# Modelling the COVID-19 epidemics in Brasil: Parametric identification and public health measures influence

**R.M. Cotta[1, 2], C.P. Naveira-Cotta[2], and P. Magal[3]**

[1] General Directorate of Nuclear and Technological Development, DGDNTM, Brazilian Navy, Rio de Janeiro, RJ, Brazil

[2] Laboratory of Nano & Microfluidics and Microsystems, LabMEMS, Mechanical Engineering Dept., POLI & COPPE, UFRJ, Federal University of Rio de Janeiro, Brazil

[3] Institut de Mathématiques de Bordeaux, Université de Bordeaux, 351 cours de la libération 33400 Talence, France

[§]Correspondence author. Email: cotta@mecanica.coppe.ufrj.br

## ABSTRACT

A SIRU-type epidemic model is proposed for the prediction of COVID-19 spreading within Brasil, and analyse the influence of public health measures on simulating the control of this infectious disease. Since the reported cases are typically only a fraction of the total number of the symptomatic infectious individuals, the model accounts for both reported and unreported cases. Also, the model allows for the time variation of both the transmission rate and the fraction of asymptomatic infectious that become reported symptomatic individuals, so as to reflect public health interventions, towards its control, along the course of the epidemic evolution. An analytical exponential behaviour for the accumulated reported cases evolution is assumed at the onset of the epidemy, for explicitly estimating initial conditions, while a Bayesian inference approach is adopted for parametric estimations employing the present direct problem model with the data from the known portion of the epidemics evolution, represented by the time series for the reported cases of infected individuals. The direct-inverse problem analysis is then employed with the actual data from China, with the first half been employed for the parametric estimation and the second half for validation of the predictive capability of the proposed approach. The full dataset for China is then employed in another parameter identification, aimed at refining the values for the average times that asymptomatic infectious individuals and that symptomatic individuals remain infectious. Following this validation, the available data on reported cases in Brasil from February 15$^{th}$ till March 29$^{th}$, 2020, is used for estimating parameters and then predict the epidemy evolution under these conditions. Finally, public health interventions are simulated, aimed at diminishing the effects of the disease spreading, by acting on both the transmission rate and the fraction of the total number of the symptomatic infectious individuals, considering time variable exponential behaviours for these two parameters, usually assumed constant in epidemic evolutions without intervention. It is demonstrated that a combination of actions to affect both parameters can have a much faster and effective result in the control of the epidemy dynamics.

## KEYWORDS

Epidemics modelling, SIRU model, Bayesian Inference, MCMC, COVID-19

## INTRODUCTION

A new human coronavirus started spreading in Wuhan, China, by the end of 2019, and turned into a pandemic disease called COVID-19 as declared by the World Health Organization on March 11th, 2020. The affected countries and cities around the world have been reacting in different ways, towards locally controlling the disease evolution. These measures 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. Naturally, China offers the longest time series on reported infected cases and the resulting effects of combining different public health interventions. As of March 26th, 2020, there were no reports in China of further internal contaminations, and all the new cases are associated with infected individuals that (re)entered in the country. Despite the apparent success of the interventions in China, each region or country might require a specific combination of measures, due to demographic spatial distribution and age structure, health system capabilities, and social-economical characteristics. In this sense, it urges to have a mathematical model that would allow for the simulation of such possible interventions on the epidemic evolution within the following few weeks or months. This article presents a collaborative research effort towards the construction of an epidemic evolution prediction tool, which combines direct and inverse problem analysis and is both reliable and easy to implement and execute, initially motivated by offering some insight into the control of COVID-19 within Brasil.

The classical susceptible-infectious-recovered (SIR) model describes the transmission of diseases between susceptible and infective individuals and provides the basic framework for almost all later epidemic models. At the onset of the coronavirus epidemy in China, there were some initial studies for the prediction of its evolution and the analysis of the impact of public health measures [1], which however did not consider in the modelling the presence of unreported infection cases, which are in practice inherent to this process. The present work is first based on the SIR-type model proposed in [2], which deals with the epidemic outbreak in Wuhan by introducing the unreported cases in the modelling, and evaluating the consequences of public health interventions. It was a direct application of previous developments [3,4] on the fundamental problem of parameter identification in mathematical epidemic models, accounting for unreported cases. This same modelling approach was more recently employed in the analysis of the epidemic outbreak in different countries, including China, South Korea, Germany, Italy,

and France [5-7]. Besides identifying unreported cases, this simple and robust model also allows for introducing a latency period and a time variable transmission rate, which can simulate a public health orientation change such as in a general isolation measure. In addition, an analytical exponential behaviour is assumed for the accumulated reported cases evolution along a second phase just following the onset of the epidemy, which, upon fitting of the available data, allows for the explicit analytical estimation of the transmission rate and the associated initial conditions required by the model.

Here, the SIR-type model in [2-7] is implemented for the direct problem formulation of the COVID-19 epidemic evolution, adding a time variable parametrization for the fraction of asymptomatic infectious that become reported symptomatic individuals, a very important parameter in the public health measure associated with massive testing and consequent focused isolation. The same analytical identification procedure is maintained for the required initial conditions, as obtained from the early stages exponential behaviour. However, a Bayesian inference approach is here adopted for parametric estimation, employing the Markov Chain Monte Carlo method with the Metropolis-Hastings sampling algorithm [8-12]. At first, the goal of the inverse problem analysis was estimating the parameters associated with the transmission rate and the fraction of asymptomatic infectious that become reported symptomatic individuals, which can be quite different in the various regions and countries and also very according to the public health measures. Then, in light of the success in this parametric identification, an extended estimation was also employed which incorporates the average time the asymptomatic infectious are asymptomatic and the average time the infectious stay in the symptomatic condition, due to the relative uncertainty on these parameters in the literature. The proposed approach was then applied to the data from China, first by taking just the first half of these data points in the estimation, while using the second half to validate the model using the estimated parameters with just the first half of the epidemy evolution, and second by employing the whole time series in the MCMC estimation procedure, thus identifying parameters for the whole evolution period. This second estimation was particularly aimed at refining the data for the average times that asymptomatic infectious individuals and that symptomatic individuals remain infectious. Upon validation of the approach through the data for China, we have proceeded to the analysis of the epidemic dynamics in Brasil, after about 35 days of collected information on reported infected individuals. First, the available data was employed in the parametric estimation, followed by the prediction of the epidemy evolution in Brasil. Then, we have

explored the time variation of both the transmission rate and the fraction of asymptomatic infectious that become reported symptomatic individuals, so as to reflect public health interventions, in simulating possible government measures, as described in what follows.

## DIRECT PROBLEM

The implemented SIR-type model [2-7] is given by the following initial value problem:

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

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

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

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

where,

$$\nu_1(t) = \nu f(t); \quad \nu_2(t) = \nu(1 - f(t)) \tag{2.a,b}$$

with initial conditions

$$S(t_0) = S_0; \quad I(t_0) = I_0; \quad R(t_0) = 0; \quad U(t_0) = U_0; \tag{3.a-d}$$

Here, $t_0$ is the beginning date of the epidemic in days, S(t) is the number of individuals susceptible to infection at time t, I(t) is the number of asymptomatic infectious individuals at time t, R(t) is the number of reported symptomatic infectious individuals (i.e., symptomatic infectious with severe symptoms) at time t, and U(t) is the number of unreported symptomatic infectious individuals (i.e., symptomatic infectious with mild symptoms) at time t. Asymptomatic infectious individuals I(t) are infectious for an average period of 1/ν days. Reported symptomatic individuals R(t) are infectious for an average period of 1/η days, as are unreported symptomatic individuals U(t). We assume 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 $\nu_1 = f\,\nu$, the rate asymptomatic infectious become unreported symptomatic is $\nu_2 = (1-f)\,\nu$, 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 1 below illustrates the infection process as a flow chart.

