Epidemic Modeling of COVID-19 in Ceará: Parameter Estimation in a Reduced SIR Model Using Computational Intelligence Metaheuristics

João Samuel Maciel de Sales
Departamento de Computação
Universidade Federal do Ceará
Fortaleza, Brasil
joaosamuel.2001@gmail.com

Ricardo Coêlho
Departamento de Estatística e Matemática Aplicada
Universidade Federal do Ceará
Fortaleza, Brasil
rcoelhos@dema.ufc.br

João Paulo do Vale Madeiro
Departamento de Computação
Universidade Federal do Ceará
Fortaleza, Ceará
jpaulo.vale@dc.ufc.br

***Abstract*—A comprehensive analysis of COVID-19 transmission dynamics in Ceará, Brazil, is presented using a reduced SIR epidemiological model calibrated with empirical data from March 2020 to March 2023. Several computational intelligence optimization algorithms were systematically applied and compared for parameter estimation. Results demonstrate that temporal segmentation, particularly with bimonthly intervals, significantly improves model accuracy, achieving the lowest mean squared errors. The temporal evolution of the basic reproduction number revealed substantial variability, reflecting periods of uncontrolled transmission as well as the impact of public health interventions. These findings reinforce the value of computational intelligence for robust and accurate epidemiological modeling, offering insights to support effective public health strategies during epidemic scenarios.**

***Index Terms*—COVID-19, SIR model, parameter estimation, computational intelligence.**

I. Introduction

Infectious diseases have long posed a significant threat to human populations, often causing considerable morbidity and mortality. Viral infectious diseases, in particular, stand out due to their high transmissibility and their potential to trigger severe public health crises [1]. The global spread of the novel coronavirus disease (COVID-19), caused by the SARS-CoV-2 virus, illustrates the catastrophic consequences caused by viral pandemics.

The COVID-19 pandemic has underscored the essential role of mathematical modeling as a key tool for analyzing epidemic dynamics, forecasting trends, and informing public health interventions [1]. Among the various epidemiological models, the compartmental Susceptible-Infected-Recovered (SIR) framework is particularly prominent due to its mathematical tractability and its ability to capture the essential dynamics of infectious disease transmission, including COVID-19 [2].

One of the main challenges in applying these models is the accurate estimation of critical parameters, such as the transmission and recovery rates. This task is often complicated by data limitations, changing epidemic conditions, and the inherent stochasticity of infectious disease outbreaks [3]. In this context, Computational Intelligence (CI) techniques, especially metaheuristic optimization algorithms like Differential Evolution (DE), Harmony Search (HS), and Particle Swarm Optimization (PSO), have demonstrated substantial effectiveness in parameter estimation, leading to more robust and accurate model calibration [2], [4]–[6].

This study aims to analyze the transmission dynamics of COVID-19 in Ceará, Brazil, through the estimation of epidemiological parameters using several CI methods. By calibrating the SIR model with empirical data from March 2020 to March 2023, this work aims to provide a quantitative assessment of the epidemic's evolution. Accurate parameter estimation is essential to generate reliable projections and support public health authorities in designing and implementing effective COVID-19 control strategies.

The structure of this study is as follows. Section II presents an overview of epidemiological models, with emphasis on compartmental models. Section III describes the systematic review conducted to identify relevant studies on CI methods for parameter estimation. Section IV details the optimization algorithms employed in this work. Section V introduces the database, experimental design, and results of the computational experiments, while Section VI discusses the conclusions and outlines directions for future work.

II. Overview of Epidemiological Models

Epidemiological models are fundamental instruments for understanding and quantifying the transmission dynamics of infectious diseases in human populations. These mathematical frameworks aim to describe, analyze, and predict the progression of diseases, thereby providing essential insights for public health interventions and evidence-based policy-making [1]. Among the broad spectrum of models available, compartmental models have become especially prominent due to their structured approach, which divides the population into distinct compartments according to disease status.

The simplest compartmental models, such as the Susceptible-Infected (SI) and Susceptible-Infected-Susceptible (SIS) frameworks, are primarily applied to diseases that confer no long-term immunity or only temporary protection upon

recovery. However, for infectious diseases such as COVID-19, which typically confer at least temporary immunity following infection, the SIR model is particularly suitable. This model captures the essential stages of disease progression, susceptibility, infection, and recovery (or removal), and its analytical tractability enables meaningful epidemiological inferences [1], [7].

More elaborate models, including Susceptible-Infected-Recovered-Susceptible (SIRS) and Susceptible-Exposed-Infected-Recovered (SEIR), introduce additional compartments to account for loss of immunity or latent periods, but the SIR model remains widely adopted due to its balance between biological realism and mathematical simplicity. Mathematically, the classical SIR model is governed by the following set of ordinary differential equations (ODEs):

$$\frac{dS(t)}{dt} = -\alpha \frac{SI}{N}, \tag{1}$$

$$\frac{dI(t)}{dt} = \alpha \frac{SI}{N} - \beta I, \tag{2}$$

$$\frac{dR(t)}{dt} = \beta I, \tag{3}$$

where $S(t)$, $I(t)$, and $R(t)$ denote the numbers of susceptible, infected, and recovered individuals at time t, respectively. The parameter α represents the transmission rate, while β denotes the recovery or removal rate.

