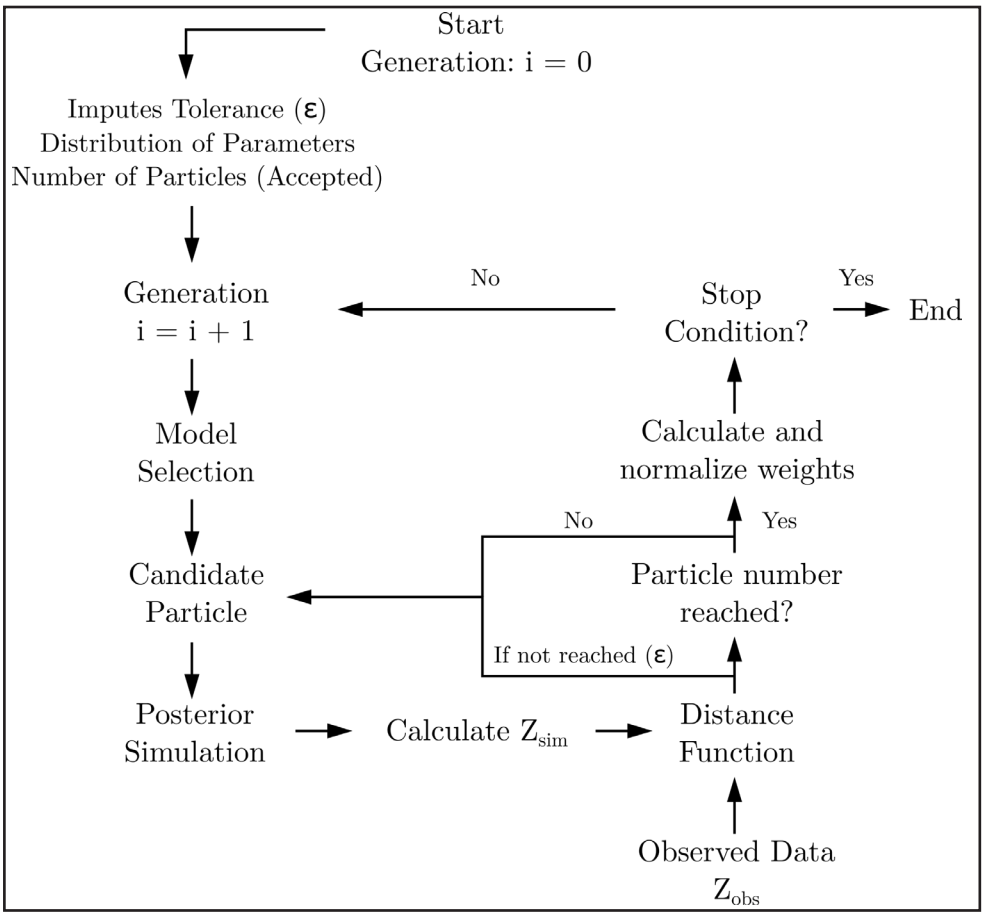
3 APPROXIMATE BAYESIAN COMPUTATION

The ABC algorithm simultaneously performs the estimation of all parameters and the selection of models. The model with the largest number of particles close to the experimental value guarantees a reasonable estimate of the parameters in the posterior distribution. Thus, the greater the number of parameters, the more costly the parameter estimation and particle adaptation will be (Toni et al., 2008; Beaumont, 2010; Sunnåker et al., 2013; Minter & Retkute, 2019).

In ABC, the simulator works as a black box feeding candidate parameters, called particles, resulting in simulated data; these values are contrasted with the observed data. They are accepted when they reach a tolerance (ϵ), previously set in the distance function (Martin, Kumar, & Lao 2021)

To reduce particle rejection, ABC is used in conjunction with a sequential Monte Carlo (SMC) method known as particle filter, which starts the simulations with a high value ϵ. If a certain number of accepted candidate particles is reached, the value of ϵ decreases, and the previous particles are used as a basis for the new generation. This process is repeated from generation to generation until the desired value ϵ is reached or this value remains constant, resulting in the maximum approximation of the simulated data set (Z_{sim}) to the observed ones (Z_{obs}) as a function of distance, as seen in Figure 1. In generation 1, the simulator determines the maximum and minimum points established for the parameters, by decreasing tolerances over time, and adjusts the weights in a uniform distribution of particles, with a fraction for the next generation (Liepe et al., 2014).

Figure 1 – ABS-SMC sequence diagram



Source: Authors

4 NUMERICAL RESULTS AND DISCUSSION

All codes used in the simulations were implemented in Spyder, which is a free and open-source scientific environment written in Python, on an i5-3337U notebook, 1.8GHz and 4Gb of RAM. As a computational implementation of the ABC algorithm, the library PYABC library was considered, which is available from the repository https://github.com/icb-dcm/pyabc.

The measurements collected were daily values, covering the period from February 25/2020 to April 19/2022, totaling 785 days. In the simulations, the number of infected people was used as Z_{obs}, the value evaluated in each model, verifying if the algorithm can return the Z_{sim} values that best represent the system.

Initially, a fragmentation of the observed data was carried out to separate milestones that directly interfered with disease behavior. Dates and data obtained from the Ministry of Health of Brazil allow seeing the first phase (I) as a period of 272 days, the second phase (II) as a period of 386 days, and the third phase (III) in Brazil after the discovery of the omicron variant represented here by 127 days in Figure 2.