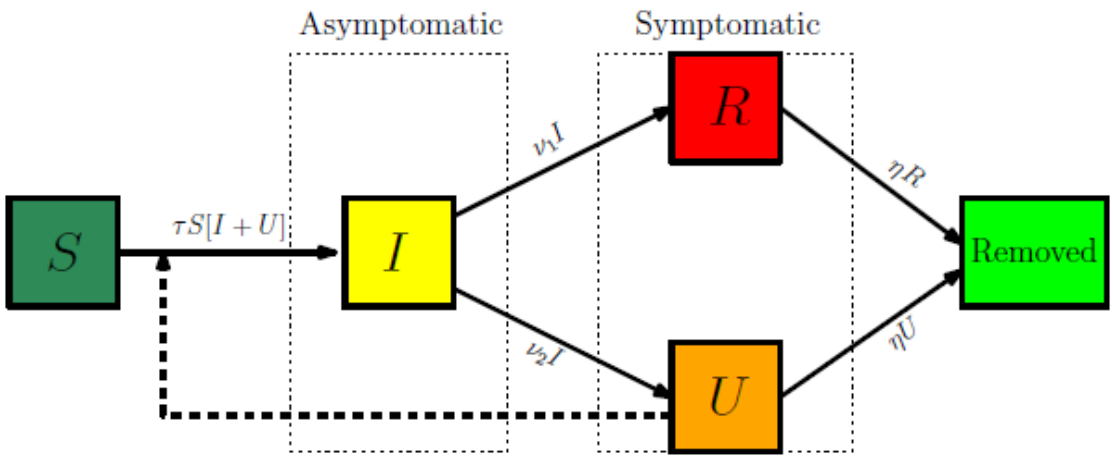



**Figure 1 – Flow chart illustrating the infection path process [3].**

The time variable coefficients, $\tau(t)$ and $f(t)$, are given by:

$$\tau(t) = \tau_0\,, \qquad 0 \le t \le N \tag{4.a}$$

$$\tau(t) = \tau_0 \exp\left(-\mu(t-N)\right), \qquad t > N \tag{4.b}$$

$$f(t) = f_0\,, \qquad 0 \le t \le N_f \tag{4.c}$$

$$f(t) = (f_{max} - f_0)\left[1 - \exp\left(-\mu_f\left(t - N_f\right)\right)\right] + f_0, \qquad t > N_f \tag{4.d}$$

These parametrized 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, $\mu$ 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, $\mu_f$ is the argument of the $f(t)$ variation between the limits $(f_0, f_{max})$. The first time variable function has been previously considered, while the second one has been introduced in the present work, so as to allow 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:

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

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

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

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

with initial conditions

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

## INVERSE PROBLEM

Inverse problem analysis is nowadays a common practice in various science and engineering contexts, in which the groups involved with experimental data and numerical simulation synergistically collaborate so as to obtain the maximum information from the available data, towards the best possible use of the modelling for the problem under study. Here, as mentioned in the introduction, we first review an analytical parametric identification described in more details in [4-7], that from the initial phases of the epidemic evolution allows to explicitly obtain the unknown initial conditions of the model, while offering a reliable estimate for the transmission rate at the onset of the epidemy. Nevertheless, even after these estimates, a few other parameters in the model remain uncertain, either due to the specific characteristics of the physical conditions or reaction to the epidemy in each specific region, or due to lack of epidemiological information on the disease itself. Therefore, an inverse problem analysis was undertaken aimed at estimating the main parameters involved in the model, as summarized in Table 1 below. First, for the dataset on the accumulated reported cases for China, the focus is on the parametrized time variation of the transmission rate ($\tau_0$ and $\mu$) and the fraction of asymptomatic infectious that become reported ($f_0$), in this case assumed constant, followed by an effort to refine the information on the average times (1/ν and 1/η) through

a simultaneous estimation of the five parameters. Then, employing the dataset for Brasil, the parametrized time variation of the transmission rate ($\tau_0$ and $\mu$) and the fraction of asymptomatic infectious that become reported ($f_0$), initially assumed constant, are estimated. In addition, due to the behaviour of the estimated CR(t) curve in this case, it is also attempted to estimate a possible time variation for the fraction of asymptomatic infectious that become reported, $f(t)$, by parametrization of an abrupt variation that requires just the estimation of $f_{max}$ and $N_f$.

Table 1 –Parameter estimates on each inverse problem analysis.

| **Country Data** | **Parameter under estimation** | **Data Range used in the estimation** |
|---|---|---|
| China | $f_0, \mu, \tau_0$ | January 19th up to February 17th |
| China | $f_0, \mu, \tau_0, 1/\nu, 1/\eta$ | January 19th up to March 25th |
| Brasil | $f_0, \mu, \tau_0$ | February 25th to March 29th |
| Brasil | $f_0, \mu, \tau_0, f_{max}, N_f$ | February 25th to March 29th |

The statistical inversion approach here implemented falls within the Bayesian statistical framework [8-12], in which (probability distribution) models for the measurements and the unknowns are constructed separately and explicitly, as shall be briefly reviewed in what follows.

As explained in previous works employing this model [4-7], it is assumed that in the early phase of the epidemic, the cumulative number of reported cases grows approximately exponentially, according to the following functional form:

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3 , \quad t \geq t_0 \tag{7.a}$$

After fitting this function to the early stages of the epidemic evolution, one may extract the information on the unknown initial conditions, in the form [4-7]:

$$t_0 = \frac{1}{\chi_2} [\ln (\chi_3) - \ln (\chi_1) \tag{7.b}$$

$$I_0 = \frac{\chi_3 \chi_2}{f_0 \nu} \tag{7.c}$$

$$U_0 = \frac{(1 - f_0)\nu}{\eta + \chi_2} I_0 \tag{7.d}$$

In addition, an excellent estimate for the initial transmission rate can be obtained from the same fitted function, in the form:

$$\tau_0 = \frac{\chi_2 + \nu}{S_0} \frac{\eta + \chi_2}{(1 - f_0)\nu + \eta + \chi_2} \tag{7.e}$$

Also, the the basic reproductive number for this initial phase model is estimated as:

$$\mathcal{R}_0 = \frac{\tau_0 S_0}{\nu} [1 + \frac{(1 - f_0)\nu}{\eta}] \tag{7.f}$$

The statistical approach for the solution of inverse problems here adopted employs the Metropolis-Hastings algorithm for the implementation of the Markov chain Monte Carlo (MCMC) method [8-9]. The MCMC method is used in conjunction with the numerical solution of the ordinary differential system, eqs.(1-3), for estimating the remaining model parameters. Consider the vector of parameters appearing in the physical model formulation as:

$$\mathbf{P}^T \equiv [P_1, P_2, ..., P_M] \tag{8}$$

where $M$ is the number of parameters. For estimating **P**, we assume that a vector of measured data is available (**Y**) containing the measurements $Y_i$ at time $t_i$, $i = 1, \ldots, I$. Bayes' theorem can then be stated as [8-9]:

$$\pi_{posterior}(\mathbf{P}) = \pi(\mathbf{P}|\mathbf{Y}) = \frac{\pi_{prior}(\mathbf{P})\pi(\mathbf{Y}|\mathbf{P})}{\pi(\mathbf{Y})} \tag{9}$$

where $\pi_{posterior}(\mathbf{P})$ is the posterior probability density, that is, the conditional probability of the parameters **P** given the measurements **Y**, $\pi_{prior}(\mathbf{P})$ is the prior density, that is, the coded information about the parameters prior to the measurements, $\pi(\mathbf{Y}|\mathbf{P})$ is the likelihood function, which expresses the likelihood of different measurement outcomes

**Y** with **P** given, and $\pi(\mathbf{Y})$ is the marginal probability density of the measurements, which plays the role of a normalizing constant. If different *prior* probability densities are assumed for the parameters, the posterior probability distribution may not allow an analytical treatment. In this case, Markov chain Monte Carlo (MCMC) methods are used to draw samples of all possible parameters, and thus inference on the posterior probability becomes inference on the samples [8-9]. The main merit of the MCMC method is about providing a picture of the posterior distribution, without solving the mathematical integrals in Bayes' rule. The idea is to approximate the posterior distribution by a large collection of samples of values. This method is especially suitable when it is unfeasible to yield an analytical solvable posterior distribution and/or a large space of parameters is involved, allowing one to do Bayesian inference even in rich and complex models. The idea behind the Metropolis-Hasting sampling algorithm is illustrated below, and these steps should be repeat until it is judged that a sufficiently representative sample has been generated.

1) Start the chain with an initial value, that usually comes from any prior information that you may have;

2) Randomly generate a proposed jump aiming that the chain will move around and efficiently explores the region of the parameter space. The proposal distribution can take on many different forms, in this work a Gaussian random walk was employed, implying that the proposed jumps will usually be near the current one;

3) Compute the probability of moving from the current value to the proposed one. Candidates moving to regions of higher probability will be definitely accepted. Candidates in regions of lower probability can be accepted only probabilistically. If the proposed jump is rejected, the current value is tally again. For more details on theoretical aspects of the Metropolis-Hastings algorithm and MCMC methods and its application, the reader should refer to [8-12].