The total population N is assumed to be constant, such that $S(t) + I(t) + R(t) = N$ for all t. An important quantity derived from these parameters is the basic reproduction number (ξ_0), defined as the expected number of secondary cases generated by a single infected individual in a fully susceptible population:

$$\xi_0 = \frac{\alpha}{\beta}. \tag{4}$$

If $\xi_0 > 1$, the disease will tend to spread within the population, while $\xi_0 < 1$ implies eventual extinction of the outbreak.

III. Systematic Review

A systematic review was conducted to identify and analyze CI methods used for the estimation of epidemiological parameters in compartmental models, with a special focus on applications to COVID-19 and the SIR model. The review was based on a transparent and reproducible methodology, employing a targeted search strategy across three major international databases: Web of Science, Scopus, and PubMed.

Three groups of descriptors were combined: (i) epidemiological modeling terms ("SIR" and "model"), (ii) disease identifiers ("COVID-19" or "SARS-CoV-2"), and (iii) parameter estimation keywords (("estimation" or "estimating") and "parameters"). These terms were applied to the titles, abstracts, and keywords of the publications.

The initial search yielded a total of 322 documents: 150 from Web of Science, 121 from Scopus, and 51 from PubMed. After removing duplicates using the Bibliometrix package in R, 157 unique articles remained. Excluding non-English publications resulted in a final set of 156 documents published between January 2020 and December 2022.

The systematic review was concluded on December 31, 2022; consequently, studies published after this date were not included. Although relevant publications may have emerged subsequently, their inclusion would have necessitated a complete new round of searching and analysis, thereby exceeding the scope of the present study. This temporal cut-off ensures methodological consistency and transparency in the conduct of the review.

After both qualitative and quantitative evaluation, it was found that CI methods, especially metaheuristic optimization algorithms, are frequently applied for parameter estimation in SIR-type models within the context of COVID-19. These approaches are particularly valuable for complex optimization scenarios where analytical solutions are impractical or unavailable. Five studies were selected on the basis of their direct application of metaheuristic algorithms to SIR parameter estimation, their rigorous empirical validation, and their representation of diverse geographic contexts. Based on this selection, the following articles were deemed especially influential for informing the CI framework adopted in this research, each illustrating a distinct application of CI in epidemiological modeling.

Reference [2] applied random search and DE algorithms to estimate the parameters of the SIR model using the real COVID-19 data from Santiago, Chile. Their results demonstrated the robustness and accuracy of DE in fitting the model to observed epidemic curves, establishing DE as a reliable approach for epidemiological parameter estimation.

Reference [4] investigated the application of stochastic evolutionary methods, specifically DE and PSO, to estimate the effective reproduction number within the SIR framework. Their research, conducted on datasets from two Malaysian islands, demonstrated the strength of these algorithms in addressing parameter uncertainty and variability in epidemic trends.

Reference [5] aimed to formulate parameter estimation as an optimization problem and to determine optimal or near-optimal parameters for the SIR model using the DE method. Furthermore, the study assesses the predictive effectiveness of this epidemiological model in the five most affected states in India, specifically those with the highest numbers of COVID-19 cases.

Reference [6] focused on the HS algorithm and its variants for optimizing parameters in compartmental models. By testing ten algorithmic variants across multiple COVID-19 datasets, they validated the versatility and competitive performance of HS in epidemiological contexts, particularly for navigating high-dimensional and complex search spaces.

Reference [8] introduced a hybrid methodology integrating the DE algorithm with the Adams-Bashforth-Moulton numerical method to estimate SIR parameters for predicting the COVID-19 outbreak in India. Their study highlighted the importance of coupling optimization with advanced numerical

integration techniques to better capture the effects of public health interventions, such as lockdowns.

Although some of these works employ more than one CI method, each approach was presented in isolation; this review, in turn, lays the groundwork for a unified comparison of all five techniques by applying them to the same dataset under a standardized evaluation protocol.

IV. Computational Intelligence Methods

The CI optimization methods have emerged as essential and versatile tools for tackling complex parameter estimation challenges, especially in fields where analytical solutions are infeasible and the objective landscape is highly nonlinear or multimodal. In epidemiological modeling, these approaches facilitate the calibration of compartmental models to empirical data, enabling accurate characterization of infectious disease dynamics. The adoption of these techniques in this research is substantiated by the findings of the systematic review, which highlighted their growing prominence and effectiveness for epidemiological parameter estimation.

