medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Monitoring and forecasting the number of reported and unreported cases of the COVID-19 epidemic in Brazil using Particle Filter

J. C. S. Dutra[1], W. B. da Silva[1] and J. M. J. da Costa[2]

[1] *Chemical Engineering Program, UFES, Alegre, 29500-000, Brazil*
[2] *Statistics Department, UFAM, Manaus, AM, CEP 69077-000, Brazil*

Abstract

In this paper, we combine algorithm of Liu & West for the Particle Filter (PF) with SIRU-type epidemic model to monitor and forecast cases of Covid-19 in Brazil from February up to September. We filter the number of cumulative reported cases and estimate model parameters and more importantly unreported infectious cases (asymptomatic and symptomatic infectious individuals). The parameters under study are related to the attenuation factor of the transmission rate and the fraction of asymptomatic infectious becoming reported as symptomatic infectious. Initially, the problem is analysed through Particle Swarm Optimization (PSO) based simulations to provide initial guesses, which are then refined by means of PF simulations. Subsequently, two additional steps are performed to verify the capability of the adjusted model to predict and forecast new cases. According to the results, the pandemic peak is expected to take place in mid-June 2020 with about 25,000 news cases per day. As medical and hospital resources are limited, this result shows that public health interventions are essential and should not be relaxed prematurely, so that the coronavirus pandemic is controlled and conditions are available for the treatment of the most severe cases.

Keywords: Epidemics modelling; SIRU model; COVID-19; Particle filter; Particle swarm optimization; Bayesian inference.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1 Introduction

The first infection by the new coronavirus (Sars-CoV-2) was detected on December 31, 2019, in Wuhan, China and a pandemic was declared by the World Health Organization on March 11, 2020 [1, 2]. Such infection causes the Covid-19, whose main symptoms include cough, fever, difficulty breathing. It presented a scary evolution with millions of infected people and thousands of deaths in the world. Because these facts, the affected countries and cities around the world have been reacting in different ways, towards locally controlling the disease evolution.

In Brazil, the first case of Covid-19 was confirmed on February 26, 2020. Ever since, the number has surged to more than 310,000 cases of the disease, according to the Brazilian Ministry of Health [3]. Such milestone makes this nation to become the third one with the highest number of infections in the world, behind only the United States and Russia. Specifically, Amazonas, Ceara, Pernambuco, Rio de Janeiro and Sao Paulo have been the worst affected states in Brazil, whose cases correspond to at least 60% of all cases in the country. The coronavirus has so far been responsible for more than 20,000 deaths. On May 21, 2020, an unenviable record was broken with 1,188 new deaths reported officially within a single day. This acceleration of the pandemic shows that Brazil tends to become a new global epicentre of the covid-19 [4].

Figure 1 comprising data up to May 21, 2020 shows that the number of confirmed cases of the Covid-19 in Brazil has significantly exceeded the number of cases in China. This extent of the disease reveals a dramatic situation, since the population of Brazil is approximately 6.5 times smaller than that of China, which implies the urgent need for epidemiological studies to develop the best strategies for public health policies.

Public strategies include general isolation through quarantine and massive testing for focused isolation, with varying degrees of success so far, as can be analysed from the limited data available [5]. In this regard, mathematical modelling is an interesting approach that can allow the evaluation of different scenarios, furnishing information for a proper support for health system decisions. In general, nonlinear dynamics of biological and biomedical systems is the objective of several researches that can be based on mathematical modelling or time series analysis [6]. Literature presents some examples related to the dynamics of infectious diseases. Mathematical equations are widely used to model the nature and impact of global pandemics in the society. In particular, coronavirus propagation can be described by a mathematical model that allows the nonlinear dynamics analysis, representing different populations related to the phenomenon [7]. Recently, the classical susceptible-infectious-recovered (SIR) model proposed in [8, 9], was employed in the analysis of the epidemic outbreak in different countries, including China, South Korea, Germany, Italy, and France [10, 11, 12]. The modelling and evaluating the consequences of public health interventions. It was a direct application of previous developments [13, 14] on the fundamental problem of parameter identification in mathematical epidemic models, accounting for unreported cases.

In this work the SIRU-type model is implemented for the direct problem formulation of the COVID-19 epidemic evolution in Brazil, adding a time variable parametrization for the fraction of asymptomatic infectious that become reported symptomatic individuals. According Cotta *et al.* [5] it is an especially important parameter in the public health measure associated with massive

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

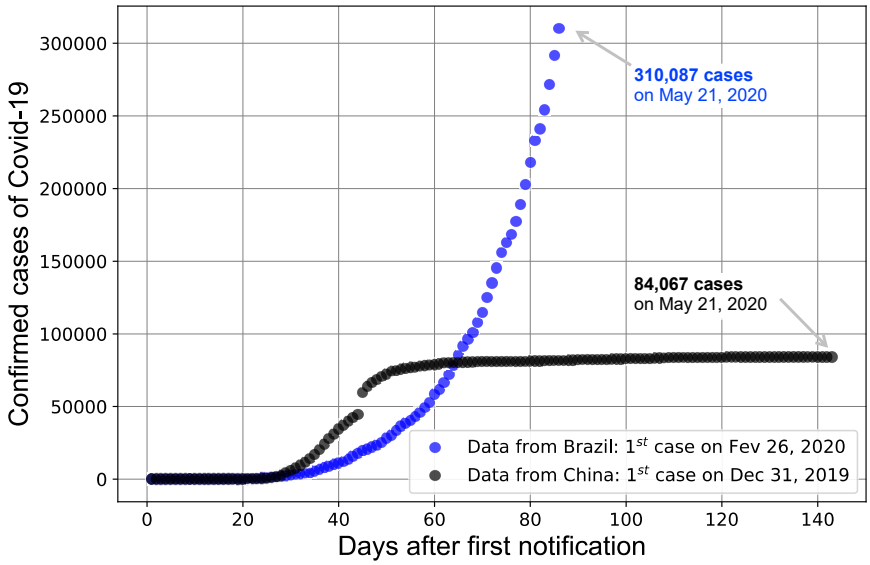

Figure 1: Comparison between the number of confirmed cases of Covid-19 in Brazil (85 days since first case) and in China (142 days since first case) up to May 21, 2020.

testing and consequent focused isolation. The same analytical identification procedure is maintained for the required initial conditions, as obtained from the early stage exponential behaviour. However, a Bayesian inference approach is here adopted for parametric estimation, employing the combined parameter and state estimation algorithm proposed by Liu and Wes and Particle Swarm Optimization (PSO).

2 Mathematical formulation

In this section, we formulate the mathematical SIR model [5, 9, 10, 11, 12, 13, 14] we analyze is the following: at time t, let $S(t)$ be the number of susceptible, $I(t)$ the number of infected, $R(t)$ the number of removed and permanently immune, $U(t)$ and number of unreported symptomatic infectious individuals. The equations for the SIR model

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$\frac{dS(t)}{dt} = -\tau(t)S(t)[I(t)+U(t)] \tag{1a}$$

$$\frac{dI(t)}{dt} = -\tau(t)S(t)[I(t)+U(t)] - \nu I(t) \tag{1b}$$

$$\frac{dR(t)}{dt} = \nu_1(t)I(t) - \eta R(t) \tag{1c}$$

$$\frac{dU(t)}{dt} = \nu_2(t)I(t) - \eta U(t) \tag{1d}$$

where,

$$\nu_1(t) = \nu f(t) \tag{2a}$$

$$\nu_2(t) = \nu(1-f(t)) \tag{2b}$$

with initial conditions

$$S(t_0) = S_0 \tag{3a}$$

$$I(t_0) = I_0 \tag{3b}$$

$$R(t_0) = 0 \tag{3c}$$

$$U(t_0) = U_0 \tag{3d}$$

where t_0 is the beginning date of the epidemic in days. Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Reported symptomatic individuals $R(t)$ are infectious for an average period of $1/\eta$ days, as are unreported symptomatic individuals $U(t)$. According to Cotta *et al.* [5] assumed that reported symptomatic infectious individuals $R(t)$ are reported and isolated immediately, and cause no further infections. The asymptomatic individuals $I(t)$ can also be viewed as having a low-level symptomatic state. All infections are acquired from either $I(t)$ or $U(t)$ individuals. The fraction $f(t)$ of asymptomatic infectious become reported symptomatic infectious, and the fraction $1-f(t)$ become unreported symptomatic infectious. The rate asymptomatic infectious become reported symptomatic is $\nu_1(t) = f(t)\nu$, the rate asymptomatic infectious become unreported symptomatic is $\nu_2(t) = (1-f(t))$, where $\nu_1(t)+\nu_2(t) = \nu$. The transmission rate, $\tau(t)$, is also allowed to be a time variable function along the evolution process. Figure 2 below illustrates the infection process as a flow chart [10].