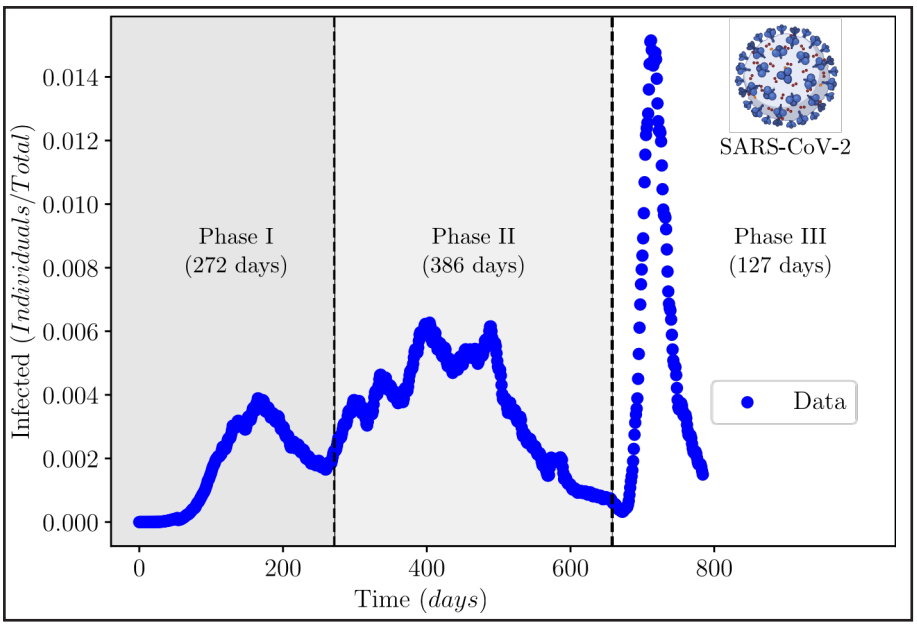
The first data set followed the stopping criterion of 10^{-4} for the minimum tolerance difference between generations; the desired minimum tolerance value of 10^{-8}, and 150,000 particles were used. By simulating data from the first phase of COVID-19 in Brazil, the number of generations and tolerance of each model can be observed in Figure 3.

In Figure 3, the SIR, SEIR, SIRS, and SEIRS models presented tolerances of 0.292; 0.37; 0.193, and 0.424. Therefore, the SIRS model had the lowest tolerance, in more significant number of generations. This result shows that it is the best model to describe the number of cases of COVID-19 during the first phase, obtaining parameters with a value of $\alpha=0.999\pm4.22\times10^{-5}$, $g=0.911\pm9.42\times10^{-5}$, and $\mu=0.026\pm7.58\times10^{-5}$, where the distribution of the 1000 simulated particles is observed from the last distribution in Table 1.

Table 1 shows the ranges of the values of each parameter simulated in the SIRS model, demonstrating a low variation between the range, referring to a good agreement between the measured values and the median of the parameters by the ABC-SMC algorithm. By contrasting the simulated values with those observed, it is possible to verify the adequacy of the adjusted model.

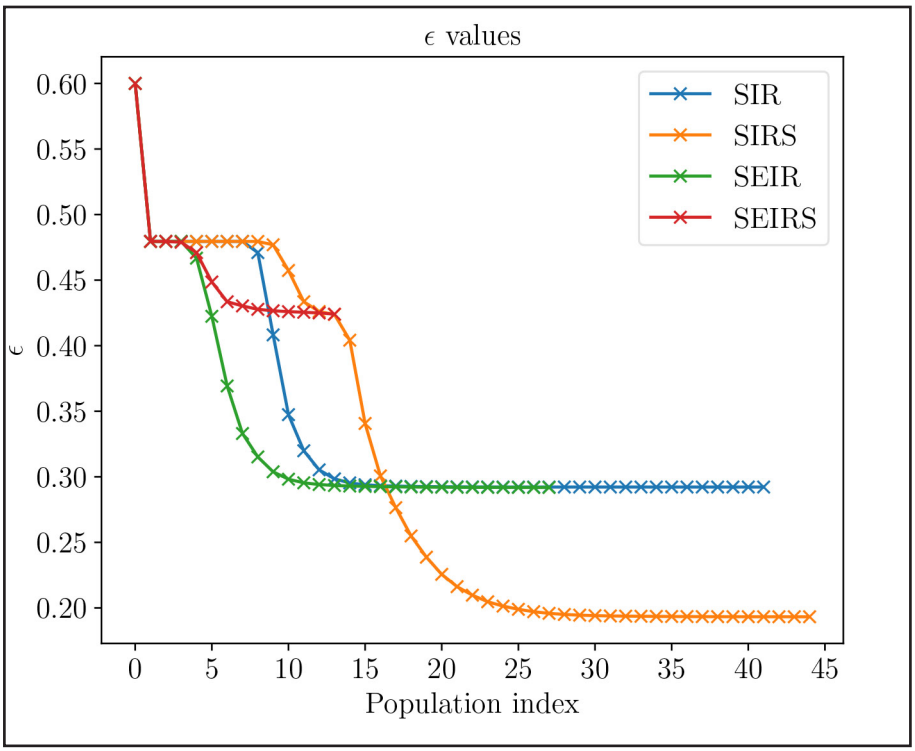
It can be seen in Figure 4 that in the first generation, with larger tolerances, the SIRS model did not fit and the particles generated parameter values with bad adjustment. However, as the generations succeed, the simulated curve tends to the real data behavior, converging at a smaller distance and providing adjusted parameter values.

Figure 2 – The number of infected individuals in Brazil divided into phases



Source:Authors

Figure 3 – Tolerance value and generations of the first data group



Source: Authors

Table 1 – Estimation of data parameters I generated from the SIRS model

Parameters	Quantile 0.01	Median	Quantile 0.99
α	0.9998	$0.999 \pm 4.22 \times 10^{-5}$	0.9999
g	0.9106	$0.911 \pm 9.42 \times 10^{-5}$	0.9111
μ	0.0257	$0.026 \pm 7.58 \times 10^{-5}$	0.0261

Source: Authors

In the second data group (phase II), the tolerance difference when simulating the models was 10^{-4}, the minimum desired tolerance value of 10^{-4}, and the maximum accepted particles of 200,000, ensuring the stopping criteria. Thus, the values of generations and tolerances described in Figure 5 were found.