This section provides an overview of the primary CI optimization methods employed in this study: Global Random Search (GRS), Local Random Search (LRS), HS, DE, and PSO. Each method is described in terms of its structure, mathematical formulation, operational mechanism, and specific suitability for epidemiological contexts.

A. Random Search Methods

Random search algorithms offer a straightforward yet effective approach to optimization by exploring the solution space stochastically. Their inherent simplicity makes them valuable as baseline or benchmark methods, facilitating comparative assessment of more advanced metaheuristics. In this research, two variants are considered: GRS and LRS.

Specifically, GRS serves primarily as an exploration mechanism, generating candidates uniformly across the entire search space, while LRS operates as an exploitation mechanism, refining solutions locally by perturbing the current best point with Gaussian noise.

GRS samples candidate solutions uniformly across the entire search space:

$$x_{\text{candidate}} \sim U(x_{\min}, x_{\max}), \tag{5}$$

where $x_{\text{candidate}}$ is a newly generated candidate solution, $U(\cdot,\cdot)$ denotes the uniform probability distribution, and $x_{\min}$ and $x_{\max}$ are the lower and upper bounds of the search space, respectively. This formulation ensures that every point within the feasible domain has an equal probability of being selected during the search process.

In contrast, LRS focuses the search in the vicinity of the current best solution, employing Gaussian perturbations to promote convergence toward local optima:

$$x_{\text{candidate}} = x_{\text{best}} + n, \quad n \sim N(0, \sigma^2 I), \tag{6}$$

where $x_{\text{candidate}}$ is the candidate solution generated in the current iteration, x_{best} denotes the best solution found so far, and n represents a noise vector sampled from a multivariate normal (Gaussian) distribution with zero mean and covariance matrix $\sigma^2 I$. Here, σ^2 is the variance controlling the step size of the local search, and I is the identity matrix. This approach enables the algorithm to explore the local neighborhood of the current best solution, increasing the likelihood of identifying improved solutions in its vicinity.

B. Harmony Search

The HS method is a population-based metaheuristic inspired by the improvisational process of musicians [6], [9], [10]. This algorithm systematically explores the solution space through key mechanisms such as harmony memory consideration rate (HMCR), pitch-adjusting rate (PAR), and bandwidth (BW). These parameters collectively balance diversification and intensification, facilitating both global and local search.

In HS algorithm, the generation of a new candidate solution $x_{i,k}$ for each decision variable (indexed by i in the k-th harmony/candidate) involves a sequence of conditional operations:

$$x_{i,k} = \begin{cases} x_{i,l} + U(0,1) \cdot (x_{i,u} - x_{i,l}) & \text{if } U(0,1) > HMCR \\ x_{i,j} & \text{otherwise} \end{cases}, \tag{7}$$

where $x_{i,k}$ denotes the i-th variable of the k-th candidate solution (the new harmony), while $x_{i,j}$ refers to the i-th variable of the j-th solution currently stored in the harmony memory, with j randomly selected among all stored solutions. The terms $x_{i,l}$ and $x_{i,u}$ represent the lower and upper bounds, respectively, for the i-th variable. The notation $U(0,1)$ corresponds to a random number drawn from the uniform distribution on the interval $[0,1]$. The parameter HMCR controls the probability of selecting a value from the harmony memory instead of generating a new random value within the variable bounds.

If the value is chosen from harmony memory (second case of (7)), it may be further adjusted:

$$x_{i,k} = x_{i,k} + (2 \cdot U(0,1) - 1) \cdot \text{BW} \quad \text{if } U(0,1) \leq \text{PAR}, \tag{8}$$

where PAR controls the probability of applying a pitch adjustment to a variable selected from harmony memory, and BW determines the maximum possible change introduced to $x_{i,k}$ during the pitch adjustment step. This sequential structure ensures that pitch adjustment is only applied to variables inherited from memory, preserving the intended exploration–exploitation balance of the HS algorithm.

C. Differential Evolution

The DE method is a robust evolutionary algorithm renowned for its ability to efficiently optimize nonlinear, high-dimensional problems [2]. DE operates via iterative processes of mutation, crossover, and selection, facilitating both the exploration and exploitation of the search space [11].

The algorithm begins by initializing a population of N_p candidate solutions, each represented by a D-dimensional vector $x_{i,G}$, where $i = 1, 2, ..., N_p$ indexes the individuals and G denotes the generation. The initial population is typically

generated by sampling each variable uniformly within its prescribed bounds.

During each iteration, mutation generates a new candidate vector $v_{i,G+1}$ by perturbing an existing solution with a scaled difference between two other randomly selected solutions:

$$v_{i,G+1} = x_{r_1,G} + F \cdot (x_{r_2,G} - x_{r_3,G}), \tag{9}$$

where $x_{r1,G}$, $x_{r2,G}$, and $x_{r3,G}$ are three distinct individuals randomly chosen from the current population, F is the mutation factor controlling the amplification of the differential variation, and $v_{i,G+1}$ is the mutant vector.