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

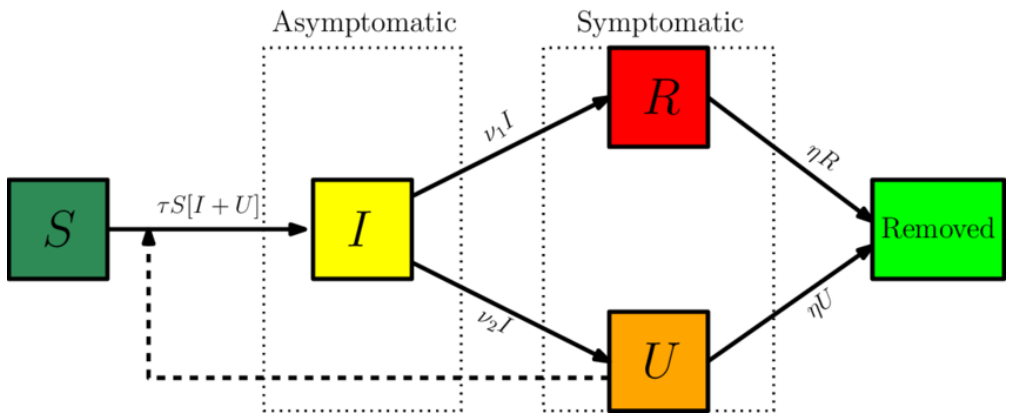

Figure 2: Flow chart illustrating the infection path process.

The time variable coefficients, $\tau(t)$ and $f(t)$, are chosen to be expressed as:

$$\tau(t) = \tau_0, \quad 0 \leq t \leq N \tag{4a}$$

$$\tau(t) = \tau_0 \exp[-\mu(t-N)], \quad t > N \tag{4b}$$

$$f(t) = f_0 \tag{4c}$$

These parameterized functions are particularly useful in interpreting the effects of public health interventions. For instance, the transmission rate, $\tau(t)$, is particularly affected by a reduced circulation achieved through a general isolation or quarantine measure, while the fraction $f(t)$ of asymptomatic infectious that become reported, thus isolated, cases can be drastically increased by a massive testing measure with focused isolation. In the above relations, μ is the attenuation factor for the transmission rate, N is the time in days for application of the public health intervention to change transmission rate, μ_f is the argument of the $f(t)$ variation between the limits (f_0, f_{max}) e first time variable function has been previously considered, while the second one has been introduced in the present work, so as to allow for the examination of combined measures.

The cumulative number of reported cases at time t, $CR(t)$, which is the quantity offered by the actual available data, and the a priori unknown cumulative number of unreported cases, $CU(t)$, are given by:

$$CR(t) = \int_{t_0}^{t} \nu_1(s) I(s) ds \tag{5a}$$

$$CU(t) = \int_{t_0}^{t} \nu_2(s) I(s) ds \tag{5b}$$

The daily number of reported cases from the model, $DR(t)$, can be obtained by computing the solution of the following equation and with initial conditions:

$$\frac{dDR(t)}{dt} = \nu f(t) I(t) - DR(t) \tag{6a}$$

$$DR(t_0) = DR_0 \tag{6b}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3 Estimation problem

The estimation problem, also known inverse problems, is an interdisciplinary task involving experimental observations and numerical analysis in many fields of science and engineering. Specifically in epidemiological studies, a crucial step is the determination of the extension of diseases. In this regard, the effort is made to infer the values of unknown states and parameters characterizing the system under investigation. The results of such studies can support developing public health policy. Here, using the available data of infected individuals in Brazil, we aim at solving a joint estimation of states and parameters (namely τ_0, f_0 and μ for the SIRU model) in order to assess the COVID-19 spread and, hence, to be able to predict locally the future progression of the pandemic.

The estimation problem, also known as non-stationary inverse problems usually requires a system dynamic model, which provides the temporal evolution of the states and parameters, an observation model, which transforms the system variables to give the observed variables, and a estimation technique that relates the all available information to the process knowledge to obtain the desired estimate [15].

$$\boldsymbol{x}_{k+1} = \boldsymbol{F}(\boldsymbol{x}_k, \boldsymbol{\theta}_k) + \boldsymbol{\varepsilon}_k \tag{7a}$$

$$\boldsymbol{y}_k = \boldsymbol{G}(\boldsymbol{x}_k, \boldsymbol{\theta}_k) + \boldsymbol{\delta}_k \tag{7b}$$

Here, $\boldsymbol{x}$ represents the state vector, $\boldsymbol{\theta}$ is the parameter vector, $\boldsymbol{y}$ is the vector of the observed variables, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are respectively the system and observation noise vectors. The function $\boldsymbol{F}$ indicates the evolution model and the function $\boldsymbol{G}$ is a observation model, both of them are nonlinear function of state variables and parameters. It is assumed that the noise vectors have zero mean and follow a normal distribution. The subscript k denotes the instant of time.

There are many numerical methods that can approach the state and parameter estimation. However, as the model considered here is nonlinear and there may be high correlation between model parameters, local optimization methods such as gradient and direct search algorithms are not suitable since they strongly depend on the initial conditions. It is necessary to resort to the use of global methods capable of handling nonlinear systems, such as:

1. Heuristic methods, including the Particle Swarm Optimization (PSO) technique which usually presents better performance than other heuristic algorithms [16, 17] ; and

2. Sequential Monte Carlo methods, as the Particle Filter (PF) algorithms which makes use of random samples for a Bayesian statistical inference.

Such methods do not suffer from instability as they do not solve the inverse problem as an optimization problem, but as a search or sampling problem [18]. In both classes of methods, a set of particles is used to find the solution for the estimation problem, i.e filtering the available data and estimating the parameter values that minimize the objective function, F_{obj}, which represents the model performance for the problem. In this work, we solve the estimation problem for the SIRU model considering the measurements as the cumulative number of reported cases ($CR(t)$) from Brazil, using PSO and PF algorithms presented below.

Before carrying on the estimation problem, useful information can be obtained if the onset of the pandemic is modelled as an exponential growth, as represented by the following equation

$$\text{CR(t)}=\chi_1 exp(\chi_2 t) - \chi_3,\ t \geq t_0 \tag{8}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

After determining χ_1, χ_2 and χ_3, it is possible to explicitly calculate the unknown initial conditions for the model variables, estimates for the initial time of the pandemics and for the initial transmission rate [10, 9]:

$$I_0 = \frac{\chi_2 \chi_3}{f_0 \nu} \tag{9a}$$

$$U_0 = \frac{(1-f_0)\nu}{\eta + \chi_2} I_0 \tag{9b}$$

$$t_0 = \frac{1}{\chi_2} \ln\left(\frac{\chi_3}{\chi_1}\right) \tag{9c}$$

$$\tau_0 = \frac{\chi_2 + \nu}{S_0} \frac{\eta + \chi_2}{(1-f_0)\nu + \eta + \chi_2} \tag{9d}$$

3.1 Particle Swarm Optimization algorithm

This technique was proposed by Kennedy and Eberhart [17] and improved by other researchers to assure the convergence to the global optimal solution. It considers that each individual of the swarm, called particle, representing a potential solution moves around in the multidimensional search space. Balancing individuality and sociability of the particles, in order to locate the optimal solution, PSO mimics the social behavior of various species adjusting dynamically the position (ξ_j^k) and velocity (v_j^k) for each particle $j = 1, ..., M$ at each iteration k, in accordance with the following equations:

$$v_j^{k+1} = \omega v_j^k + \phi_p r_p (B_j - \xi_j^k) + \phi_g r_g (B_g - \xi_j^k) \tag{10a}$$

$$\xi_j^{k+1} = \xi_j^k + v_j^{k+1} \tag{10b}$$

In this formulation, j denotes the particle, k represents the iteration number, v is the velocity and ξ is the position of the particle. Regarding the regions of the search space where the objective function $F_{obj}(\xi_j)$ shows optimum values, B_j is the best value found by the particle itself and B_g is the best value found by the entire swarm. The coefficients r_p and r_g are uniform random numbers between 0 and 1; $0 < \omega < 1$ is related to inertial weight ; and finally ϕ_p and ϕ_g denote respectively individual cognition and the social parameters that must be chosen.