After the simulation, with values represented in Figure 5, the tolerance for each model was 0.1453; 0.145; 0.1456, and 0.1466 in the SIRS, SEIR, SIRS, and SEIRS models, respectively, indicating similar values. Taking into account the criterion of constant tolerance value, the SEIR model presented slightly the lowest final tolerance value, thus representing the best fit; although the SIRS model has the highest number of generations. Therefore, the parameters to describe the problem behavior were $\alpha=0.156\pm3.9\times10^{-4}$, $g=0.994\pm0.012$, and $\beta=0.138\pm3.25\times10^{-4}$, where the distribution of the 1000 simulated particles, from the last distribuition, is observed in Table 2.

Table 2 – Estimation of the parameters of data II generated from the SEIR model

Parameters	Quantile 0.01	Median	Quantile 0.99
α	0.155	$0.156 \pm 3.9 \times 10^{-4}$	0.157
g	0.949	0.994 ± 0.012	0.999
β	0.137	$0.138 \pm 3.25 \times 10^{-4}$	0.139

Source: Authors

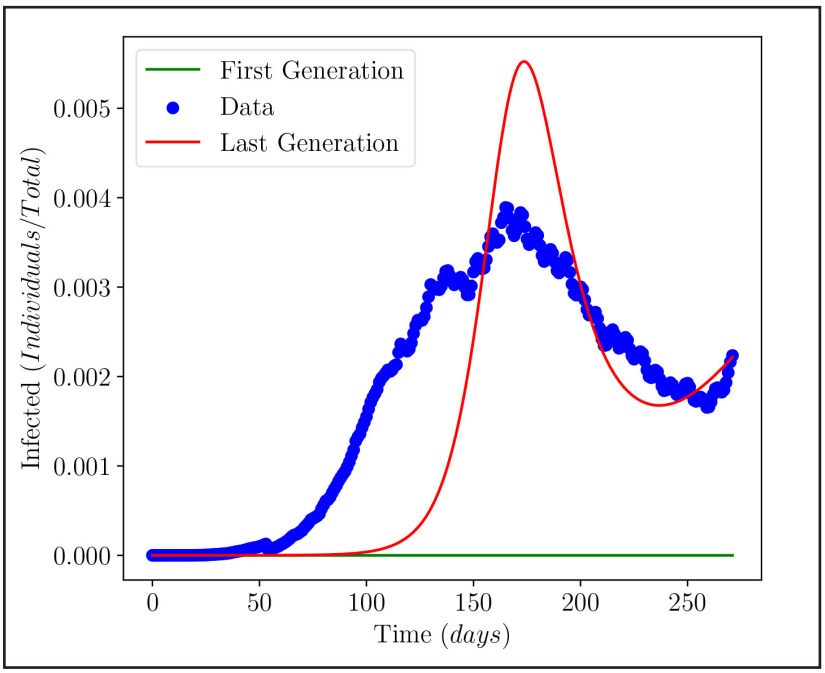
Table 2 shows the ranges of the values of each parameter simulated in the SEIR model, demonstrating a low variation in the range, referring to a good agreement between the measured values and the median of the parameters by the ABC-SMC algorithm.

In Figure 6, comparing the first and last generation, it is noticed that, when the tolerance is high, there is no adjustment of the curve. However, as the generations increase, the simulator generates better candidate parameters, and the curve represents a regression that fits the data.

In the simulations considering the period of Omicron variant discovery (that is, phase III), the number of generations of each model and the tolerance reached with this data group can be verified in Figure 7. As a stopping condition, the criterion had values of 10^{-4} for the difference in the minimum tolerance between generations. The minimum desired tolerance value of 10^{-4} and the maximum of 200,000 particles used.

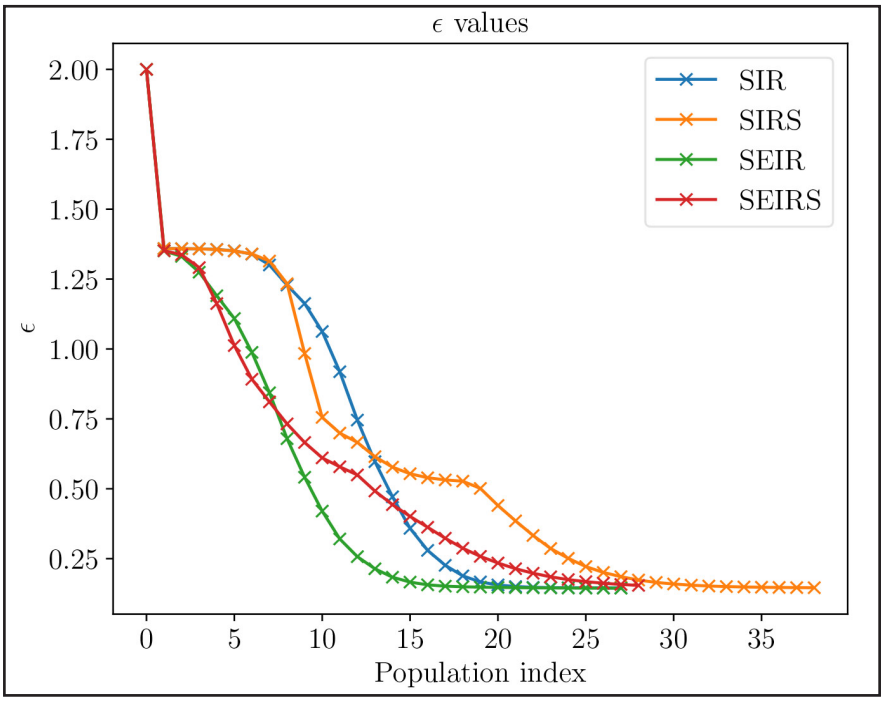
In Figure 7, the tolerance values obtained were 0.164, 0.122, 0.14, and 0.083 for the SIR, SEIR, SIRS and SEIRS models, respectively. Observing the results of the tolerance values according to the generations, the lowest value reached by the SEIRS, SEIR SIRS, and SIR model can be seen consecutively. Therefore, the closest fitted model to the observed data was SEIRS, with parameters α=0.704±6.36x10^{-3}, g=0.997±8.73x10^{-4}, β=0.730±1.18x10^{-3}, and μ=0.0127±1.18x10^{-4}, where the distribution of the 1000 simulated particles is shown in Table 3.