Next, crossover creates a trial vector $u_{i,G+1}$ by mixing components from the mutant vector $v_{i,G+1}$ and the target vector $x_{i,G}$:

$$u_{j,i,G+1} = \begin{cases} v_{j,i,G+1}, & \text{if } rand_{j,i} \leq Cr \text{ or } j = I_{rand}, \\ x_{j,i,G}, & \text{otherwise} \end{cases} \tag{10}$$

where $u_{j,i,G+1}$ is the j-th component of the trial vector for individual i at generation $G+1$, Cr is the crossover rate, $rand_{j,i}$ is a random number uniformly distributed in $[0,1]$ for each component, and I_{rand} ensures that at least one component is inherited from the mutant vector.

Finally, selection compares the trial vector and the original target vector, retaining the one with the better fitness value:

$$x_{i,G+1} = \begin{cases} u_{i,G+1}, & \text{if } f(u_{i,G+1}) \leq f(x_{i,G}) \\ x_{i,G}, & \text{otherwise} \end{cases}, \tag{11}$$

where $f(\cdot)$ denotes the objective (fitness) function. In this formulation, it is presented under the convention of minimization; for maximization problems, the comparison criterion would be reversed accordingly.

This population-based approach, with its simple yet powerful mechanisms for exploration and exploitation, has demonstrated strong performance in parameter estimation for SIR-type epidemiological models [2], [4], [5], [8].

D. Particle Swarm Optimization

The PSO method is a swarm-based metaheuristic inspired by the collective behaviors of natural populations, such as bird flocking and fish schooling [12]. The algorithm begins by initializing a swarm of N_p particles, where each particle i is characterized by a position vector $x_{i,t}$ and a velocity vector $v_{i,t}$ in the D-dimensional search space at iteration t. The initial positions are typically sampled uniformly within the predefined bounds for each dimension, and initial velocities are often set to zero or to small random values.

At each iteration, the velocity of each particle is updated according to three components: inertia, which preserves the previous velocity; a cognitive term, which attracts the particle toward its own historical best position; and a social term, which pulls the particle toward the global best position found by any particle in the swarm. The velocity update equation is given by

$$v_{i,t+1} = w \cdot v_{i,t} + c_1 \cdot r_1 \cdot (x_{\mathrm{b},i,t} - x_{i,t}) + c_2 \cdot r_2 \cdot (x_{\mathrm{b},g,t} - x_{i,t}), \tag{12}$$

where $v_{i,t}$ is the velocity of particle i at iteration t, w is the inertia weight controlling the influence of the previous velocity, c_1 and c_2 are non-negative coefficients representing the cognitive and social acceleration factors, respectively, and r_1 and r_2 are independent random variables sampled uniformly from the interval $[0,1]$. The term $x_{best,i,t}$ denotes the best position visited so far by particle i (personal best), while $x_{\text{best},g,t}$ indicates the best position found by the entire swarm (global best) up to iteration t.

After updating the velocity, the new position of each particle is determined by

$$x_{i,t+1} = x_{i,t} + v_{i,t+1}, \tag{13}$$

where $x_{i,t}$ is the current position of particle i, and $v_{i,t+1}$ is the updated velocity. This process allows particles to dynamically balance the exploration of new areas in the search space and the exploitation of previously found high-quality solutions.

V. Computational Experiments

This section details the COVID-19 database utilized for the case study and describes the computational framework adopted for parameter estimation and numerical simulation experiments.

A. COVID-19 Database

The COVID-19 case data for Ceará were sourced from the official records published by the Brazilian Ministry of Health [13]. The dataset comprises daily counts of new and cumulative confirmed cases, spanning from March 17, 2020, to March 16, 2023, for a total of 1,095 days, corresponding to the initial three years of the pandemic in the state.

Specific preprocessing procedures were undertaken to ensure data consistency and reliability. Initially, four dates (September 5, 2020; February 19, 2021; September 21, 2021; and March 4, 2023) exhibited negative daily case counts, resulting in a decrease in the cumulative series on those days. For each quarter containing an anomaly, the total negative deviation was calculated and then proportionally redistributed across all days of that period, repeating the adjustment until the cumulative deviation was fully eliminated.

Subsequently, every interval of one or more consecutive days with zero reported new cases, whether arising from the above correction or occurring elsewhere, was addressed by taking the total cases reported on the next nonzero day, dividing that value by the length of the preceding zero-case interval, and allocating the resulting mean evenly across those days. This ensures that the overall incident case count is preserved while removing flat segments from the cumulative trajectory.

These two sequential data-cleaning steps are essential for preventing reporting artifacts from biasing model calibration and ensuring that the parameter estimation process accurately reflects the underlying epidemic dynamics in Ceará. No referenced study has applied an identical imputation for zero-case intervals; consequently, the procedure was defined empirically following general time-series smoothing and imputation practices.