A general description of the PSO algorithm follows [17, 16]:

Algorithm 1: The Particle Swarm Optimization.

1: For each $j = 1, \cdots, M$, set the initial conditions for the swarm: the position (ξ_j^0) and velocity (v_j^0) are randomly generated, given suitable ranges for the variables.
2: Evaluate the objective function for each particle of the swarm, $F_{obj}(\xi_j)$, and update eventually the best value found by the particle itself, B_p, and the best value found by the entire swarm, B_g.
3: Update of the velocity of each particle of the swarm, v_j^{k+1}.
4: Update of the position of each particle of the swarm, in order to obtain the new position, ξ_j^{k+1}.
5: Check the stopping criteria; if it is not verified, return to step (2) for the next iteration.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3.2 Particle Filter algorithm

The estimation problem with PF methods is tackled in the form of a stochastic process, in which a set particles (or samples) with associated weights is to approximate the posterior density of the states and parameters. Within the Bayesian framework, all available information (measurements and phenomena involved) is used as a source of knowledge to determine the unknown variables. The prior distribution provides the necessary information to draw the first set of particles, which are updated through the evolution model and transformed by the observation model to give the estimates. The likelihood function is then used to compare such estimates with the experimental measurements, and incorporates more information, via particle weights, in order to determine the posterior distribution. Sequentially, all new information is combined with the previous information to create the basis for the statistical procedure. Such steps describe briefly the Sequential Importance Sampling (SIS) filter [19, 15, 20, 21].

The performance of the SIS filter may be negatively affected by the degeneracy phenomenon, in which few particles contribute to the approximation of the posterior probability distribution [19, 20]. In this regard, the Sampling Importance Resampling (SIR) and Auxiliary Sampling Importance Resampling (ASIR) filters consider a resampling step to deal with the elimination of particles originally with low weights and the replication of particles with high weights, at each instant t_k or whenever the number of effective particles falls below a certain threshold [22, 23, 24, 25, 26]. However, such filters are used to obtain estimates of state variables [19, 20].

In this work, we are particularly interested in the joint estimation of parameters and state variables, since estimating the parameters that regulate the dynamics of growth in the number of infected individuals is also very important to understand the phenomenon. For this purpose, we used the algorithm proposed by Liu and West [27]. This algorithm is a generalization of the ASIR filter, and the inference is made on the joint posterior density $\pi(\boldsymbol{x}_k, \boldsymbol{\theta}|\boldsymbol{D}_k)$,where $\boldsymbol{\theta}$ is the vector of parameters, $\boldsymbol{D}_k$ contains the measurements up to time k and $\boldsymbol{X}_k$ the state variables. Time evolves to $k+1$, we observe $\boldsymbol{y}_{k+1}$ and now want to generate a sample from $P(\boldsymbol{x}_{k+1}, \boldsymbol{\theta}|\boldsymbol{D}_{k+1})$. Bayes' theorem gives this as

$$P(\boldsymbol{x}_{k+1}, \boldsymbol{\theta}|\boldsymbol{D}_{k+1}) \propto P(y_{k+1}|\boldsymbol{x}_{k+1}, \boldsymbol{\theta})P(\boldsymbol{x}_{k+1}|\boldsymbol{\theta}, \boldsymbol{D}_k)P(\boldsymbol{\theta}|\boldsymbol{D}_k) \tag{11}$$

The filter is based on the hypothesis of [28], which assumes that, for a vector of static parameters $\boldsymbol{\theta}$, the posterior density estimation $P(\boldsymbol{\theta}|\boldsymbol{D}_k)$ is done by smoothed density via kernel, [27]

$$P(\boldsymbol{\theta}|\boldsymbol{D}_k) \approx \sum_{i=1}^{N} w_k^i N(\boldsymbol{\theta}|\boldsymbol{m}_k^i, h^2\boldsymbol{V}_k), \tag{12}$$

thus the artificial distribution of the parameters $\boldsymbol{\theta}$ is represented by the w_k^i weighted mixture of normal distributions with mean $\boldsymbol{m}_k$ and variance $\boldsymbol{V}_k$, where $N(\cdot|\boldsymbol{m}, \boldsymbol{S})$ is a multivariate normal density mean $\boldsymbol{m}$ and variance matrix $\boldsymbol{S}$, where,h is a smoothing parameter, be chosen as a smooth

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and decreasing function of N, e

$$\boldsymbol{m}_k = \boldsymbol{\theta}_k^i + (1-a)\bar{\boldsymbol{\theta}}_{\boldsymbol{k}} \tag{13}$$

$$a = \sqrt{(1-h^2)} \tag{14}$$

$$\bar{\boldsymbol{\theta}} = \frac{1}{N}\sum_i^k \frac{\boldsymbol{\theta}_k^i}{N} \tag{15}$$

$$\boldsymbol{V}_k = \frac{1}{N}\sum_i^k \left(\boldsymbol{\theta}_k^i - \bar{\boldsymbol{\theta}}\right)\left(\boldsymbol{\theta}_k^i - \bar{\boldsymbol{\theta}}\right)^T \tag{16}$$

where T denotes the transpose of the matrix,$\bar{\boldsymbol{\theta}}$ is the mean of the posterior particles. In addition, a is related to a discount factor δ as follows,[27]

$$a = \frac{3\delta - 1}{2\delta} \tag{17}$$

where $0.95 < \delta < 0.99$

The general particle filter algorithm for combined estimation of parameters and state variables in time step $k-1$, based on the measurements available in time k is given below, [27]

4 Numerical results and discussion

In all simulations presented here, we assumed the estimates $1/\nu = 6.20798$ and $1/\eta = 11.2784$ days [5] for the average times in the model, and also set $S_0 = 210.15 \cdot 10^6$ as the total of susceptible individuals in Brazil. Firstly, we used the data points related to the first month of the cumulative cases in Brazil, from February, 26 2020 up to March 25, 2020 (30 days), to fit the early exponential growth. In the sequence, we estimated the model parameters via inverse problem analysis applying data from February 26, 2020 up to May 21, 2020 (85 days).

Besides that, we also considered that isolation measures were taken from March 17, 2020, when domestic isolation, remote work and shutting down temporarily businesses and services were recommended. The data points recorded from May 8, 2020 until May 21, 2020 (14 days), were used to verify the predictive capability of the proposed approach. This way, we used about 84% of the data set to estimate the SIRU model parameters (learning step) and 16% for validation, changing the estimated values, if necessary, of f_0 or μ to improve the model performance (prediction step). Such computational studies for direct and inverse problems were performed in PYTHON on a computer with Intel Core i5 processor and 8 GB RAM.

4.1 Early exponential growth model

The adjustment for the early growth of the cumulative infectious cases in Brazil is displayed in Figure 3. We can see a satisfactory agreement between the data and the model, making clear that this phase of the pandemic is purely exponential. The estimated parameters for the model $\chi_1 = 4.185$, $\chi_2 = 0.231$, and $\chi_3 = 1.0$. Such result allowed obtaining an initial time for the growth

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Algorithm 2: The Liu and West Algorithm.