## RESULTS AND DISCUSSION

### Model Validation: China

Before proceeding to the analysis of the COVID-19 epidemic evolution within Brasil, the major concern in the present contribution, the need was felt in validating the proposed direct-inverse problem analysis approach. In this sense, due to the largest available dataset

on this pandemic, we have chosen to use the information from China in terms of the accumulated confirmed infectious cases. The data for China was extracted from [6], complemented by the most recent data from [13] up to March 25th, 2020. The exponential fit for the early phase of the China CR(t) dataset provided the estimates of the three parameters, $\chi_1 = 0.14936, \chi_2 = 0.37680, \chi_3 = 1.0$, from which we have estimated $t_0 = 5.046$. The remaining data for the initial conditions, $I_0$ and $U_0$, and the early stage transmission rate, $\tau_0$, are in fact recalculated from within the MCMC algorithm, since the changing values of *f* will affect them, according to eqs. (7.c-e). The average times in the model were taken as 1/ν=7 and 1/η=7 days and the isolation measures were taken at N=29 days [6]. First, experimental data from China from the period of January 19th up to February 17th was employed in demonstrating the estimation of three parameters, $f_0, \mu,$ and $\tau_0$, assuming there is no significant time variation in the function *f*(*t*) ($\mu_f = 0$). In the absence of more informative priors, uniform distributions were employed for all three parameters under estimation. Table 1 presents the prior information and the initial guesses for the parameters. If the initial guesses were used to predict the CR(t) behavior, an over-estimation of the accumulated reported infected individuals would occur, especially in the long term, as can be noticed in Figure 1, confirming the need for a proper parameter estimation.

Table 1 – Prior distributions and initial guesses for the parameters to be estimated $f_0, \mu,$ and $\tau_0$ (China).

| **Parameter** | **Prior distribution** | **Initial Guess** |
|---|---|---|
| $f_0$ | $U[0,1]$ | 0.5 |
| $\mu$ | $U[0,5]$ | 0.1 |
| $\tau_0$ | $U[0, 1 \times 10^{-6}]$ | $4.478 \times 10^{-8}$ |

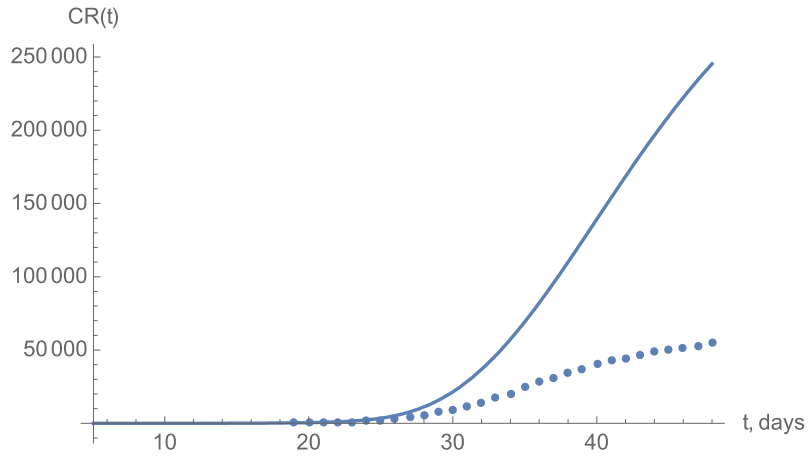
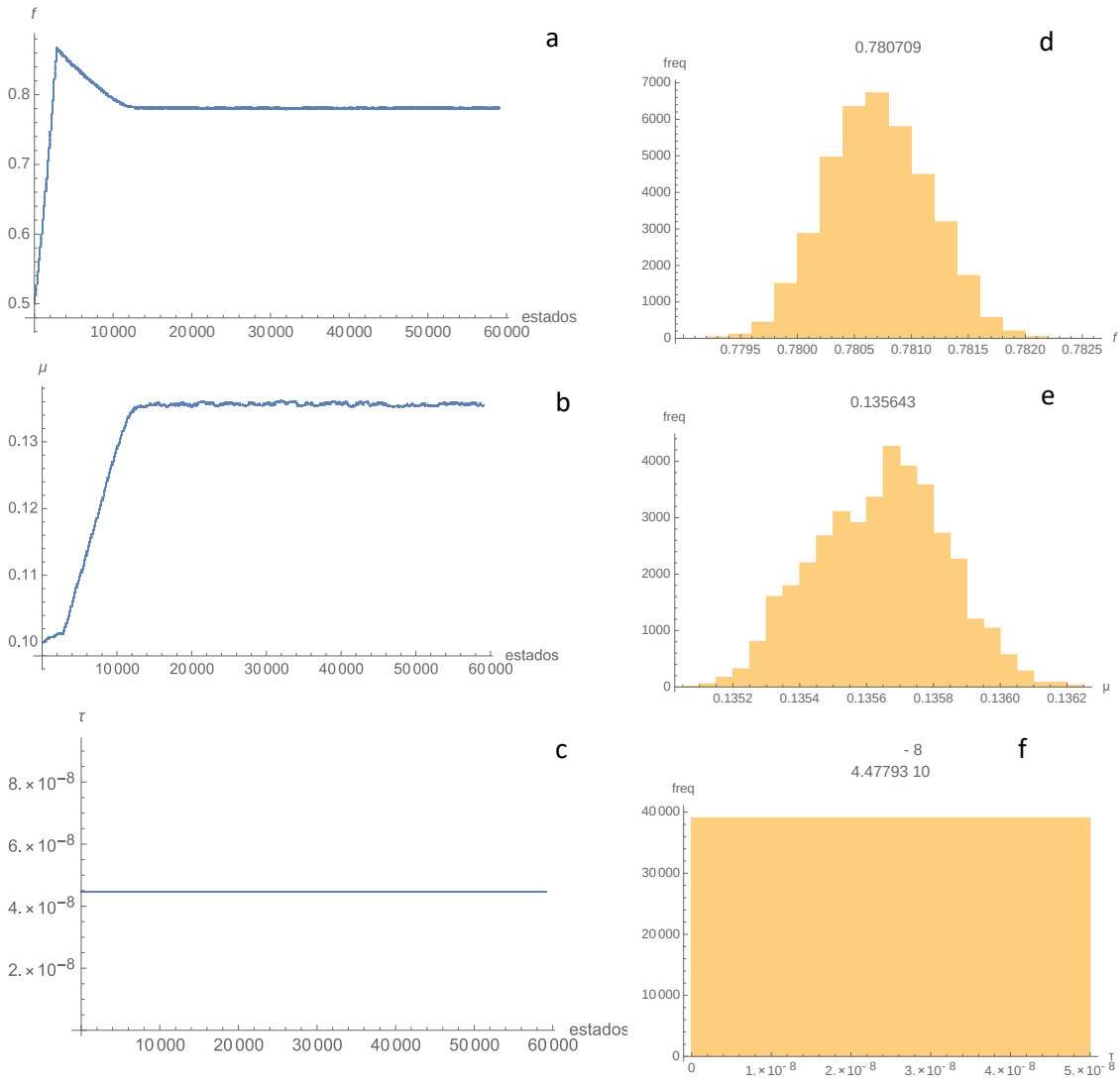



Figure 1 – Comparison of the model prediction (solid line) for the accumulated reported cases, CR(t), using the initial guesses from Table 1 against actual data from China from January 19th up to February 17th (dots).

Figures 2.a to 2.c show the complete Markov Chains for each estimated parameter, respectively, $f_0, \mu,$ and $\tau_0$. The central tendency (average value) of the posteriors here sampled, after neglecting the first 20,000 burning states of the chain, are called the estimated values. Both the estimated values and their 95% confidence intervals are presented in Table 2. It should be mentioned that these values are fairly close to those employed in [6], where $\tau_0$ was estimated as $4.51 \times 10^{-8}$. Once a value of $f_0 = 0.8$ was assumed, which means that 20% of symptomatic infectious cases go unreported, it led to a good agreement with the data by taking $\mu$=0.139. Figures 2.e to 2.f, complement the analysis offering an overview of the respective histograms of the sampled posteriors of $f, \mu$ and $\tau 0$, where the acceptance rate was about 55%.