Figure 4 – Adjustment of SIRS in the first and last generation in the second data group



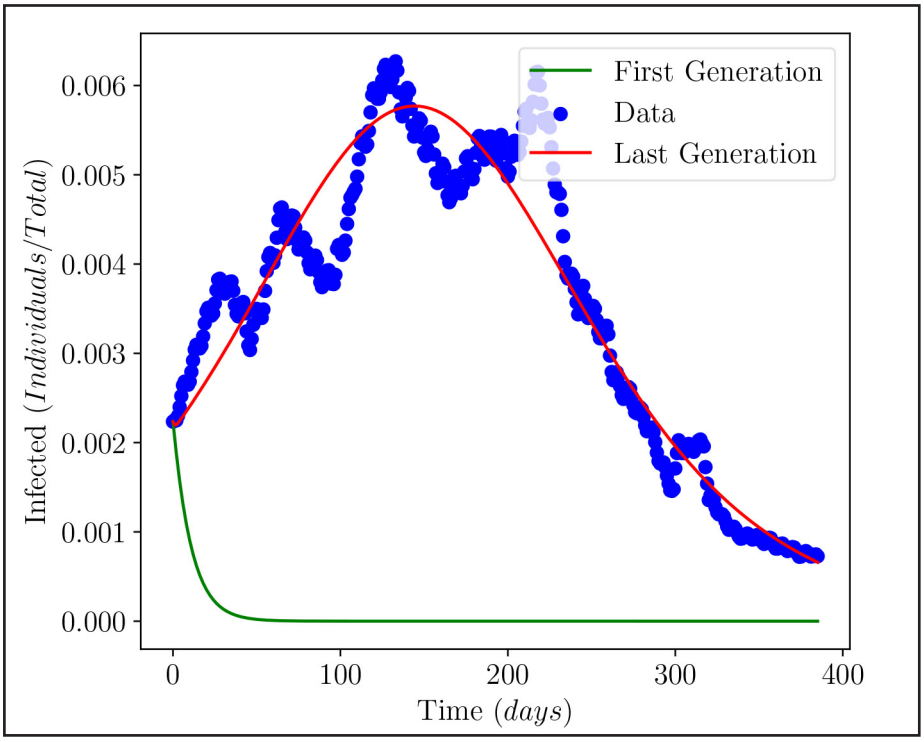
Source: Authors

Figure 5 – Tolerance value and generations of the second data group



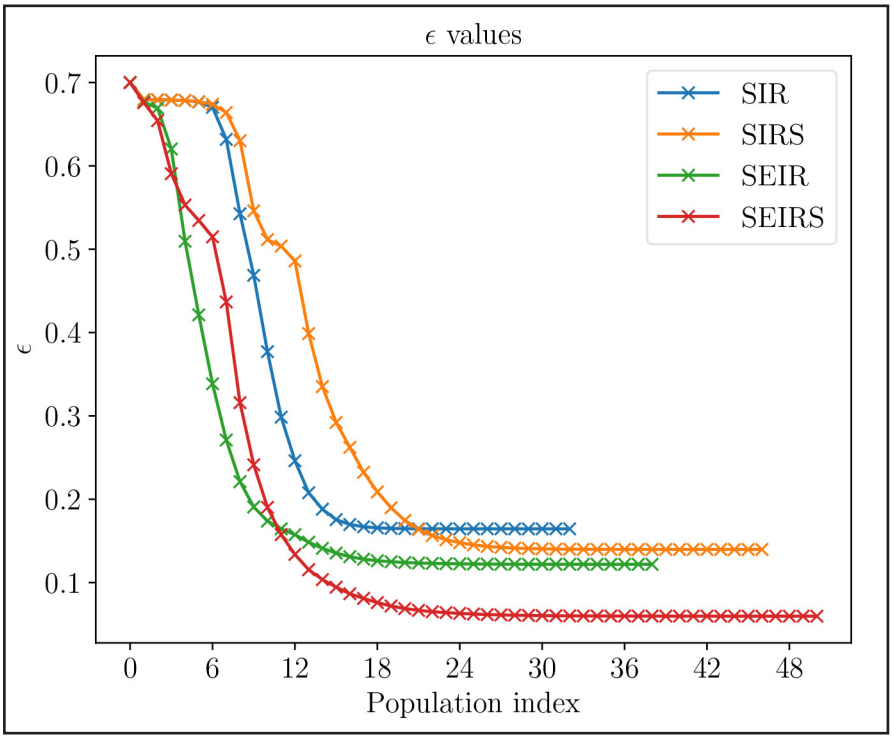
Source: Authors

Figure 6 – Adjustment of SEIR in the first and last generation in the second data group