1: For each $j = 1, \cdots, N$, identify the prior point estimates of $(\boldsymbol{x}_k, \boldsymbol{\theta})$ given by $(\boldsymbol{\mu}_{k+1}^j, \boldsymbol{m}_k^j)$ where;

$$\boldsymbol{\mu}_{k+1}^j = E\left(\boldsymbol{x}_{k+1} | \boldsymbol{x}_k^j, \boldsymbol{\theta}_k^j\right)$$

may be computed from the state evolution density and $\boldsymbol{m}_k^j = a\boldsymbol{\theta}_k^j + (1-a)\bar{\boldsymbol{\theta}}_k$ is the jth kernel location.

2: Sample an auxiliary integer variable from the set $\{1, \cdots, N\}$ with probabilities proportional to

$$g_{k+1}^j \propto w_k^j P\left(\boldsymbol{y}_{k+1} | \boldsymbol{\mu}_{k+1}^j, \boldsymbol{m}_k^j\right);$$

call the sampled index z.

3: Sample a new parameter vector $\boldsymbol{\theta}_{k+1}^z$ from the zth normal component of the kernel density, namely

$$\boldsymbol{\theta}_{k+1}^z \sim N\left(\cdot | \boldsymbol{m}_k^z, h^2 \boldsymbol{V}_k\right).$$

4: Sample a value of the current state vector $\boldsymbol{x}_{k+1}^z$ from the system equation

$$P\left(\cdot | \boldsymbol{x}_k^z, \boldsymbol{\theta}_{k+1}^z\right).$$

5: Evaluate the corresponding weight

$$w_{k+1}^z \propto \frac{P\left(\boldsymbol{y}_{k+1} | \boldsymbol{x}_{k_1}^z, \boldsymbol{\theta}_{k+1}^z\right)}{P\left(\boldsymbol{y}_{k+1} | \boldsymbol{\mu}_{k+1}^z, \boldsymbol{m}_k^z\right)}.$$

6: Repeat step $(2) - (5)$ a large number of times to produce a final *posterior* approximation $\left(\boldsymbol{x}_{k+1}^z, \boldsymbol{\theta}_{k+1}^z\right)$ with weights w_{k+1}^z, as required.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

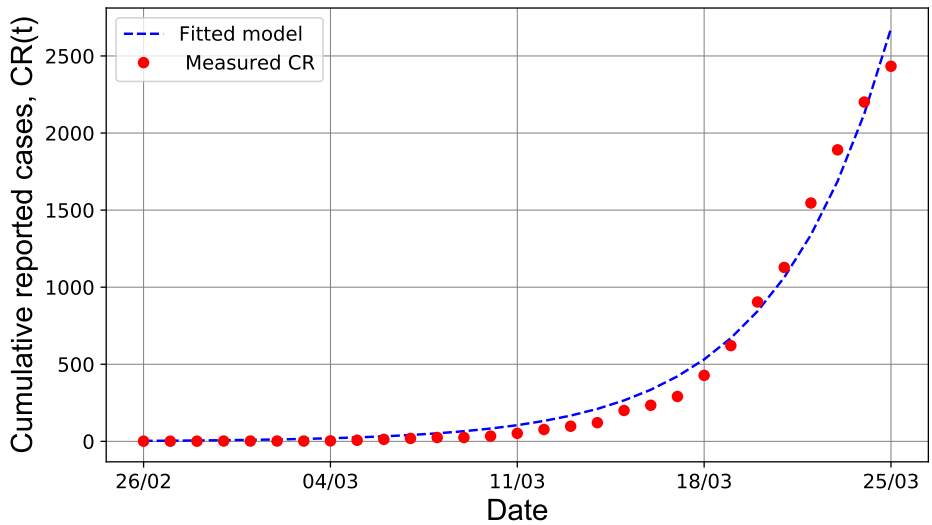

Figure 3: Early stage fitting for the cumulative reported cases, $CR(t) = \chi_1 exp(\chi_2 t) - \chi_3$, in Brazil from February, 26 2020 up to March 25, 2020, with $\chi_1 = 4.185$, $\chi_2 = 0.231$, and $\chi_3 = 1.0$.

at $t_0 = -6.203$. According to this model, a negative value for t_0 means that the starting time of the epidemic in Brazil took place about six days before the first reported case, what may have happened on February 20, 2020. Although such modelling cannot represent subsequent cases, we can use its parameters to find the remaining initial conditions, I_0, U_0 and τ_0, which have to be recalculated from within both PSO and PF algorithms.

4.2 Estimation from PSO simulations

To solve the estimation problem through the PSO technique, we applied a time weighted cost function as the objective function $F_{obj} = \int t \cdot (CR_{meas} - CR_{model})^2 dt$, where CR_{meas} and CR_{model} are the values of the measured and estimated cumulative cases at time t. We used 100 particles and 50 independent calls of the code. To tune the algorithm, we set the parameters $\omega = 0.9$ and $\phi_p = \phi_g = 2.0$[16].

Firstly, to monitor the number of reported and unreported cases and infectious cases, we used the full data set (85 days) for estimation of the SIRU model parameters, f_0, μ and $\tau(t) = g(t, \tau_0, \mu)$. The adjustment with PSO simulations led to the parameters $f_0 = 0.39123$, $\mu = 0.0109$ and $\tau_0 = 1.426 \cdot 10^{-9}$. In Figure 4, it is possible to note an excellent agreement of the simulated values with the measured values of the cumulative reported cases (Figure 4-a). However, the number of reported cases is not represented accurately, what can also be better observed in the Figure 4-b for the daily cases. In addition, the low value of the fraction f_0 indicates that about 40% of asymptomatic individuals became reported symptomatic infectious, meaning that a vast majority of the asymptomatic population was not diagnosed. However, the significant under reporting of cases

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

are caused by low number of tests being carried out, delay in the official processing of information and more seriously the lack of test kits. Such reasons make it difficult to fight Covid-19 in Brazil.

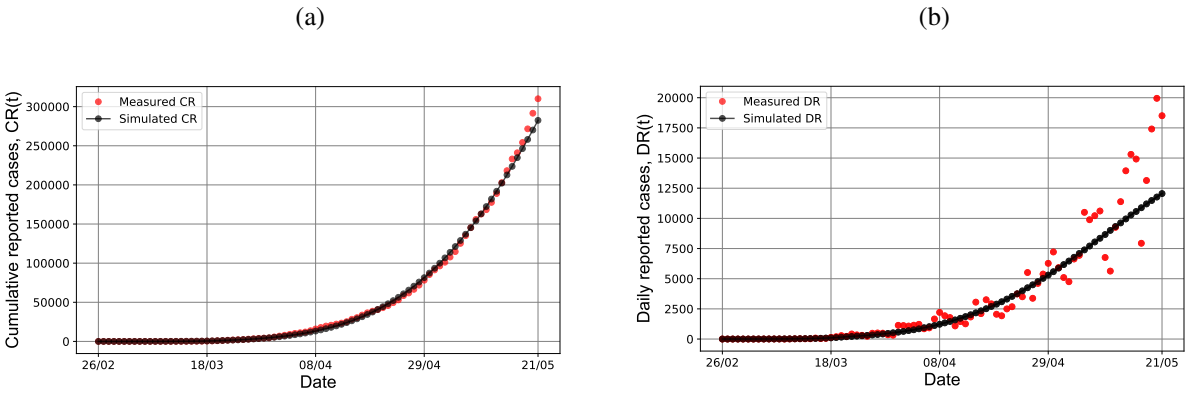

Figure 4: Monitoring results for cumulative (a) and daily (b) reported cases from PSO simulations, considering data from February, 26 2020 up to May 21, 2020 in Brazil, with $f_0 = 0.39123$, $\mu = 0.0109$ and $\tau_0 = 1.426 \cdot 10^{-9}$.