Table 2 – Estimated values and 95% confidence intervals for three parameters, $f_0, \mu,$ and $\tau_0$ (China).

| **Parameter** | **Estimated values** | **95% confidence interval** |
|---|---|---|
| $f$ | 0.780709 | $[0.779638, 0.781853]$ |
| $\mu$ | 0.135643 | $[0.135193, 1.136101]$ |
| $\tau_0$ | $4.47793 \times 10^{-8}$ | $[4.47793 \times 10^{-8}, 4.47793 \times 10^{-8}]$ |




Figures 2 – Markov chains for the three estimated parameters, 2.a-c, and respective histograms of the sampled posteriors, 2.e-f.

Figure 3 demonstrates the markedly improved agreement of the model results and actual data within this portion of the dataset, once the estimated values in Table 2 are employed in the direct problem solution, as can be seen from the excellent agreement between the estimated CR(t) (solid line) and the experimental data from China (dots).

The desired model validation is then illustrated in Figure 4, confirming the excellent agreement of China's full dataset (period from January19th till March 22nd) with the mathematical model predictions, after adopting the estimated values for the parameters in Table 2. It should be recalled that non-informative priors were adopted for the three parameters, as presented in Table 1, and except for the transmission rate, when eq.(7.e) provides an excellent initial guess, the remaining guesses were completely

arbitrary, such as in the analysis for a less complete dataset, as will be discussed in the next section.

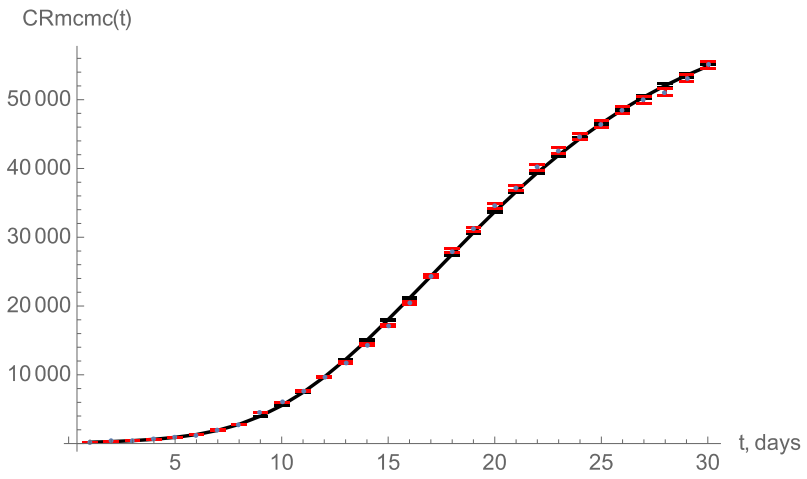



Figure 3 – Comparison of the estimated CR(t) (black line) with its 95% confidence interval limits against the actual data for China from January19th up to February17th (red dots).

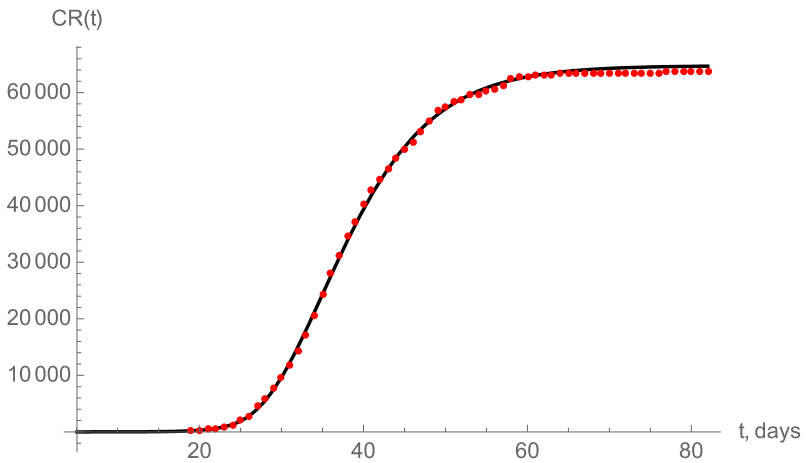



Figure 4 – Comparison of the theoretical model for CR(t) with the three estimated parameter values (solid line), against the complete dataset for China from January 19th up to March 25th (red dots).

Although the present estimated parameters have led to a good prediction of the second half of the China epidemic evolution data, there are still uncertainties associated with the average times here assumed both equal to 7 days, according to [6]. This choice was based on early observations of the infected asymptomatic and symptomatic patients in Wuhan, but more recent studies have been refining the information on the epidemic evolution and

the disease itself, such as in [14-17]. For this reason, we have also implemented a statistical inverse analysis with the full dataset of China, but now seeking the estimation of five parameters, so as to simultaneously estimate the average times (1/ν and 1/η). Both uniform and Gaussian distributions were adopted for the two new parameters, with initial guesses of 1/ν=7 days and 1/η=7 days, and N=29 days, as employed in [6]. Table 3 provides the estimated values and 95% confidence intervals for all five parameters, with Gaussian priors for the two average times with data obtained from [14,17]. The most affected parameter in comparison with the previous estimates is the average time 1/η, which is also the one with widest confidence interval. This behaviour is also evident from the Markov chains for this parameter, now simultaneously estimated. Figure 5 compares the theoretical predictions with the model incorporating the five estimated parameters as in Table 3, against the full CR(t) dataset for China, confirming the improved agreement.

Table 3 – Estimated values and 95% confidence intervals for five parameters, $f_0, \mu, \tau_0, 1/\nu$ and $1/\eta$ (China).

| **Parameter** | **Estimated values** | **95% confidence interval** |
|---|---|---|
| $f$ | 0.718491 | $[0.711595, 0.723138]$ |
| $\mu$ | 0.132032 | $[0.131789, 0.13227]$ |
| $\tau 0$ | $4.47793 \times 10^{-8}$ | $[4.47793 \times 10^{-8}, 4.47793 \times 10^{-8}]$ |
| $1/\nu$ | 6.20798 | [6.12574, 6.25764] |
| $1/\eta$ | 11.2784 | [10.4379, 12.3593] |

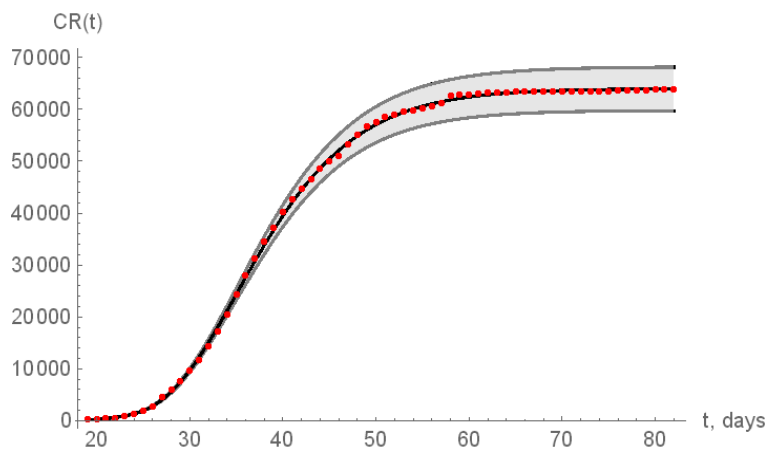



Figure 5 – Comparison of the theoretical model for CR(t) with the five estimated parameter values (black line), against the complete dataset for China from January 19th up to March 25th (red dots).