Source: Authors

Figure 7 – Tolerance value and generations of the third data group



Source: Authors

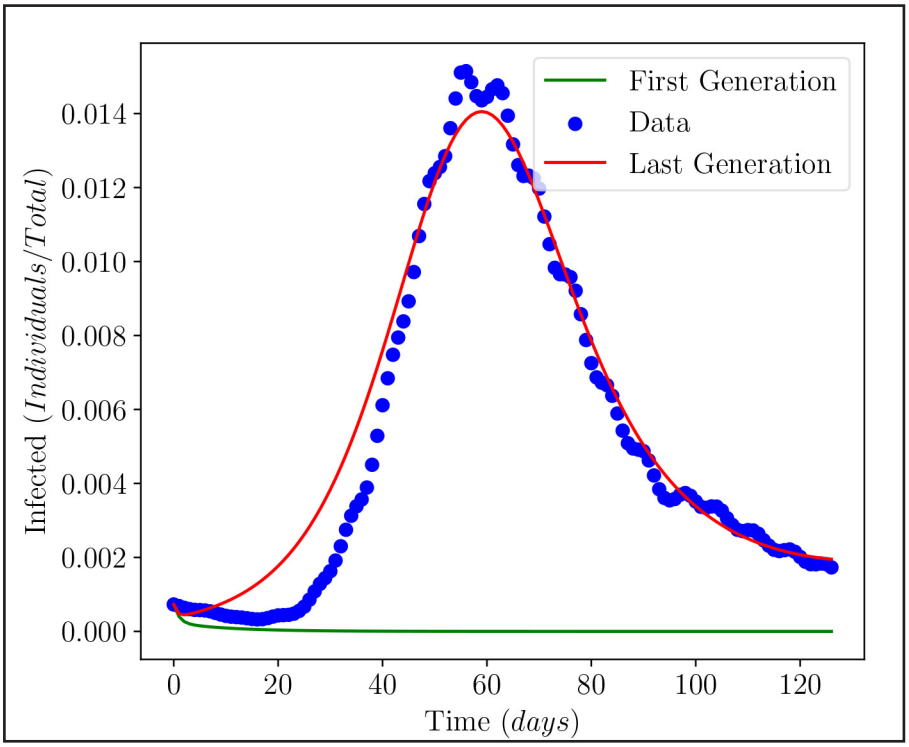
Table 3 – Estimation of data parameters III generated from the SEIRS model

Parameters	Quantile 0.01	Median	Quantile 0.99
α	0.678	$0.704 \pm 6.36 \times 10^{-3}$	0.71
g	0.995	$0.997 \pm 8.73 \times 10^{-4}$	0.999
β	0.727	$0.730 \pm 1.18 \times 10^{-3}$	0.733
μ	0.012	$0.0127 \pm 1.18 \times 10^{-4}$	0.013

Source: Authors

Table 3 shows the ranges of values for each parameter of the SEIRS model. There can be seen a wide variation in the parameters α and μ compared to the variation of g and β. This refers to the greater dispersion of particles in α and μ, measured by the ABC-SMC algorithm. By contrasting the simulated values with those observed, it is possible to verify the adequacy of the fitted model.

Figure 8 – Adjustment of SEIRS in the first and last generation of the discovery period of the Omicron variant



Source: Authors

In Figure 8, for the first generation, which has larger tolerances, the parameter values were inadequate. However, as generations succeed, the simulated behavior was improved and approximated the observed data.

In order to compare the values obtained in each group (phase I, II and III), a table was created for the selected models, comprising their parameters and tolerance values, as seen in Table 4.

Table 4 – Values of the estimated parameters in each phase

Group	Model	α	g	β	μ	ϵ	R_0
1st phase	SIRS	$0.999 \pm 4.22 \times 10^{-5}$	$0.911 \pm 9.42 \times 10^{-5}$	-	$0.026 \pm 7.58 \times 10^{-5}$	0.193	1.097
2nd phase	SEIR	$0.156 \pm 3.9 \times 10^{-4}$	0.994 ± 0.012	$0.138 \pm 3.25 \times 10^{-4}$	-	0.145	0.157
Discovery of the omicron	SEIRS	$0.704 \pm 6.36 \times 10^{-3}$	$0.997 \pm 8.73 \times 10^{-4}$	$0.730 \pm 1.18 \times 10^{-3}$	$0.0127 \pm 1.18 \times 10^{-4}$	0.083	0.706

Source: Authors

According to the Table 4, the best fit of the SIRS model in the first stage refers to the interference of the parameters of α, g, and μ, showing that the reinfection rate interfered directly at the beginning of the pandemic. In the second period, when the SEIR model was selected, the parameters α, g, and β were present, demonstrating the interference of the exposure rate. After the omicron variant was discovered, all parameters interfered with the data.

Among the selected models, it is still possible to verify the values of ϵ in each period. The low value of ϵ in the third group makes it possible to certify the best data adequacy. This could be related to the smaller oscillation in the data, possibly due to the greater availability of patient testing kits; and also the nonoccurrence of waves of infection, which may be linked to the fact that the omicron variant was more dominant.