The result for the number of asymptomatic and symptomatic infectious cases given by the model states, $I(t)$, $R(t)$ and $U(t)$, which are not measured variables, is also obtained (Figure 5). Infections may be transmitted from either asymptomatic individuals or unreported symptomatic individuals, which sum up more than 350,000 at this stage (May, 21 2020) of the pandemic in the country. It is known that most healthy people are likely to experience asymptomatic or mild cases; however, this fact makes the coronavirus spread control even harder.

Finally, in order to verify the predictive capability of this approach, we divided the problem into the learning and prediction steps as explained above. The best result of the simulations provided $f_0 = 0.39123$, $\mu = 0.0109$ and $\tau_0 = 1.426 \cdot 10^{-9}$. In Figure 6, from May 8, 2020, the prediction for 14 days is compared to the measured (or observed) values of the reported cases. We can see again that the SIRU model can be adjusted to represent the Brazil data set, since there is an excellent agreement with the data used in the learning step. However, the prediction step shows a poor performance when one desires to use the estimated parameters in the learning step. This is probably because any pandemics is a dynamic process depending on many factors, such as individual behaviour against isolation enforcement, application rate of test kits and processing of information. Since we used a PSO algorithm that does not consider time-varying parameters and uncertainties, this approach failed to exhibit a different trend which was not present in the training data.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

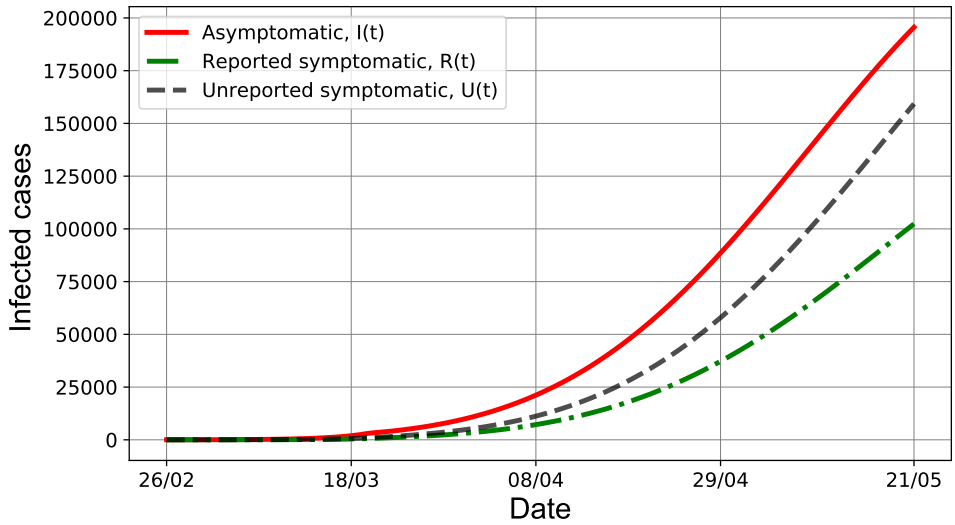

Figure 5: Monitoring results for reported and unreported cases from PSO simulations, considering data from February 26, 2020 up to May 21, 2020 in Brazil, for asymptomatic and symptomatic infectious cases, $I(t)$, $R(t)$ and $U(t)$, with $f_0 = 0.39123$, $\mu = 0.0109$ and $\tau_0 = 1.426 \cdot 10^{-9}$.

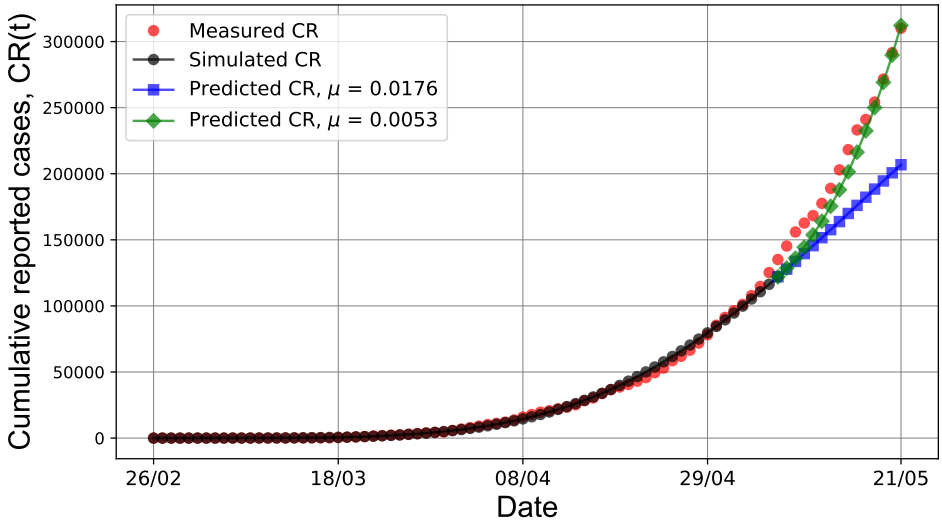

Figure 6: Simulation and prediction results from PSO simulations compared with the measured number of reported cases in Brazil from May 08, 2020 up to May 21, 2020. Different changes on were carried out to improve the performance of the prediction step.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The development of a reliable forecasting model is fundamental to allow public policies to be adopted in advance, in order to avoid further increasing the number of infected individuals. So, proposing a different scenario to predict properly may be interesting to show that the SIRU model can represent this pandemic. In this sense, we consider that the most uncertain parameter of the model would be the attenuation factor μ, which depends on the quarantine and isolation measures to decrease the infection rate. Keeping f_0 constant at the estimated value, we checked for different changes on μ to find a scenario in which the performance of the prediction phase step could be improved. According to the result in Figure 6, we found that reducing substantially the attenuation factor by 70% would suffice to improve the prediction for the reported cases. Such result highlights the importance of considering state and parameter uncertainties. In this regard, we expect that the particle filter allows reaching a better prediction performance, what would make forecast analysis more reliable (or, at least, provide better understanding of the dynamic behaviour of the parameters)or, at least, understanding the time variation and range of the parameters. Anyway, the best result of the PSO algorithm is used merely here for comparison with the PF simulations, and to provide an initial estimate of the model parameters.

4.3 Estimation from PF simulations

We used the algorithm of Liu & West for the particle filter technique in the estimation problem. Such algorithm is suitable for the joint estimation of state variables and model parameters. The number of particles N_{part} was set arbitrarily to 5000 and the results from the PSO simulations were considered as initial guesses for the parameters.

Gaussian prior probability densities were assumed for the parameters f_0 and μ with standard deviations corresponding to approximately 1% of the means of each parameter. For the effect of the model uncertainty, we tuned throughout the simulations 0.001% for the number of individuals susceptible to infection, $S(t)$, and 5% for the remaining states, $I(t)$, $R(t)$ and $U(t)$ – such percentage is related to the previous value of the respective state. Besides that, the uncertainty level for the observed cumulative cases, $CR(t)$, changes over time. It started at $\sigma_{CR} = 30\%$ in relation to the measured value and, as of March 17, 2020 when isolation was enforced, it was decreased in the simulation by a step-change to $\sigma_{CR} = 3\%$. To present the results, we calculated the estimation error, $RMSE = \sqrt{\sum_k (CR_{meas}(k) - CR_{model}(k))^2 / \sigma_{CR}^2(k)}$, and 95% credibility level for each state variable and parameter was calculated as the 2.5% and 97.5% quantiles .

The posterior distributions of each state and parameters estimated on May 07, 2020 (last day of the learning step) provided the respective initial condition for the prediction step. Lastly, we also performed an additional step to forecast new Covid-19 cases during 120 days (from May 22, 2020 to September 17, 2020). In the prediction and forecast horizons, the model parameters are updated via random walk model.