**Model Application: Brasil**

The CR(t) data for the accumulated reported infectious in Brasil, from February 25th, when the first infected individual was reported, up to March 29th, is presented in the Appendix. First, the exponential phase of the evolution was fitted, taking the data from day 10 to 25, yielding the estimates of the three parameters, $\chi_1 = 0.42552, \chi_2 = 0.293696, \chi_3 = 3.2335$, from which we have estimated $t_0 = 6.90514$. The remaining data for the initial conditions, $I_0$ and $U_0$, and the early stage transmission rate, $\tau_0$, are in fact recalculated from within the MCMC algorithm, since the changing values of $f_0$ will affect them, according to eqs. (7.c-e). The average times in the model were taken as 1/ν= 6.21 days and 1/η= 11.28 days, which were obtained from the MCMC simulation on the full dataset for China, as discussed in the previous section. Also, the Brazilian government took isolation measures starting on N=21 days. Then, the statistical inverse problem analysis was employed to estimate the three parameters, $f_0, \mu,$ and $\tau_0$, again assuming there is still no significant time variation in the function $f(t)$ ($\mu_f = 0$). Once more, in the absence of more informative priors, uniform distributions were employed for all three parameters under estimation. Table 4 presents the estimated values and 95% confidence intervals for the three parameters. It is clear that the transmission rate attenuation factor, $\mu$, is much less pronounced in the Brazilian case, in comparison to the China data behavior, which is possibly due to mild public health measures of isolation, or low overall adhesion to more severe proposed sanitary measures, at this early phase of the epidemic evolution. For instance, in the analysis of the Italy epidemic evolution reported in [6], with data from January 31st to March 8th, a comparable low attenuation factor of $\mu = 0.032$ was identified. It is also possible to observe the lower value of the parameter $f_0$, in comparison to the value obtained for the China dataset, which represents that only around 30% of the infected symptomatic individuals become in fact reported cases. This result could reflect an initial protocol of not thoroughly testing the mildly symptomatic individuals or just a lack of enough testing kits. This fact shall be discussed again further ahead when the impact of public health measures is analysed. Figure 6 illustrates the good agreement of Brasil's full dataset (period from February 25th till March 29th) with the mathematical model predictions, after adopting the estimated values for the parameters in Table 4. The theoretical CR(t) curve is plotted together with the 95% confidence interval bounds for this simulated evolution. It should be recalled that non-informative priors were adopted for the three parameters, as in the China example, and except for the transmission rate,

when eq. (7.e) provides an excellent initial guess, the remaining guesses were completely arbitrary. The initial conditions of the SIRU model are themselves dependent on the estimated parameters, thus the resulting initial values become $I_0$= 19.6146, $U_0$= 5.77965, $\mathcal{R}_0$= 4.96694. The initial value for the susceptible individuals is $S_0$= 211.3x10$^6$,

Table 4 – Estimated values and 95% confidence intervals for three parameters, $f_0, \mu,$ and $\tau_0$ (Brasil).

| **Parameter** | **Estimated values** | **95% confidence interval** |
|---|---|---|
| $f_0$ | 0.300567 | $[0.298584, 0.302429]$ |
| $\mu$ | 0.0554277 | $[0.0548846, 0.0561299]$ |
| $\tau_0$ | $1.66755 \times 10^{-9}$ | $[1.66755 \times 10^{-9}, 1.66755 \times 10^{-9}]$ |

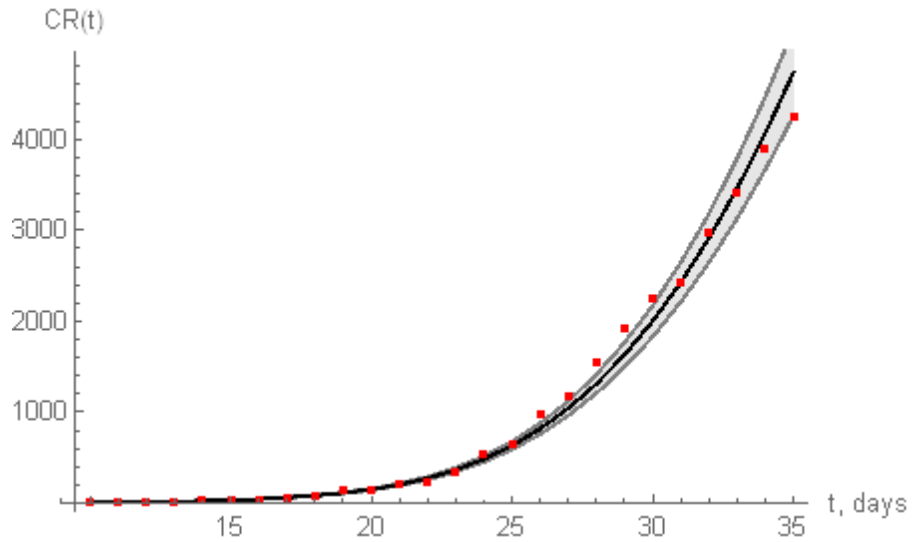



Figure 6 – Comparison of the theoretical model for CR(t) with the three estimated parameter values (black line) and the respective 95% confidence intervals (gray area), against the complete dataset for Brasil from February 25$^{th}$ up to March 29$^{th}$ (red dots).

Next, this parameter estimation is employed in the prediction of the COVID-19 evolution in Brasil. Five scenarios were here explored: (i) the present public health interventions remain unchanged; (ii) a stricter isolation is implemented from now on, further reducing the transmission rate; (iii) an attenuation on the social isolation policy, leading to an increased transmission rate; (iv) an increment on the fraction of reported cases, through a massive blood testing campaign, for instance, forcing more unreported cases to become

reported ones, thus isolating them earlier; (v) a combination of public health measures acting on both reducing the transmission rate and on increasing the conversion factor of unreported to reported cases;

In the first scenario, it is assumed that no further public health interventions are implemented, other than those already reflected by the data which should be fully maintained throughout the control period, and the epidemics should evolve from the present stage, under the parameters above identified. Figure 7.a shows the evolution of the accumulated reported, CR(t), and unreported, CU(t), infectious individuals up to 150 days. Due to the fairly low value of $f_0 \approx 0.30$, the number of unreported infectious cases is quite high, reaching around 84,968 individuals, while the reported cases should reach 36,514 individuals, thus a total of infected symptomatic individuals of 121,482. No predictions on casualties are here proposed, since these are highly dependent on age structure, social-economical conditions, and health system response. Figure 7.b presents the predicted evolution of the daily reported infectious cases, which shows a peak at around t=47 days of about 1,067 reported cases.

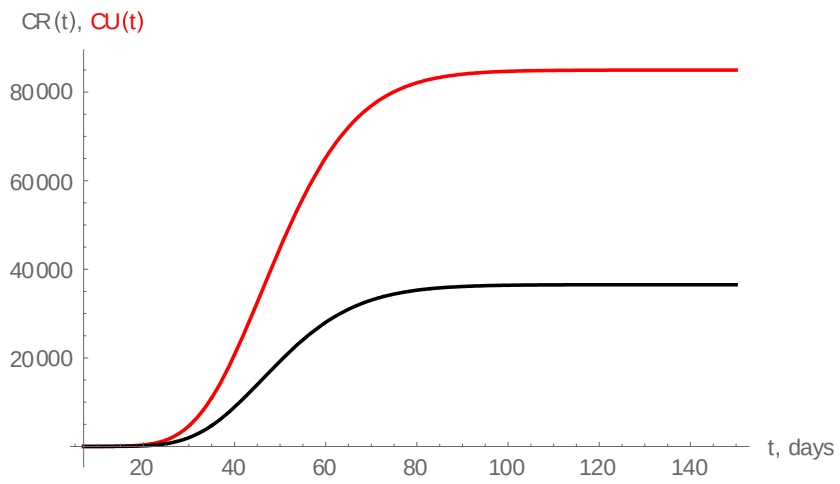



Figure 7.a – Comparison of the theoretical model for CR(t) (black curve) and CU(t) (red curve) with the three estimated parameter values from the available dataset for Brasil from February 25th up to March 29th.

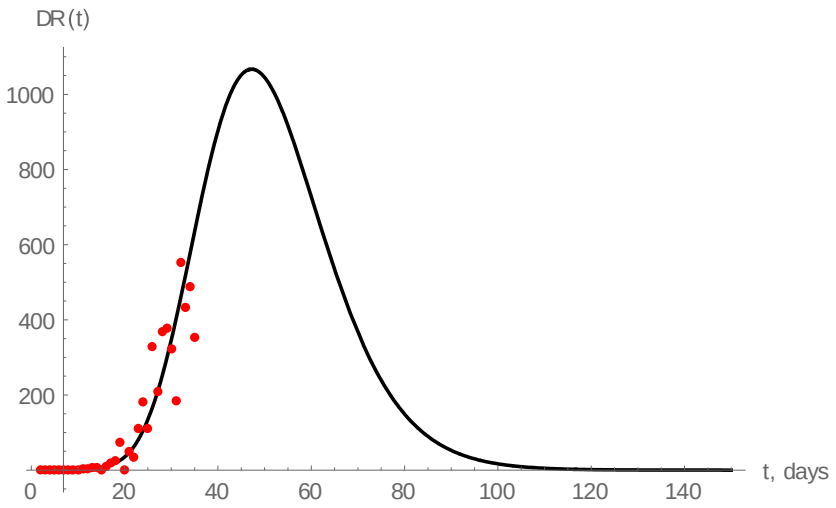



Figure 7.b – Prediction of the daily reported data distribution, DR(t), with the three estimated parameter values from the available daily reported cases dataset for Brasil from February 25th up to March 29th (red dots).