B. SIR Model Configuration

In this study, special attention is given to a reduced formulation of the SIR epidemiological model, with the objective of accurately describing the cumulative COVID-19 cases in Ceará. While the standard SIR model is typically expressed as a system of three coupled differential equations [14], algebraic manipulation allows for its simplification into a single ordinary differential equation [2].

The reduced SIR model investigated in this work is defined as follows:

$$\begin{aligned} \frac{dv(t)}{dt} &= -\alpha \cdot v(t) + \nu \cdot \xi_0 \cdot \left[1 - \exp\left(-\frac{\alpha}{\nu} v(t)\right)\right] + I_0, \\ v(t_0) &= 0 \end{aligned} \tag{14}$$

where $v(t)$ denotes the cumulative number of infected individuals at time t, and $v(t_0)$ is the initial cumulative number of infections at the starting time t_0. The parameters of the model are defined as follows: α is the recovery rate; β is the transmission rate; and $\nu = \alpha/\beta$ is the ratio between recovery and transmission. The parameter ξ_0 represents the initial reproduction number, while I_0 denotes the initial number of infectious individuals at t_0.

This one-equation model is derived from integrating and simplifying the two-equation version of the SIR system, leveraging the conservation property of the total population. Its primary advantage lies in computational efficiency, as it directly models the cumulative incidence and requires the estimation of fewer state variables. Further details regarding the derivation, equivalence, and practical application of this reduced SIR model can be found in [2], [15] and related literature.

C. Parameter Estimation

The parameter estimation process for the reduced SIR model, expressed as a single differential equation, was formulated as an optimization problem. In this context, the decision variables $x = \{x_1, \ldots, x_D\}$ correspond to the model parameters to be estimated, with D denoting the total number of parameters in the system. Parameter estimation was performed by minimizing the following objective function:

$$\min f(v) = \frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{v(t_n)}{v_n} - 1 \right)^2, \tag{15}$$

where N is the total number of days in the dataset, $v(t_n)$ is the numerically simulated cumulative number of infected individuals at time t_n, and v_n is the observed value on day t_n. The numerical solution of the model was obtained using a fourth-order predictor–corrector method [16].

The objective function follows the formulation in [2]. This formulation corresponds to the mean squared relative error between simulated and observed values, providing a dimensionless, scale-invariant metric that balances fitting across low- and high-incidence periods. Lower and upper bounds were established for each parameter to guarantee biological plausibility and methodological consistency, using intervals reported in [2]. These bounds are presented in Table I.

TABLE I
PARAMETER BOUNDS FOR SIR MODEL

Parameter	Interval
α	$[10^{-3}, 10^{-1}]$
ν	$[1.000, 30.000]$
t_0	$[1, 50]$
I_0	$[0, 30]$
ξ_0	$[1, 7]$

TABLE II
IC METHODS CONTROL PARAMETER INITIALIZATION

Parameter	GRS	LRS	HS	DE	PSO
N_p	15	15	–	15	15
N_{max}	1000	1000	1000	1000	1000
HMS	–	–	15	–	–
$HMCR$	–	–	0.95	–	–
PAR	–	–	0.3	–	–
BW	–	–	0.01	–	–
F	–	–	–	$U[0.5, 1]$	–
C_r	–	–	–	0.7	–
w	–	–	–	–	0.5
c_1	–	–	–	–	2
c_2	–	–	–	–	2
r_1	–	–	–	–	$U[0, 1]$
r_2	–	–	–	–	$U[0, 1]$

Parameter estimation employed several CI methods, each initialized with control settings drawn from the literature to promote convergence and maintain solution diversity. Table II provides the complete configuration for each algorithm. Population size ($N_p = 15$) and maximum iterations ($N_{\max} = 1000$) were adopted from [2], with harmony memory size (HMS) for HS method set equal to N_p. DE method uses a mutation factor $F \sim U[0.5, 1]$ and a crossover rate $C_r = 0.7$ [2]. HS parameters, memory consideration rate ($HMCR = 0.95$), pitch adjusting rate ($PAR = 0.3$) and bandwidth ($BW = 0.01$), were taken from [6]. PSO employs inertia weight $w = 0.5$, cognitive coefficient $c_1 = 2$, social coefficient $c_2 = 2$, and random factors $r_1, r_2 \sim U[0, 1]$, determined empirically. For the DE method, a tournament selection strategy was implemented: In each generation, four individuals are randomly chosen (with replacement), and the one with the lowest mean squared error is selected for the next generation. Additionally, the two best and three worst individuals (in terms of mean squared error) from each generation are automatically carried over to the next. The remaining ten individuals are selected via a tournament.

D. Results

Computational experiments were systematically performed to assess the efficiency of CI methods in estimating parameters of the reduced SIR model for the COVID-19 epidemic in Ceará. All code was developed in Python 3.13.1 and executed in Visual Studio Code, utilizing a Lenovo 83EU laptop with a 12th Gen Intel Core i5-12450H processor (2.0 GHz, 8 cores, 12 threads) and 16 GB RAM, running a 64-bit Windows system.