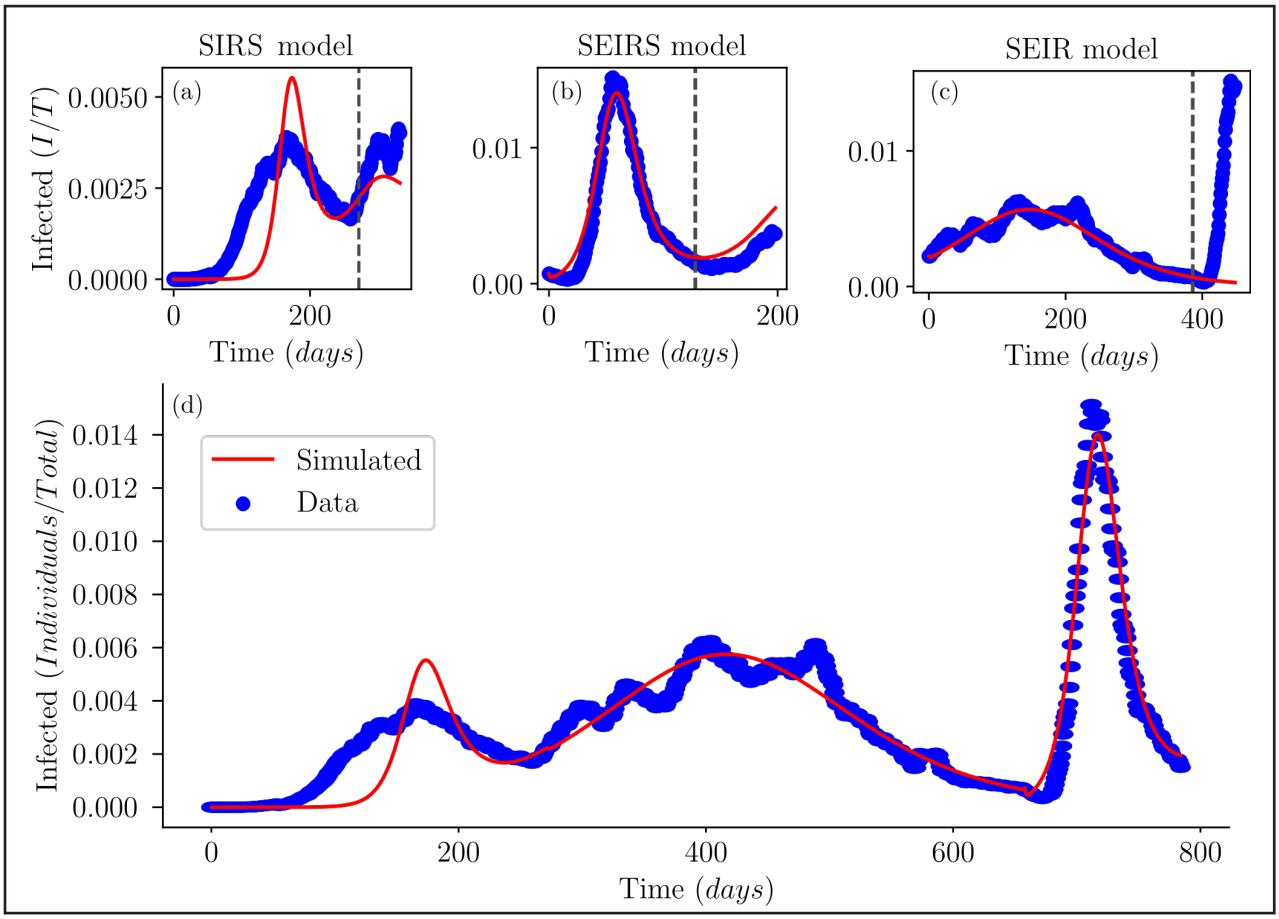
For the rate R_0, the values of 1.097, 0.157, and 0.706 can be found for each phase according to the selected models. The high value of R_0 predicts an increase of infection cases, and the low value demonstrates the tendency to control of the disease spreading. Therefore, it is noted that the infection was more uncontrolled in the first phase. In the other two periods, the number of infected individuals was under a better control policy (Qiu, *et al*. 2022).

So far, SIRS has been verified as the best model for the first phase, SEIR for the second phase, and SEIRS after discovering the omicron variant. A forecasting test was performed to test the result of the ABC-SMC algorithm. Models were simulated for the 2 month period with the values of the estimated parameters. The result was compared with the actual number of infected that were reported.

For the data of phases I and III seen respectively in the Figure 9 (a,b), it can be observed that the forescasting results using SIRS and SEIRS showed future estimates close to the real data, considering two months after the period used in the model calibration. However, the result for the SEIR model in Figure 9 (c), which was adjusted for the data of phase II, failed to predict the increase in new infected individuals. This result was expected since the omicron variant was about to emerge in phase III.

Finally, an overall evaluation in Figure 9 (d) shows that the set of models selected and calibrated using the ABC-SMC provides a promising response to describe the behavior of COVID-19 in Brazil.

Figure 9 – Forecasting test for the period of 2 months: (a) SIRS, (b) SEIR, (c) SEIRS (d) response of the models in all time periods



Source: Authors

5 CONCLUSIONS

In this work, the estimation of parameters for the SIR, SIRS, SEIR, and SEIRS models, applied to epidemiological data from Brazil, revealed promising results. However, the estimation of precise parameters, which can predict the behavior of the disease, becomes unlikely due to the terrible unpredictability of the data. The selected

models consider constant parameters throughout the infection time, not representing samples with large deviations and different scenarios in a short time.

The application of ABC-SMC in epidemiological systems revealed a powerful potential to fit data. Demonstrating that with suitable models following the same methodology, even with highly dispersed data, it was possible to generate promising results by incorporating various data sources. It was still possible to evaluate different models, which showed that the best model is not always the one with more parameters. However, the characteristics of the data, such as the size of the point dispersion, must be studied to better understand the behavior and its predictability potential when these points are changed.

Anyway, even in the presence of errors to approximate the real data, a pattern for the future expectation can be revealed by simulating the fitted models, provided that there are no significant changes in the virus strains.

ACKNOWLEDGEMENTS

This paper was partially funded by the following Brazilian agencies: FAPES (Fundação de Amparo à Pesquisa do Espirito Santo) and FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais).

REFERENCES

Apergis, N. (2022). COVID-19 and cryptocurrency volatility: Evidence from asymmetric modelling. *Financ Res Lett*, V.*47*, P. 102- 659. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S1544612321005894. Acesso em: Julho de 2023.

Beaumont, M. A. (2010). Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, *41*(1), 379–406. Disponível em: https://doi.org/10.1146/annurev-ecolsys-102209-144621. Acesso em: Julho de 2023.