Next, the second scenario explores the implementation of more strict isolation and sanitary measures to reduce the transmission rate by assuming, after day $N_2$=40 (eq.10.c), 50% improvement with respect to the value of $\mu$ here identified , thus around, $\mu_2$=0.0831, still below that achieved in China (0.132), hopefully still feasible in Brasil. The time variable transmission rate is then computed from:

$$\tau(t) = \tau_0 \,, \qquad 0 \le t \le N \tag{10.a}$$

$$\tau(t) = \tau_0 \exp\big(-\mu(t-N)\big), \qquad N < t \le N_2 \tag{10.b}$$

$$\tau(t) = \tau_0 \exp\big(-\mu(N_2-N)\big) \exp\big(-\mu_2(t-N_2)\big), \qquad t > N_2 \tag{10.c}$$

The changes in the accumulated reported and unreported cases, as shown in Figure 8, are quite significant. The predicted number of unreported symptomatic infectious cases is now much lower reaching after 150 days around 67,360 individuals, while the reported cases should reach 28,947 individuals, with an impressive reduction to a total of around 96,307 infectious cases. The predicted evolution of the daily reported infectious cases would then show a peak at around t=45 days of about 1,013 reported cases.

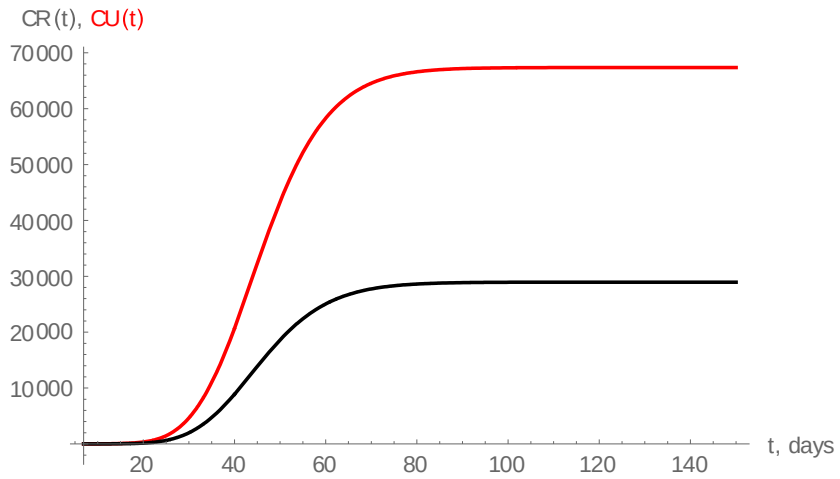



Figure 8 – Comparison of the theoretical model for CR(t) (blue curve) and CU(t) (red curve) with the three estimated parameter values from the available dataset for Brasil from February 25th up to March 29th, and improving by 50% the transmission rate attenuation parameter, to become $\mu_2 = 0.0831$, after day $N_2$=40.

Through the third scenario, one can predict the consequences of relaxing the public health measures that affect transmission rate, for instance by relaxing somehow the isolation and sanitary measures. This is simulated here by reducing the identified transmission rate attenuation factor, by assuming, after day $N_2$=40, half the value of $\mu$ here identified, thus around, $\mu_2$=0.0277. The changes in the accumulated reported and unreported symptomatic cases, as shown in Figure 9, are marked changed to worse. The predicted number of unreported infectious cases is now much lower reaching after 150 days around 147,815 individuals, while the reported cases should reach 63,521 individuals, with a drastic increase to a total of around 211,336 infectious cases. The predicted evolution of the daily reported infectious cases would then show a peak at around t=56 days of about 1,300 reported cases.

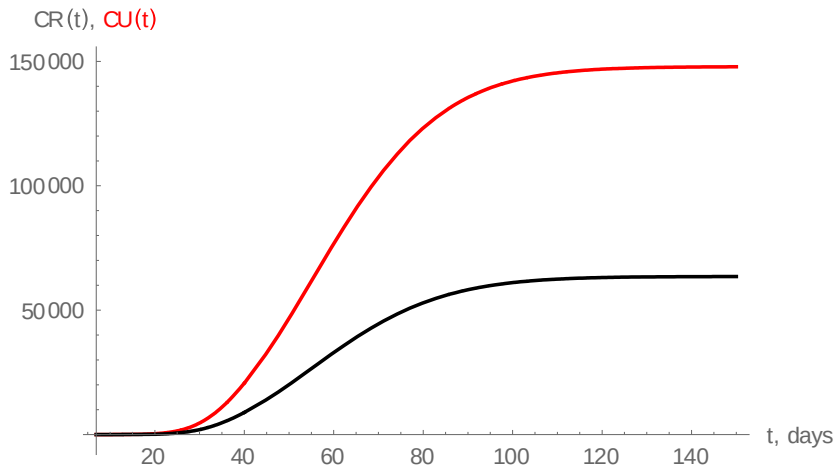



Figure 9 – Comparison of the theoretical model for CR(t) (black curve) and CU(t) (red curve) with the three estimated parameter values from the available dataset for Brasil from February 25th up to March 29th, and halving the transmission rate attenuation parameter, to become $\mu_2 = 0.0277$, after day $N_2$=40.

Besides acting on the transmission rate along time, public health measures may also be effective in reducing the ratio of reported to unreported infectious case, since the reported cases are directly isolated and thus interrupting the contamination path, as analyzed in the fourth scenario. For instance, increasing the fraction of reported and unreported infectious cases parameter, to become $f = 0.7185$, the value previously obtained from the China dataset. Therefore, Figure 10 shows the behavior of both CR(t) and CU(t), which according to the value of $\mu_f = 0.5$, occurring after the day $N_f$=40, leads to the crossing of reported and unreported cases that can be observed. The predicted number of unreported infectious cases is now reaching, after 150 days, around 40,894 individuals, while the reported cases should reach 60,531 individuals, with an also marked reduction to a total of around 101,425 infectious cases. The predicted evolution of the daily reported infectious cases would then show a peak at around t=47 days of about 2,341 reported cases. Although this peak value is higher than for the base case (1,067), before further public health intervention, a number of these are of mild symptomatic cases that were moved from the unreported to the reported cases evolution, thus isolated earlier.

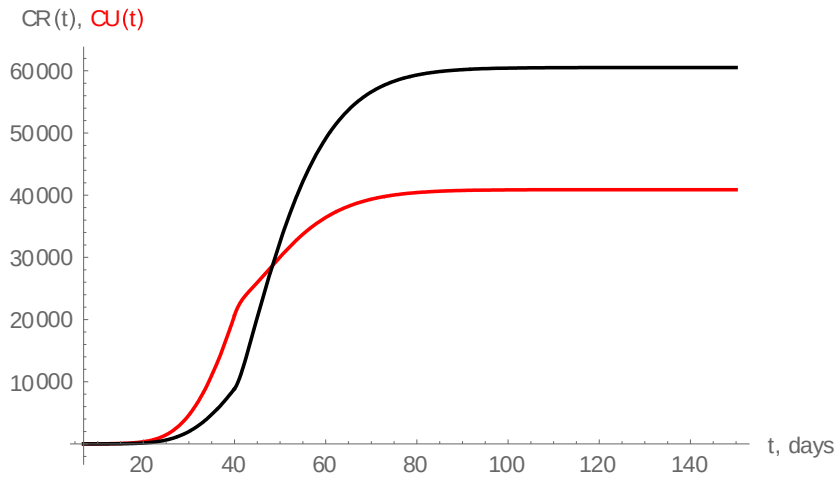



Figure 10 – Comparison of the theoretical model for CR(t) (black curve) and CU(t) (red curve) with the three estimated parameter values from the available dataset for Brasil from February 25th up to March 29th, and increasing the fraction of reported and unreported symptomatic infectious cases parameter, to become $f = 0.7185$, starting after day $N_2$=40 with $\mu_f$=0.5.