The simulation protocol explored several temporal segmentation strategies, including analyses for the entire period,

annual, semiannual, quarterly, and bimonthly intervals. These partitions were chosen to address two main challenges: (i) the full-period simulation frequently resulted in a sigmoidal plateau, making it difficult for the model to accurately reflect the evolving epidemic curve; and (ii) COVID-19 is known for its seasonality and shifting epidemiological patterns in Brazil, which warranted a more granular temporal approach. The same random seed was used for all methods to ensure fair comparison and reproducibility.

Each CI method was executed ten times for each segment within the chosen temporal partitioning. For example, in the annual partitioning (comprising three years), each year received 10 runs, totaling 30 executions for this segmentation. Similarly, there were 60 runs for the semiannual partition (six semesters), 90 for the four-monthly (nine periods), 120 for the quarterly (twelve quarters), and 180 for the bimonthly (eighteen periods). Including the 10 runs for the full period, this amounted to 2,450 simulations across all segmentations and methods.

The effectiveness of each optimization strategy was evaluated using the Mean Squared Error (MSE) between simulated and observed data, enabling the selection of optimal parameter sets for each temporal segmentation.

Table III presents the total execution time required to complete all runs for each temporal partition, rather than the average per run. Each partition underwent multiple executions (for example, the bimonthly partition comprised 180 runs), and the values in Table III correspond to the cumulative duration of those runs. Methods such as GRS, LRS, HS, and PSO consistently demonstrated low computational overhead, with total execution times per partition remaining below 17 seconds. In contrast, DE exhibited higher computational demands, with cumulative times ranging from approximately 24 to 46 seconds depending on the segmentation granularity. This difference underscores the higher computational demands of DE in comparison to the other approaches. The experimental framework, which integrated diverse temporal segmentation, computational resources, and error analysis, provided a systematic basis for comparing CI algorithms and refining parameter estimation for COVID-19 dynamics in Ceará.

Table IV presents the best MSE achieved by each CI method across different temporal segmentation scenarios for the reduced SIR model. The results reveal a clear trend: As the temporal segmentation becomes finer, from the entire period to bimonthly intervals, the best-case MSE values decrease consistently for all methods. This demonstrates that segmenting the analysis period allows the model to better adapt to local epidemic fluctuations, improving the overall fit to the observed data. For the bimonthly segmentation, Table IV shows that both GRS, LRS, HS, and PSO achieved the lowest MSE, with a value of 0.0024.

Table V presents the average MSE for each CI method across all temporal segmentations. The results indicate a general trend of decreasing average MSE as the segmentation becomes more granular, reflecting improved model adaptation to the epidemic's local fluctuations. In the bimonthly segmentation, the methods HS and PSO achieved identical average MSE values, while LRS and DE obtained very similar results. This indicates that, under the finest segmentation, the choice of CI algorithm becomes less critical, as all tested methods are capable of producing highly accurate parameter estimates for the reduced SIR model.

Table VI summarizes the worst MSE values obtained for each CI method under different segmentation strategies. The results indicate a general trend of decreasing worst-case errors as the temporal segmentation becomes more granular, reflecting improved model adaptability. An exception occurs with the LRS method, which shows a slight increase in worst MSE from four-monthly to quarterly segmentation. For the bimonthly interval, HS and PSO consistently achieve the lowest worst MSE values, indicating greater robustness in delivering accurate parameter estimates even under less favorable runs. In contrast, GRS and LRS present slightly higher worst-case errors, reinforcing the advantage of metaheuristics such as HS and PSO in finely segmented epidemic modeling scenarios.

A comprehensive comparison of the best-case, average-case, and worst-case MSE values across temporal segmentations reveals that the bimonthly segmentation consistently produced the lowest error values for all CI methods, indicating that finer temporal partitioning enhances the model's fit to real COVID-19 data in Ceará.

In this most granular scenario, HS and PSO outperformed the other algorithms, achieving the lowest best MSE and sharing the lowest average and the lowest worst MSE values. These results demonstrate that, when applied to short bimonthly intervals, HS and PSO provide the most robust and accurate parameter estimation for the reduced SIR model. In summary, the use of bimonthly segmentation combined with HS or PSO proved to be the most effective approach for epidemic modeling in this context.