Bekker, R.; uit het Broek, M.; Koole, G. (2023). Modeling COVID-19 hospital admissions and occupancy in the Netherlands. *Eur J Oper Res*, *304*(1), 207–218, Disponível em: https://www.sciencedirect.com/science/article/pii/S0377221721011000.Acesso em: Julho de 2023.

Chan, Y. T. (2022). The macroeconomic impacts of the COVID-19 pandemic: A SIR-DSGE model approach. *China Econ Rev*, V.*71*, 101-725. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S1043951X21001437. Acesso em: Julho de 2023.

Chen, Z.; Feng, L.; Lay, H. A.; Furati, K.; Khaliq, A. (2022). SEIR model with unreported infected population and dynamic parameters for the spread of COVID-19. Math Comput Simul, Vol.198, P. 31–46, Disponível em: https://www.sciencedirect.com/science/article/pii/S0378475422000787. Acesso em: Julho de 2023.

Cooper, I.; Mondal, A.; Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, V.*139*, P. 110-057. Disponível em: https://linkinghub.elsevier.com/retrieve/ pii/S0960077920304549. Acesso em: Julho de 2023.

Das, A.; Dhar, A.; Goyal, S.; Kundu, A.; Pandey, S. (2021). COVID-19: Analytic results for a modified SEIR model and comparison of different intervention strategies. *Chaos, Solitons & Fractals*, Vol.*144*, p. 110-595. Disponível em: https://www.sciencedirect.com/science/article/pii/S0960077920309863. Acesso em: Julho de 2023.

Deo, V.; Grover, G. (2021). A new extension of state-space SIR model to account for Underreporting – An application to the

COVID-19 transmission in California and Florida. *Results Phys*, Vol.*24*, p. 104-182. Disponível em https://www.sciencedirect.com/science/article/pii/S2211379721003302. Acesso em: Julho de 2023.

Dhar, M.; Bhattacharya, P. (2019). Analysis of SIR Epidemic Model with Different Basic Reproduction Numbers and Validation with HIV and TSWV Data. *Iran J Sci Technol Trans A Sci*, *43*(5), 2385–2397. Disponível em: http://link.springer.com/10.1007/s40995-019-00701-9. Acesso em: Julho de 2023.

Franco, N. (2021). COVID-19 Belgium: Extended SEIR-QD model with nursing homes and long-term scenarios-based forecasts. *Epidemics*, Vol.*37*, p. 100-490. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S1755436521000414.

Kozyreff, G. (2021). Hospitalization dynamics during the first COVID-19 pandemic wave: SIR modelling compared to Belgium,

France, Italy, Switzerland and New York City data. *Infect Dis Model*, Vol.*6*, P. 398–404. Disponível em: https://www.sciencedirect.com/science/article/pii/S2468042721000099. Acesso em: Julho de 2023.

Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P., Stumpf, M. P. H. (2014). A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nat Protoc*, *9*(2), 439–456. Disponível em: http://www.nature.com/articles/nprot.2014.025. Acesso em: Julho de 2023.

Lima, J. P. d. O. (2021). Sistemas complexos aplicado a modelos epidemiológicos. *Physicae Organum - Revista dos Estudantes de Física da UnB*, *7*(1), 59–71. Disponível em: https://periodicos.unb.br/index.php/physicae/article/view/36012 Acesso em: Julho de 2023.

Liu, J.; Ong, G. P.; Pang, V. J. (2022). Modelling effectiveness of COVID-19 pandemic control policies using an Area-based

SEIR model with consideration of infection during interzonal travel. *Transp Res Part A Policy Pract*, V.*161*, 25–47. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0965856422001197. Acesso em: Julho de 2023.

López, L.; Rodó, X. (2021). A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. *Results Phys*, V.*21*, 103-746. Disponível em: https://linkinghub.elsevier.com/ retrieve/pii/S2211379720321604. Acesso em: Julho de 2023.

Marinho, P. R. D.; Cordeiro, G. M.; Coelho, H. F.C.; Brandão, S. C. S. (2021). Covid-19 in Brazil: A sad scenario. *Cytokine Growth Factor Reviews*, *58*, 51–54. Disponível em: https://doi.org/10.1016/j.cytogfr.2020.10.010. Acesso em: Julho de 2023.

Marinov, T. T.; Marinova, R. S. (2022). Inverse problem for adaptive SIR model: Application to COVID-19 in Latin America. *Infect Dis Model*, *7*(1), 134–148. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2468042721000828. Acesso em: Julho de 2023.

Martin, O. A., Kumar, R., Lao, J. (2021). *Bayesian Modeling and Computation in Python*. Chapman and Hall/CRC, Boca Raton.

Minter, A.; Retkute, R. (2019). Approximate Bayesian computation for infectious disease modelling. *Epidemics*, v. *29*, 100-368. Disponível em: https://doi.org/10.1016/j.epidem.2019.100368. Acesso em: Julho de 2023.

Morando, N.; Sanfilippo, M.; Herrero, F.; Iturburu, M.; Torti, A.; Gutson, D.; Pando, M. A.; Rabinovich, R. D. (2022). Evaluación de intervenciones durante la pandemia COVID-19: desarrollo de un modelo basado en subpoblaciones con distintas tasas de contacto. *Revista Argentina de Microbiología*, Vol. *54*(2), p. 81–94. Disponível em: https://doi.org/10.1016/j.ram.2021.04.004. Acesso em: Julho de 2023.

Morens, D. M., Fauci, A. S. (2020). Emerging Pandemic Diseases: How We Got to COVID-19. *Cell*, *182*(5), 1077–1092. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0092867420310126. Acesso em: Julho de 2023.