Finally, in the fifth scenario, the combination of public health measures acting on both the transmission rate and on the conversion factor or unreported to reported cases is analyzed for Brasil. Therefore, let us consider after day $N_2$=40, 50% improvement with respect to the value of $\mu$ here identified, thus around, $\mu_2$=0.0831, and simultaneously increase the fraction of reported and unreported infectious cases, to become $f = 0.7185$, also starting after day $N_2$=40, with $\mu_f$=0.5. The changes in the accumulated reported and unreported cases, as shown in Figure 11, are the most encouraging in the present analysis. The predicted number of unreported infectious cases is now reaching after 150 days around 36,770 individuals, while the reported cases should reach 50,006 individuals, with a marked decrease to a total of around 86,777 infectious cases, about 30% reduction with respect to the base case. The predicted evolution of the daily reported infectious cases would then show a peak at around t=46 days of about 2,196 reported cases. Again, though this peak value is higher than for the base case, before the public health improvements, a number of these are of mild symptomatic cases that were moved from the unreported to the reported cases evolution, thus moved to monitored isolation earlier, and not necessarily requiring hospitalization.

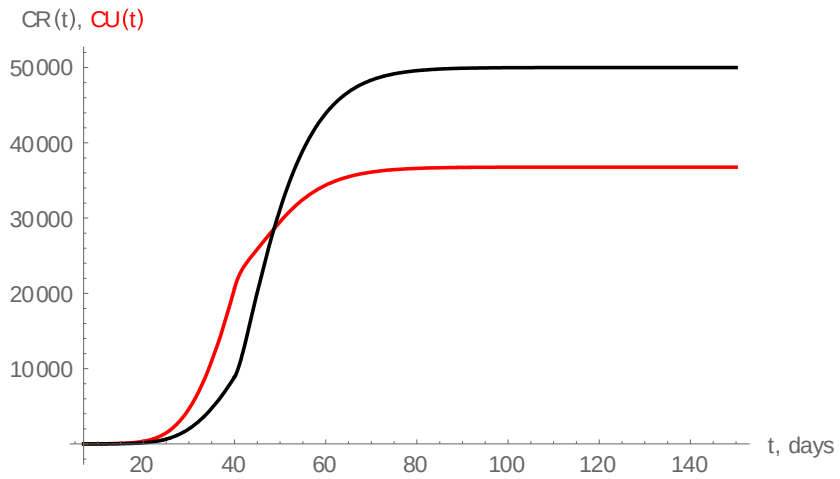



Figure 11 – Comparison of the theoretical model for CR(t) (black curve) and CU(t) (red curve) with the three estimated parameter values from the available dataset for Brasil from February 25th up to March 29th, and combining public health measures, towards improving by 50% the transmission rate attenuation parameter, to become $\mu_2 = 0.0831$, after day $N_2$=40, and increasing the fraction of reported and unreported symptomatic infectious cases parameter, to become $f = 0.7185$, starting after day $N_2$=40 with $\mu_f$=0.5.

Though the three parameters estimation provides a fairly good reproduction of the behaviour of the CR(t) curve for Brasil, one may observe a change in the pattern of the evolution around day 30, that could not be entirely followed by the proposed model. It is also a known fact that the initial amount of kits for blood testing that were purchased by the Brazilian government were finished around this time, and before being fully supplemented, there could have been a reduction on the number of executed exams of the symptomatic individuals, that might have affected the partition of reported to unreported cases by the end of this period covered by the present dataset. Therefore, the more general model including the time variation of the partition f(t), eqs.(4.c,d), is here implemented for a more refined inverse problem analysis. It is then expected that a reduction on the f value can be identified ($f_{max}<f_0$), with an abrupt variation on the exponential behaviour, here assumed as a sharp functional time dependence (large $\mu_f$). Therefore, an additional statistical inverse problem analysis is undertaken, this time for estimating five parameters, namely, $f_0, \mu, \tau_0, f_{max}$, and $N_f$ , aimed at improving the overall agreement with the CR(t) data behaviour, with a possible reduction on the partition of the reported and unreported

infectious cases. With uniform distributions for all five parameters, taking the previous estimates for the three first parameters, an arbitrary guesses for $f_{max}$, and $N_f$, the five estimated parameters are shown in Table 5, together with the 95% confidence interval for each parameter. Figure 12 shows the theoretical CR(t) curve obtained with the five parameters estimation, plotted together with the 95% confidence interval bounds for this simulated evolution.

Table 5 – Estimated values and 95% confidence intervals for five parameters, $f_0, \mu, \tau_0, f_{max}$, and $N_f$ (Brasil).

| **Parameter** | **Estimated values** | **95% confidence interval** |
|---|---|---|
| $f$ | 0.303671 | $[0.302624, 0.304697]$ |
| $\mu$ | 0.0389639 | $[0.0388438, 0.0390961]$ |
| $\tau 0$ | $1.66755 \times 10^{-9}$ | $[1.66755 \times 10^{-9}, 1.66755 \times 10^{-9}]$ |
| $f_{max}$ | 0.156734 | [0.156146, 0.157217] |
| $N_f$ | 30.4197 | [30.3522, 30.4915] |

One can see the marked reduction on the *f*(*t*) parameter from the estimates in Table 5, which results in the increase of the unreported to reported infectious cases, as is shown in Figure 13.a for CR(t) and CU(t) predictions up to 150 days. Clearly, the reduction on the testing, and thus on the isolation of reported infectious individuals, leads to an impressive increase on the total number of infected individuals after 150 days (723,698 cases), including unreported (609,125) and reported cases (114,572). Figure 13.b presents the predicted evolution of the daily reported infectious cases, which shows a peak at around t=61 days of about 2,672 reported cases.

Hopefully this difficulty with the availability of enough testing kits that occurred around day 30 has been already solved and the desirable increase on the number of tests and reported cases will be apparent from the next few entries in the accumulated reported cases. From the present results it is quite clear that the reduction on the testing has unfortunate consequences on the epidemic evolution. At the end of this report, the predicted results for CR(t) provided the value of 5438 reported cases, in comparison to the officially announced value of 5717 cases on March 31st, 2020.

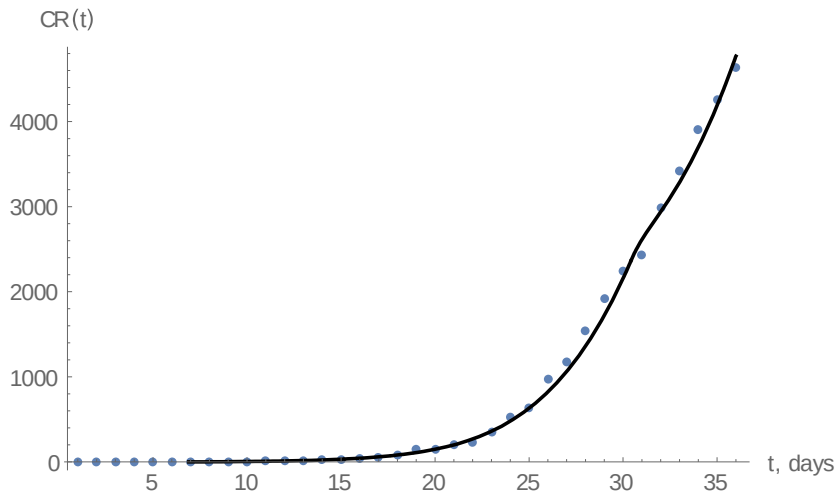



Figure 12 – Comparison of the theoretical model for CR(t) with the five estimated parameter values (black line) and the respective 95% confidence intervals (gray area), against the complete dataset for Brasil from February 25th up to March 29th (red dots).

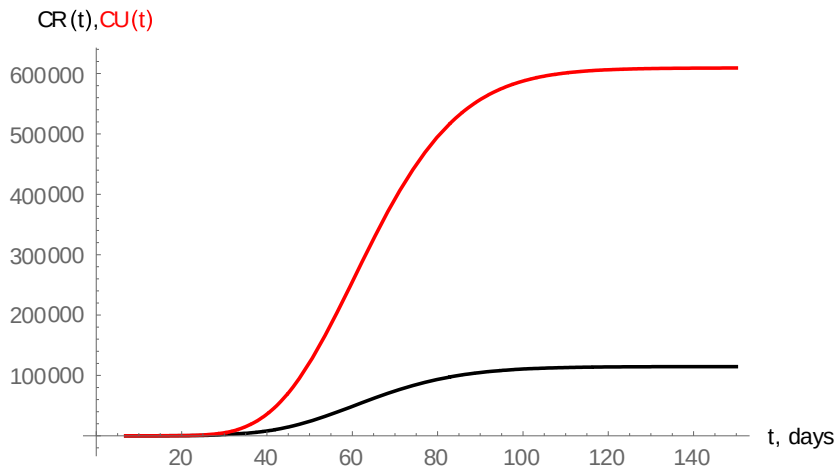



Figure 13.a – Comparison of the theoretical model for CR(t) (black curve) and CU(t) (red curve) with the five estimated parameter values from the available dataset for Brasil from February 25th up to March 29th.