TABLE III
EXECUTION TIME (SECONDS) ACROSS TEMPORAL SEGMENTATIONS

Segmentation	GRS	LRS	HS	DE	PSO
Entire	7.8	8.9	9.9	24.5	8.1
Annual	7.7	8.4	7.8	26.2	7.9
Semiannual	8.4	11.5	12.2	29.9	11.8
Four-monthly	8.4	9.8	10.2	44.6	13.4
Quarterly	8.4	12.2	9.5	38.9	15.3
Bimonthly	9.1	11.5	9.9	45.3	16.5

TABLE IV
BEST-CASE MEAN SQUARED ERROR

Segmentation	GRS	LRS	HS	DE	PSO
Entire	0.1362	0.1315	0.1525	0.1341	0.1442
Annual	0.0289	0.0252	0.0298	0.0293	0.0249
Semiannual	0.0080	0.0084	0.0109	0.0135	0.0120
Four-monthly	0.0047	0.0044	0.0047	0.0045	0.0044
Quarterly	0.0035	0.0039	0.0034	0.0036	0.0037
Bimonthly	0.0024	0.0024	0.0024	0.0025	0.0024

TABLE V
AVERAGE-CASE MEAN SQUARED ERROR

Segmentation	GRS	LRS	HS	DE	PSO
Entire	0.1368	0.1469	0.1611	0.1397	0.1542
Annual	0.0290	0.0331	0.0316	0.0309	0.0289
Semiannual	0.0125	0.0147	0.0147	0.0164	0.0121
Four-monthly	0.0050	0.0050	0.0054	0.0052	0.0053
Quarterly	0.0038	0.0040	0.0036	0.0037	0.0039
Bimonthly	0.0027	0.0026	0.0025	0.0026	0.0025

TABLE VI
WORST-CASE MEAN SQUARED ERROR

Segmentation	GRS	LRS	HS	DE	PSO
Entire	0.1644	0.1665	0.1653	0.1626	0.1653
Annual	0.0389	0.0401	0.0346	0.0393	0.0371
Semiannual	0.0231	0.0207	0.0173	0.0207	0.0130
Four-monthly	0.0086	0.0051	0.0061	0.0061	0.0060
Quarterly	0.0055	0.0054	0.0037	0.0037	0.0040
Bimonthly	0.0036	0.0036	0.0026	0.0028	0.0026

E. Analysis of the Reproduction Number in Ceará

The temporal evolution of the basic reproduction number (ξ_0) for COVID-19 in Ceará was analyzed using the reduced one-equation SIR model and focusing exclusively on the bimonthly segmentation, which provided the most accurate results according to previous analyses. Among the two CI methods that achieved the lowest MSEs in this segmentation, HS and PSO, the HS algorithm was selected to perform the parameter estimation for this analysis.

Table VII presents the estimated ξ_0 values for each bimonthly interval throughout the study period. According to Table VII, the ξ_0 in Ceará displayed considerable fluctuation throughout the pandemic.

During the initial bimonthly periods, ξ_0 remained elevated, with values of 2.75 in the first period and 2.93 in the second, reflecting the early uncontrolled phase of COVID-19 transmission.

TABLE VII
ESTIMATED ξ_0 FOR EACH BIMONTHLY PERIOD IN CEARÁ

Bimonthly Period	Start Date	End Date	ξ_0 Value
1	03/17/2020	05/16/2020	2.7468
2	05/17/2020	07/16/2020	2.9297
3	07/17/2020	09/16/2020	1.4915
4	09/17/2020	11/16/2020	1.3858
5	11/17/2020	01/16/2021	1.5006
6	01/17/2021	03/16/2021	4.6360
7	03/17/2021	05/16/2021	2.0357
8	05/17/2021	07/16/2021	2.2111
9	07/17/2021	09/16/2021	2.7228
10	09/17/2021	11/16/2021	2.1610
11	11/17/2021	01/16/2022	4.1743
12	01/17/2022	03/16/2022	1.2717
13	03/17/2022	05/16/2022	3.2950
14	05/17/2022	07/16/2022	3.5413
15	07/17/2022	09/16/2022	1.3166
16	09/17/2022	11/16/2022	3.6808
17	11/17/2022	01/16/2023	5.4844
18	01/17/2023	03/16/2023	1.8598

After this initial stage, a decline is observed, particularly in periods 3 and 4, where ξ_0 drops to 1.49 and 1.39, respectively, likely as a result of intensified public health interventions and increased population awareness [17]. Striking increases in ξ_0 appear in certain periods.

Notably, a sharp spike to 4.64 occurs in period 6, which may be associated with the second wave of infections in Ceará. Other peaks are observed in periods 11, 13, 14, 16, and 17. These surges may be linked to increased social activity, travel, or relaxation of preventive measures during school holidays and festive seasons, as suggested by local news sources [17].

Conversely, periods of lower ξ_0, such as 12 and 15, reflect moments of better epidemic control, potentially due to the reimplementation of public health interventions, increased vaccination coverage, or reduced community transmission.

This temporal pattern demonstrates that the transmissibility of COVID-19 in Ceará was highly dynamic, with periods of intense spread interspersed with phases of relative control. These findings emphasize the importance of ongoing monitoring and adaptive public health strategies throughout the epidemic.