Using the full data set (February 26, 2020 to May 21, 2020), the estimation with the algorithm of Liu & West significantly improved the agreement of the simulated values with the measured values (Figure 7), especially for those measured points near the end of the considered period. This was possible due to the dynamic estimation of the parameters (Figure 8 a and b), which were varied over time from the initial guess. The attenuation factor was increased about 30% from the initial guess, showing that isolation has been effective at some extent decreasing the transmission rate,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$\tau(t)$. The fraction of report shows that over 60% of symptomatic infectious cases were not reported in this period. The improvement obtained in comparison to the PSO result can be seen from the estimation error *RMSE* (Figure 8d), which was consistently decreased. Regarding the infectious cases, the estimation is alarming since the group of potential transmitters is increasing (Figure 9). The number of the infectious population that may not have been diagnosed (that is, asymptomatic and unreported symptomatic individuals) is estimated to be more than 450,000 cases. Hence, it is worthy to point out that massive, rapid tests and restrictive measures are paramount to curb the spread of the virus.

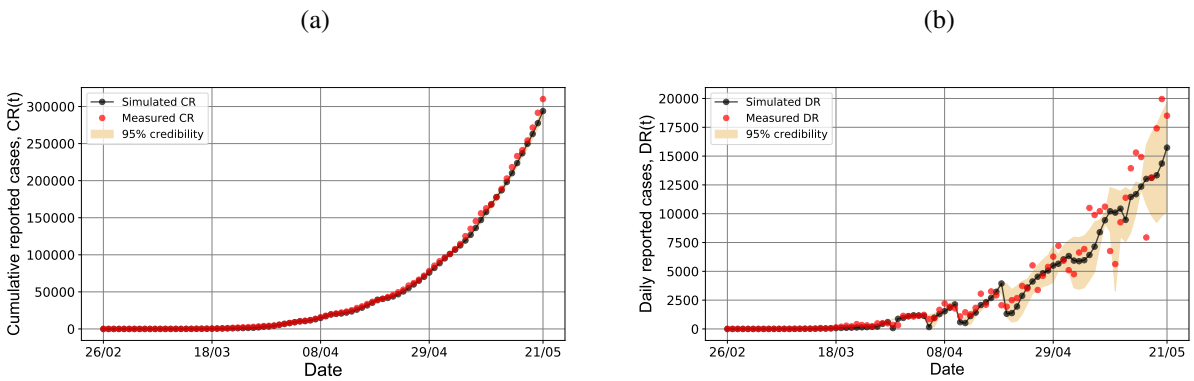

Figure 7: Monitoring results for (a) cumulative cases and (b) daily reported cases from the PF simulations, considering data from February 26, 2020 up to May 21, 2020 in Brazil.

The capability of the model to predict was analyzed using the available data of reported cases in Brazil from February 26, 2020 up to May 7, 2020. According to Figure 10, the adjustment of the parameters found in the learning step is suitable to predict accurately the pandemic evolution up to May 18, 2020 (10 days ahead), which would have presented in advance the notification of more than 115,000 cases.

From this result, we simulated the model considering 120 days from May 21, 2020 to forecast new cases. The dynamic behaviour of the parameters in both steps can be seen in Figure 11. The forecasted values for the daily reported cases and for the infectious cases are shown in Figures 12 and 13, in which we highlighted each analysis step (estimation, prediction and forecast). The result suggests that the epidemic peak in Brazil are due to be reached around mid-June, 2020,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

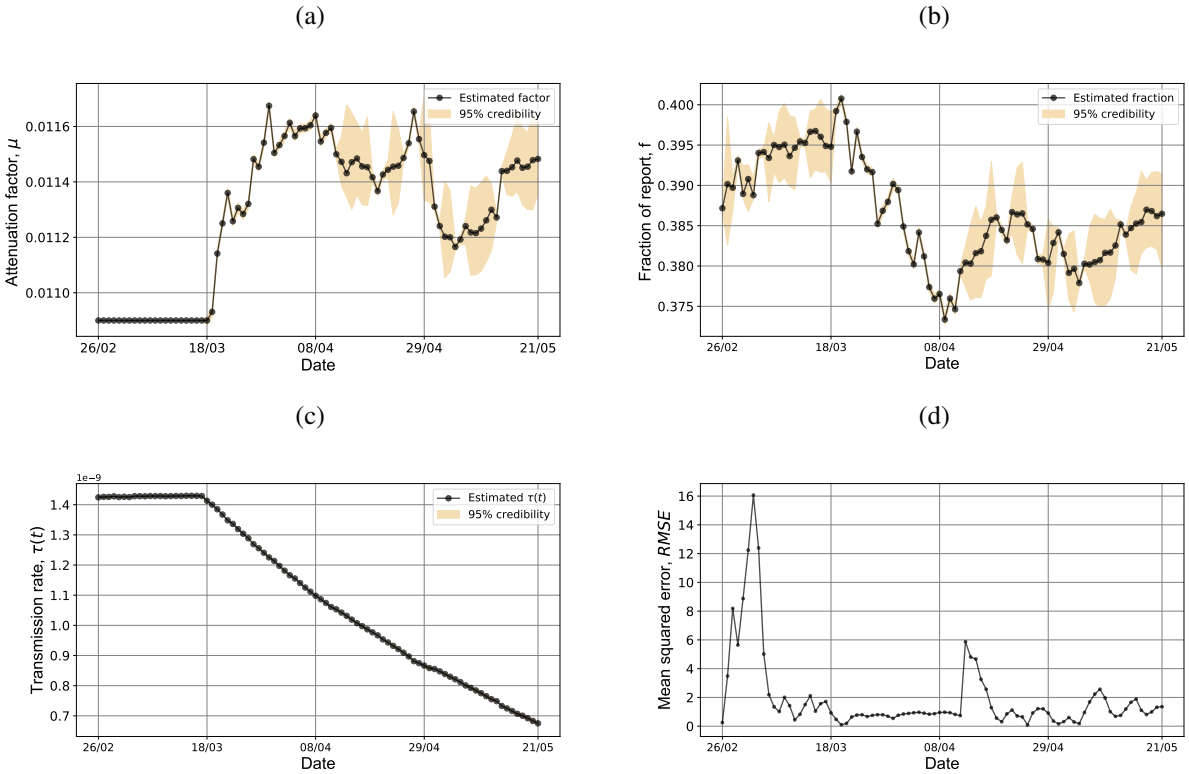

Figure 8: Dynamic results of the SIRU model parameters: (a) attenuation factor, μ ; (b) fraction of report, $f(t) = f_0$; (c) transmission rate, $\tau(t) = \tau_0 exp[-\mu(t-N)]$; and (d) the mean squared error, RMSE, from the PF simulations, considering data from February 26, 2020 up to May 21, 2020 in Brazil.

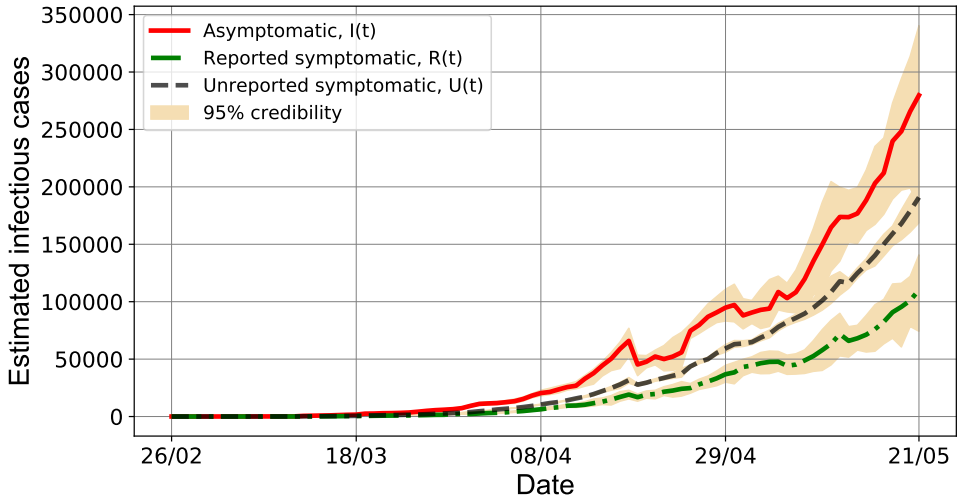

Figure 9: Monitoring results for reported and unreported cases from the PF simulations, considering data from February 26, 2020 up to May 21, 2020 in Brazil, for asymptomatic and symptomatic infectious cases, $I(t)$, $R(t)$ and $U(t)$.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