Muñoz-Fernández, G. A., Seoane, J. M., Seoane-Sepúlveda, J. B. (2021). A SIR-type model describing the successive waves of COVID-19. *Chaos, Solitons & Fractals*, V.*144*, 110-682. Disponível em: https://www.sciencedirect.com/science/article/pii/S0960077921000357. Acesso em: Julho de 2023.

Ng, T. W.; Turinici, G., Danchin, A. (2003). A double epidemic model for the SARS propagation. *BMC Infectious Diseases*, *3*(1), Disponível em: https://doi.org/10.1186/1471-2334-3-19. Acesso em: Julho de 2023.

Paul, A. K., Kuddus, M. A. (2022). Mathematical analysis of a COVID-19 model with double dose vaccination in Bangladesh. *Results Phys*, v.*35*, 105-392. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2211379722001589. Acesos em:

Postnikov, E. B. (2020). Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals*, v.*135*, p. 109-841. Disponível em: https://www.sciencedirect.com/science/article/pii/S0960077920302411. Acesso em: Julho de 2023.

Qiu, H., Wang, Q., Wu, Q., Zhou, H. (2022). Does flattening the curve make a difference? An investigation of the COVID-19 pandemic based on an SIR model. *Int Rev Econ Financ*, Vol. *80*, p. 159–165. Disponível em https://www.sciencedirect.com/science/article/pii/S1059056022000831. Acesso em: Julho de 2023.

Rao, V. S. H.; Upadhyay, R. K. (2013). Modeling the Spread and Outbreak Dynamics of Avian Influenza (H5N1) Virus and Its Possible Control. Em: *Dyn. Model. Infect. Dis.*, Springer New York, New York, NY, pp. 227–250. Disponível em: http: //link.springer.com/10.1007/978-1-4614-9224-5{_}9. Acesso em: Julho de 2023.

Singh, A., Arquam, M. (2022). Epidemiological modeling for COVID-19 spread in India with the effect of testing. *Phys A Stat Mech its Appl*, v.*592*, p. 126-774. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0378437121009596. Acesso em: Julho de 2023.

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., Dessimoz, C. (2013). Approximate bayesian computation. *PLoS Computational Biology*, v.*9*(1), e1002,803. Disponível em: https://doi.org/10.1371/journal.pcbi.1002803. Acesso em: Julho de 2023.

Toni, T.; Welch, D.; Strelkowa, N.; Ipsen, A.; Stumpf, M. P. (2008). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface*, v.*6* n°31, 187–202. Disponível em: https://doi.org/10.1098/rsif.2008.0172. Acesso em: Julho de 2023.

Ujvari, S. C. (2020). *História Das Epidemias*. Contexto.

Umar, M.; Sabir, Z.; Raja, M. A. Z.; Sánchez, Y. G. (2020). A stochastic numerical computing heuristic of SIR nonlinear model based on dengue fever. *Results Phys*, v.*19*, 103-585. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/ S2211379720320258. Acesso em: Julho de 2023.

Weinstein, S. J., Holland, M. S., Rogers, K. E., Barlow, N. S. (2020). Analytic solution of the SEIR epidemic model via asymptotic approximant. *Phys D Nonlinear Phenom*, Vol.*411*, p. 132-633. Disponível em: https://www.sciencedirect.com/science/article/pii/S016727892030453X. Acesso em: Julho de 2023.

Wintachai, P.; Prathom, K. (2021). Stability analysis of SEIR model related to efficiency of vaccines for COVID-19 situation.

Heliyon, *7*(4), e06,812. Disponível em: https://doi.org/10.1016/j.heliyon.2021.e06812. Acesso em: Julho de 2023.

Authorship contributions

1 – Lucas Martins Inez

Graduated in Biochemistry, Bachelor's degree in Chemical Engineering. Currently pursuing a Master's degree in Chemical Engineering
https://orcid.org/0009-0007-8781-3942 • lucasinez94@gmail.com
Contribution: Conceptualization; Data curation; Formal Analysis; Investigation; Methodology; Software; Visualization; Writing – original draft

2 – Carlos Eduardo Rambalducci Dalla

Graduated and master in Chemical Engineering, Currently, Ph.D student in Mechanical Engineering
https://orcid.org/0000-0002-8078-6554 • carloseduardodalla@gmail.com
Contribution: Conceptualization; Data curation; Formal Analysis; Investigation; Methodology; Software; Visualization; Writing – original draft

3 – Wellington Betencurte da Silva

Professor at Department of Rural Engineering. Graduated in Mathematics, master in Mechanical Engineering and a doctor in Mechanical Engineering
https://orcid.org/0000-0003-2242-7825 • wellingtonufes@gmail.com
Contribution: Conceptualization; Funding acquisition; Supervision; Writing – review & editing

4– Julio Cesar Sampaio Dutra

Professor at Department of Rural Engineering. Bachelor's degree in Chemical Engineering and a Ph.D in Chemical Engineering
https://orcid.org/0000-0001-6784-4150 • julio.dutra@ufes.br
Contribution: Conceptualization; Methodology; Validation; Writing – review & editing

5 – José Mir Justino da Costa

Professor, graduated in mathematics, master in mathematics and doctor in Mechanical Engineering
https://orcid.org/0000-0001-5719-4377 • zemir@ufam.edu.br
Contribution: Conceptualization; Writing – review & editing

How to quote this article

Inez, L. M.; Dalla, C. E. R; Silva, W. B. da.; Dutra, J. C. S.; & Costa, J. M. J. da.(2023) Selection of models and parameter estimation for monitoring the COVID-19 epidemic in Brazil via bayesian inference. *Ciência e Natura*, Santa Maria, *45*(spe. 3) e73812. DOI: https://doi.org/10.5902/2179460X73812. Available from: https://doi.org/10.5902/2179460X73812. Accessed in: day month abbr. year.