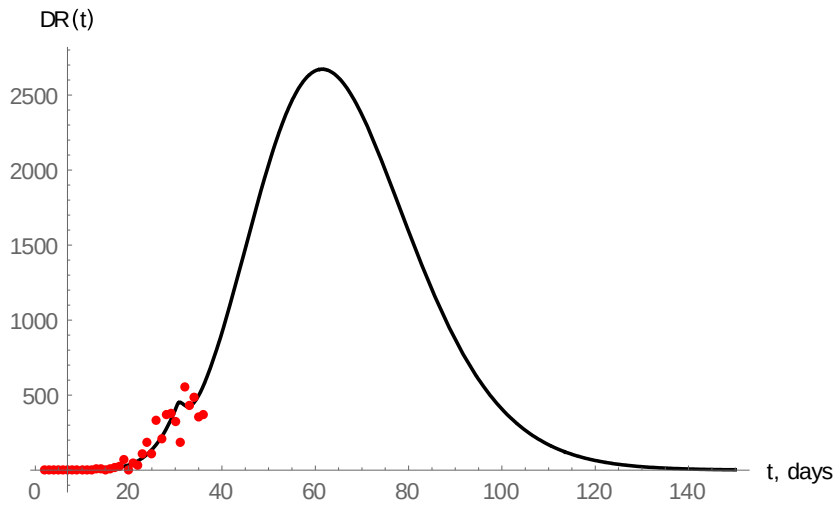



Figure 13.b – Prediction of the daily reported data distribution, DR(t), with the five estimated parameter values from the available daily reported cases dataset for Brasil from February 25th up to March 29th (red dots).

## CONCLUSIONS

The present work implements a mixed analytical-statistical inverse problem analysis to the prediction of epidemics evolution, with focus on the COVID-19 progression in Brasil. A SIRU-type model is implemented for the direct problem solution, while a mixture of an analytical parametric estimation for the early phase epidemic exponential behavior with a Bayesian inference approach for the entire period, are considered for the inverse problem analysis. The evolution of the COVID-19 epidemy in China is considered for validation purposes, by taking the first part of the dataset to estimate parameters, and retaining the rest of the evolution data for direct comparison with the predicted results, with excellent agreement. Then, the same approach is applied to the Brazilian case, this time employing the available time series so far for the parametric estimates, and then offering an evolution prediction. Also, some public health intervention measures are critically examined, in addition to those already implemented, permitting the inspection of their impact on the overall dynamics of the disease proliferation. Clearly, a combination of public health interventions can offer a considerable impact reduction on the disease progression within Brasil, as illustrated by the implemented modelling. It was also analyzed the negative impact due to the scarcity of testing kits during a period, which if not solved and even incremented, would lead to an increase on the ratio of unreported to reported symptomatic cases, and consequently on a dramatic epidemic evolution.

Further improvement on the modelling is envisioned by enriching the model with latency effects, age structure discrimination, spatial demographic distribution dependence, and recovery factor differentiation among isolated and non-isolated patients.

**ACKNOWLEDGEMENTS**

The authors are deeply grateful to Dr. Tania Mattos Petraglia, MD, for the valuable information on COVID-19 pathology and treatment.

**REFERENCES**

[1] Tang, B.; Wang, X.; Li, Q.; Bragazzi, N.L.; Tang, S.; Xiao, Y.; Wu, J. Estimation of the Transmission Risk of 2019-nCov and Its Implication for Public Health Interventions. J. Clin. Med. 2020, 9, 462. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3525558

[2] Z.Liu, P.Magal, O.Seydi, and G.Webb, Understanding unreported cases in the 2019 - nCov epidemic outbreak in Wuhan, China, and the importance of major public health interventions, MPDI Biology 2020, 9 (3), 50.

[3] Magal, P.; Webb, G. The parameter identification problem for SIR epidemic models: Identifying Unreported Cases. J. Math. Biol. 2018, 77, 1629–1648.

[4] Ducrot, A.; Magal, P.; Nguyen, T.;Webb, G. Identifying the Number of Unreported Cases in SIR Epidemic Models. Math. Med. Biol. J. IMA 2019.

[5] 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, SSRN (2020)

[6] P.Magal and G.Webb, Predicting the number of reported and unreported cases for the COVID - 19 epidemic in South Korea, Italy, France and Germany, SSRN (2020)

[7] Z.Liu, P.Magal, O.Seydi, and G.Webb, A COVID - 19 epidemic model with latency period SSNR (2020)

[8] Kaipio, J., and E. Somersalo, Statistical and Computational Inverse Problems, Applied Mathematical Sciences, V. 160, Springer-Verlag, 2004.

[9] Gamerman, D., and H.F. Lopes, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd ed., Chapman & Hall/CRC, Boca Raton, FL, USA, 2006.

[10] Orlande, H.R.B., O. Fudym, D. Maillet, R.M. Cotta, Thermal Measurements and Inverse Techniques, CRC Press, Boca Raton, FL, USA, 2011.

[11] Kruschke, J.K., Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan, Academic Press, NY, 2015.
[12] Orlande, H.R.B., Inverse Problems in Heat Transfer: New Trends on Solution Methodologies and Applications, ASME J. Heat Transfer, V.134, pp.031011, 2012.
[13] https://www.worldometers.info/coronavirus/country/#countries
[14] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, and J. Lessler, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, Annals of Internal Medicine, published at Annals.org on 10 March 2020. doi:10.7326/M20-0504
[15] Q. Li et al., Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia, The New England Journal of Medicine, published on January 29, 2020, and last updated on January 31, 2020, at NEJM.org. doi: 10.1056/NEJMoa2001316
[16] Neil M Ferguson et al., Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, Imperial College COVID-19 Response Team, March 16th, 2020.
[17] B. Tang, X. Wang, Q. Li., N.L. Bragazzi, Sa.Tang, Y. Xiao, J Wu, Estimation of the transmission risk of 2019-nCov and its implication for public health interventions, SSRN, 20 Pages, Posted: 27 Jan 2020.

**APPENDIX**

**Table A.1 - Data for Brasil - accumulated reported cases, CR(t), and casualties.**

| DATE | Death | Infected |
|---|---|---|
| 24/02/2020 | 0 | 0 |
| 25/02/2020 | 0 | 1 |
| 26/02/2020 | 0 | 1 |
| 27/02/2020 | 0 | 1 |
| 28/02/2020 | 0 | 1 |
| 29/02/2020 | 0 | 2 |
| 01/03/2020 | 0 | 2 |
| 02/03/2020 | 0 | 2 |
| 03/03/2020 | 0 | 2 |
| 04/03/2020 | 0 | 3 |
| 05/03/2020 | 0 | 8 |
| 06/03/2020 | 0 | 13 |
| 07/03/2020 | 0 | 19 |
| 08/03/2020 | 0 | 25 |
| 09/03/2020 | 0 | 25 |
| 10/03/2020 | 0 | 34 |
| 11/03/2020 | 0 | 52 |
| 12/03/2020 | 0 | 77 |
| 13/03/2020 | 0 | 151 |
| 14/03/2020 | 0 | 151 |
| 15/03/2020 | 0 | 200 |
| 16/03/2020 | 0 | 234 |
| 17/03/2020 | 1 | 346 |
| 18/03/2020 | 4 | 529 |
| 19/03/2020 | 7 | 640 |
| 20/03/2020 | 11 | 970 |
| 21/03/2020 | 18 | 1178 |
| 22/03/2020 | 25 | 1546 |
| 23/03/2020 | 34 | 1924 |
| 24/03/2020 | 46 | 2247 |
| 25/03/2020 | 57 | 2433 |
| 26/03/2020 | 77 | 2985 |
| 27/03/2020 | 92 | 3417 |
| 28/03/2020 | 111 | 3904 |
| 29/03/2020 | 136 | 4256 |