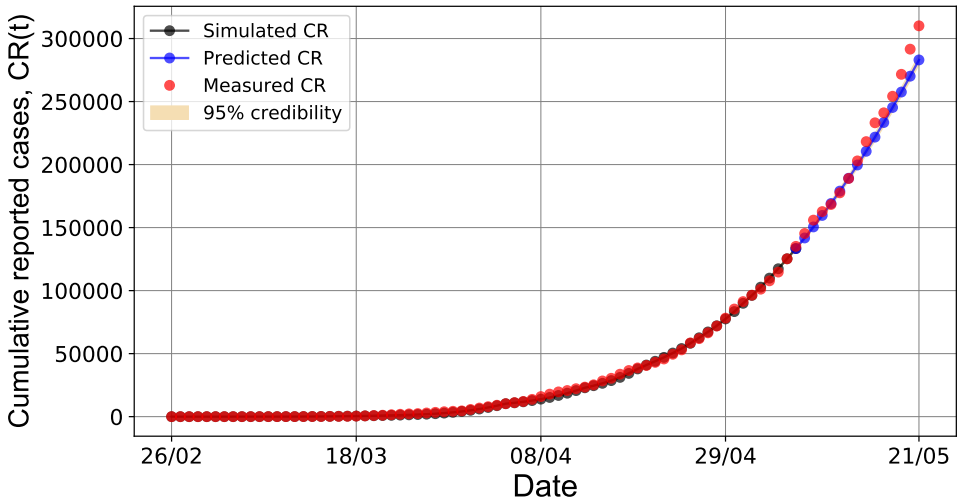

Figure 10: Validation of the predictive capability of the PF simulations for the measured number of reported cases in Brazil, considering data from May 08, 2020 up to May 21, 2020.

with almost 25,000 cases being notified a day. Specifically, it can inferred that the number of asymptomatic individuals may increase up to 400,000 cases and, more importantly, the number of unreported symptomatic cases around 460,000 may be twice greater than the number of reported symptomatic. In brief, this means that number of people requiring medical assistance will increase dramatically in the future. According to the forecast, the pandemics in Brazil is expected to slow down beyond the peak of daily number of cases and the number of cumulative reported cases to be more than 1,500,000 (Figure 14). So, if such numbers impose a burden on the health system, it is important to enhance isolation measures, as quarantine and social distancing. As Brazil is a very large country, the spread of the virus is supposed to vary from region to region, such way that community containment measures may be established locally at different moments.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

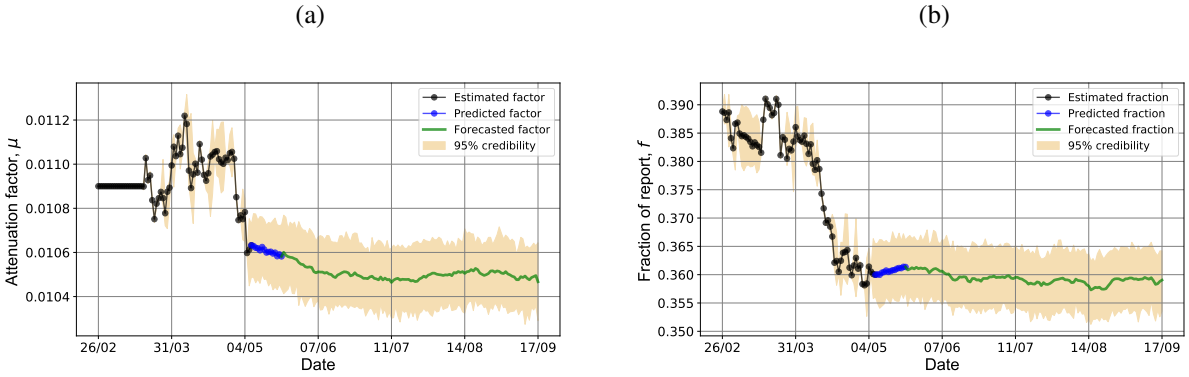

Figure 11: Dynamic behaviour of the parameters during the prediction and forecasting steps: (a) attenuation factor, μ ; (b) fraction of report, $f(t) = f_0$.

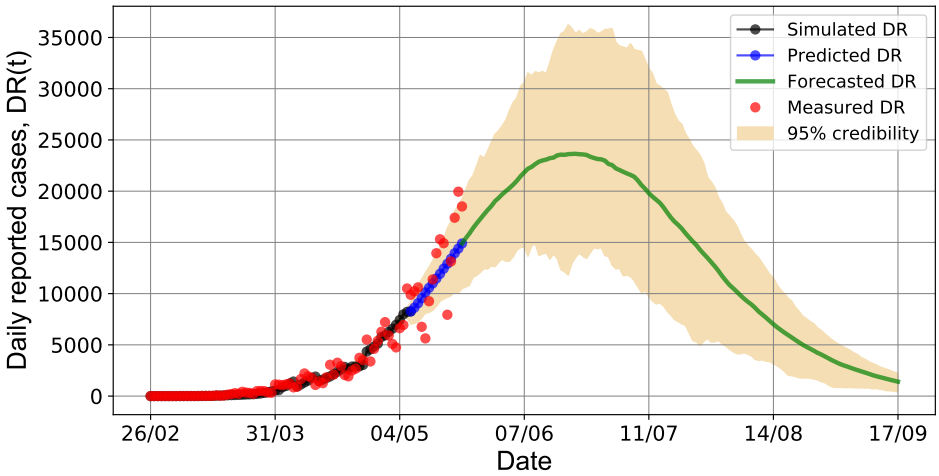

Figure 12: Forecast result for the daily reported cases in Brazil up to September 17, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

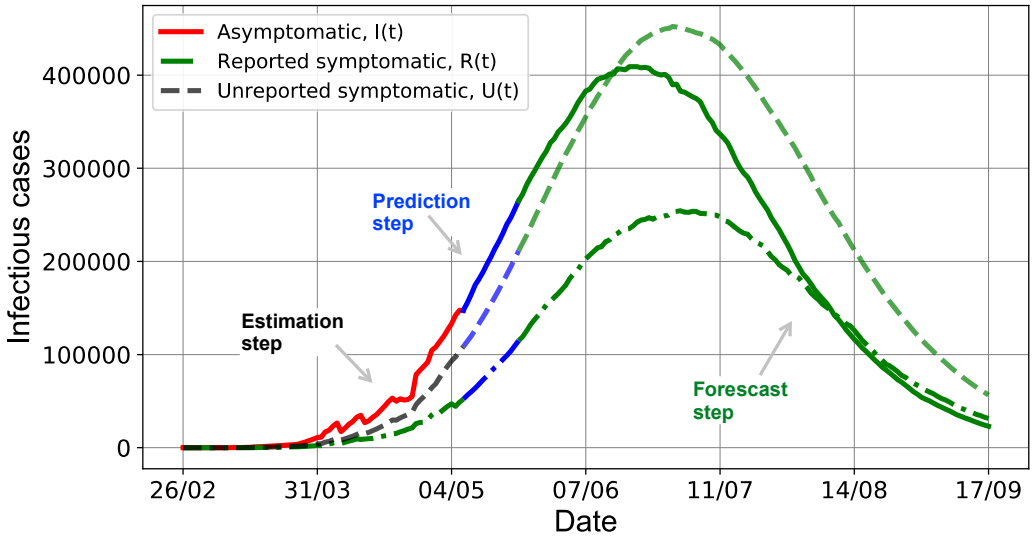

Figure 13: Forecast result for the number of infectious cases up to September 17, 2020.