VI. CONCLUSION

This study provided a comprehensive analysis of the transmission dynamics of COVID-19 in Ceará, Brazil, by leveraging a reduced SIR epidemiological model calibrated with empirical data from March 2020 to March 2023. Through the systematic application and comparison of CI optimization methods, including GRS, LRS, HS, DE, and PSO, the research demonstrated the effectiveness and robustness of metaheuristics for parameter estimation in epidemic modeling.

The experimental results highlighted the critical role of temporal segmentation in improving model accuracy. Notably, bimonthly segmentation enabled the SIR model to adapt more closely to the dynamic fluctuations of the COVID-19 epidemic curve in Ceará. In this scenario, the HS and PSO algorithms consistently achieved the lowest mean squared errors, indicating superior performance and robustness among the tested approaches.

The temporal evolution of the basic reproduction number revealed substantial variability across the study period, with marked increases and decreases that reflected both uncontrolled transmission phases and the impact of public health interventions. These findings emphasize the necessity for ongoing monitoring and flexible strategies to address epidemic waves and changing transmission patterns.

The methodological framework and results presented here reinforce the value of CI-based approaches for epidemiological parameter estimation, especially in complex and rapidly evolving scenarios such as COVID-19. Future work may extend this framework to incorporate additional data sources, more complex compartmental models, or alternative optimization strategies, further enhancing the accuracy and utility of epidemic forecasting tools to support public health decision-making.

As future work, a more methodologically robust evaluation framework would involve withholding a portion of the empir-

ical data for validation, thereby assessing the generalization capacity of the parameters obtained via the metaheuristic algorithms. By calibrating the reduced SIR model on a training subset and subsequently testing its predictive performance on unseen data, it would be possible to determine whether the finely tuned parameters maintain low error rates outside the calibration window. Such an out-of-sample validation protocol is especially critical in epidemic modeling, where the ultimate goal is not merely to reproduce past trends but to forecast future dynamics with reliability. Incorporating this train-test split would thus strengthen the study's conclusions on the practical applicability of CI-based parameter estimation in real-time outbreak scenarios.

ACKNOWLEDGMENT

The following Brazilian research agencies have supported this work: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) — Proc: 140788/2024-7, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) - Proc: 88882.461678/2019-01 e Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) — Proc: BMD-0008-00739.01.03/24.

REFERENCES

[1] S. Schecter, V. Matos, D. Marchesin, and M. López, *Equações diferenciais e modelos epidemiológicos*. Brasil: Instituto de Matemática Pura e Aplicada, 2021.

[2] S. Rica and G. A. Ruz, "Estimating SIR model parameters from data using differential evolution: an application with COVID-19 data," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Chile, 2020, pp. 1–6.

[3] G. Servidone, S. Ternes, M. Rossi, and M. F. Nogueira, "Fuzzy theory use in deterministic epidemiological models," in *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, Brazil, 2017.

[4] W. K. Wong and F. H. Juwono, "Estimating effective reproduction number for sir compartmental model: a stochastic evolutionary approach," *Journal of Social Computing*, no. 2, 2022.

[5] S. S. Panigrahi, A. J. Muthukumar, S. Thangavelu, G. Jeyakumar, and C. Shunmuga Velayutham, *A Comparative Study on Parameter Estimation of COVID Epidemiological Models Using Differential Evolution Algorithm*. Singapore: Springer Nature Singapore, 2022, pp. 241–263.

[6] K. Gopal, L. S. Lee, and H.-V. Seow, "Parameter estimation of compartmental epidemiological model using harmony search algorithm and its variants," *Applied Sciences*, no. 3, 2021.

[7] M. S. Cecconello, M. T. Mizukoshi, and W. A. Lodwick, "Interval nonlinear initial-valued problem using constraint intervals: theory and an application to the SARS-CoV-2 outbreak," *Information Sciences*, 2021.

[8] A. Singh, M. Mehra, and S. Gulyani, "A modified variable-order fractional SIR model to predict the spread of COVID-19 in India," *Mathematical Methods in the Applied Sciences*, 2021.

[9] T. Zhang and Z. W. Geem, "Review of harmony search with respect to algorithm structure," *Swarm and Evolutionary Computation*, 2019.

[10] A. A. Al-Omoush, A. A. Alsewari, H. S. Alamri, and K. Z. Zamli, "Comprehensive review of the development of the harmony search algorithm and its applications," *IEEE Access*, 2019.

[11] R. Storn and K. Price, "Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, 1997.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, 1995.

[13] "Covid-19 casos e óbitos," Ministry of Health, 2023.

[14] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London*, no. 772, 1927.

[15] H. Berestycki, J.-M. Roquejoffre, and L. Rossi, "Propagation of epidemics along lines with fast diffusion," *Bulletin of mathematical biology*, no. 1, p. 2, 2020.

[16] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical analysis*. Boston: Cengage Learning, 2015.

[17] A. Vieira, "Linha do tempo da COVID-19 no Ceará," 2022.