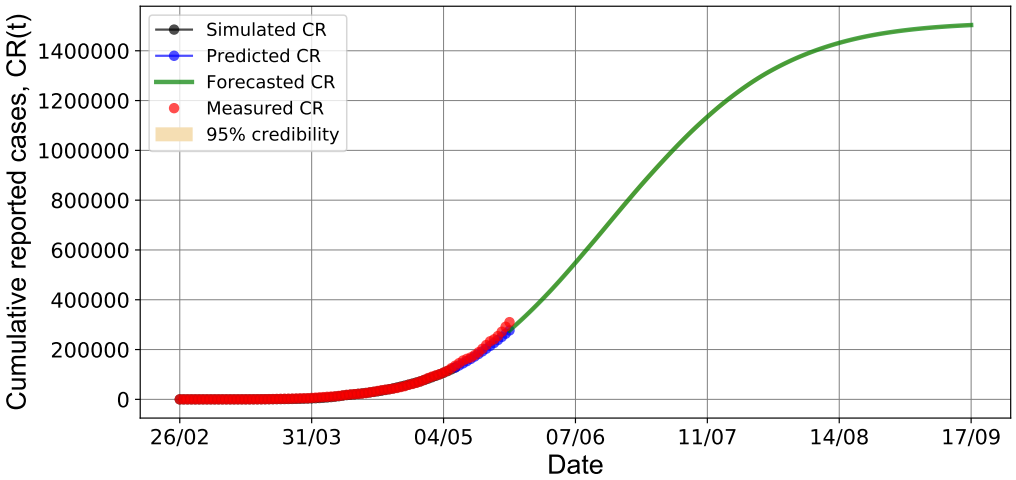

Figure 14: Forecast result for the cumulative reported cases up to September 17, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

5 Conclusions

Since the coronavirus outbreak has been evolving rapidly, this work considered a numerical application to predict the evolution of a Covid-19 epidemic in Brazil. We implemented the model SIRU to represent the coronavirus spread and used the cumulative reported cases as measurement. For the estimation problem, an inverse problem was solved under the Bayesian framework with particle filter. Specifically, we used the algorithm of Liu & West to perform a joint estimation of states and parameters. Besides that, particle swarm optimization based simulations provided initial guesses for this problem. The model was calibrated to represent the number of reported cases and validated using two-week real data. Afterwards, we performed a forecast step to predict new cases of the disease. It was found that the pandemic peak is expected to take place in mid-June, 2020 with about 25,000 news cases a day and over 1,500,000 individual will have tested positive for Covid-19 up to September, 2020. Such result reinforces the need for social distancing measures and for diagnostic tests to decrease the transmission rate and the number of infected people in severe condition.

References

[1] P. Zhou, X. Yang, X-G Wang, B. Hu, L. Zhang, W. Zhang, H-R Si, Y. Zhu, B. Li, C-L. Huang, H-D. Chen, J. Chen, Y. Luo, H. Guo, R-D. Jiang, M-Q. Liu, Yi. Chen, X-R. Shen, and X. Wang. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 03 2020.

[2] Y. Wang, M-H. Hu, Q. Li, X-P. Zhang, G. Zhai, and N. Yao. Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with covid-19 in an accurate and unobtrusive manner. 02 2020.

[3] https://covid19.who.int/region/amro/country/br. Acessado: 25 mai. 2020.

[4] P. L. Menezes, D. M. Garner, and V. E. Valenti. Brazil is projected to be the next global covid-19 pandemic epicenter. 05 2020.

[5] R. M. Cotta, C. P. , and P. Magal. Parametric identification and public health measures influence on the covid-19 epidemic evolution in brazil. *Biology*, (Special Issue "Theories and Models on COVID-19 Epidemics"), 2020.

[6] M. A. Savi. Chaos and order in biomedical rhythms", journal of the brazilian society of mechanical sciences and engineering. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 27:157–169, 2005.

[7] B. Tang, X. Wang, Q. Li, N. Bragazzi, S. Tang, Y. Xiao, and J. Wu. Estimation of the transmission risk of the 2019-ncov and its implication for public health interventions. *Journal of Clinical Medicine*, 01 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[8] R. M. Anderson and R. M. May. *Infectious Diseases of Humans*. Oxford University Press, Oxford, 1991.

[9] Z. Liu, P. Magal, O. Seydi, and G. Webb. Understanding unreported cases in the covid-19 epidemic outbreak in wuhan, china, and the importance of major public health interventions. *Biology*, 9:50, 03 2020.

[10] Z. Liu, P. Magal, O. Seydi, and G. Webb. Predicting the cumulative number of cases for the covid-19 epidemic in china from early data. *medRxiv*, (submitted), 2020.

[11] e Magal, P. and G. Webb. Predicting the number of reported and unreported cases for the covid-19 epidemic in south korea, italy, france and germany. 03 2020.

[12] Z. Liu, P. Magal, O. Seydi, and G. Webb. A covid-19 epidemic model with latency period. *Infectious Disease Modelling*, 04 2020.

[13] P. Magal and G. Webb. The parameter identification problem for sir epidemic models: identifying unreported cases. *Journal of Mathematical Biology*, 77:1–20, 01 2018.

[14] A. Ducrot, P. Magal, H. T. Nguyen, and G Webb. Identifying the number of unreported cases in sir epidemic models. *Mathematical medicine and biology : a journal of the IMA*, 07 2019.

[15] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer-Verlag, Berlin, 2004.

[16] Marcio Schwaab, Evaristo Chalbaud Biscaia Jr., José Luiz Monteiro, and José Carlos Pinto. Nonlinear parameter estimation through particle swarm optimization. *Chemical Engineering Science*, 63(6):1542 – 1552, 2008.

[17] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995.

[18] José Pallero, María Fernández-Muñiz, Ana Cernea, Óscar Álvarez-Machancoses, Luis Pedruelo-González, Sylvain Bonvalot, and Juan Fernández-Martínez. Particle Swarm Optimization and Uncertainty Assessment in Inverse Problems. *Entropy*, 20(2):96, jan 2018.

[19] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188, 2001.

[20] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter*. Artech House, Boston, 2004.

[21] N. Gordon, D. Salmond, and A. F. M. Smith. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. *Institution of Electrical Engineers*, 140:107–113, 1993.

[22] W. B. Silva, J. C. S. Dutra, J. M. J. Costa, L. A. S. Abreu, D. C. Knupp, and A. J. Silva Neto. A hybrid estimation scheme based on the sequential importance resampling particle filter and the particle swarm optimization (PSO-SIR). In *Computational Intelligence, Optimization and*

medRxiv preprint doi: https://doi.org/10.1101/2020.05.27.20115212; this version posted May 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Inverse Problems with Applications in Engineering, pages 247–261. Springer International Publishing, 2019.

[23] W. B. Silva, M. C. Rochoux, H. R. B. Orlande, M. J. Colaço, O. Fudym, M. El-Hafi, B. Cuenot, and S. Ricci. Application of particle filters to regional-scale wildfire spread. *High Temperatures High Pressures*, 43:415–440, 2014.

[24] L. S. F. Salardani, L. P. Albuquerque, J. M. J. Costa, W. B. Da Silva, and J. C. S. Dutra. Particle filter-based monitoring scheme for simulated bio-ethylene production process. *Inverse Problems in Science and Engineering*, 27:648–668, 2018.

[25] A. C. S. R. Dias, W. B. Da Silva, and J. C. S. Dutra. Propylene polymerization reactor control and estimation using a particle filter and neural network. *Macromolecular Reaction Engineering*, 11:1–20, 2017.

[26] R. Marques, . Silva, W.B, R. Hoffmann, J. S. Dutra, and F. Coral. Sequential state inference of engineering systems through the particle move-reweighting algorithm. *Computational and Applied Mathematics*, 37:220–236, 2017.

[27] J. Liu and M. West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer, 2001.

[28] Mike West. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):409–422, 1993.