

# Scrutinizing the heterogeneous spreading of COVID-19 outbreak in Brazilian territory

Rafael M. da Silva<sup>1,\*</sup>, Carlos F.O. Mendes<sup>2,†</sup>, and Cesar Manchein<sup>3,‡</sup>

<sup>1</sup>Departamento de Física, Universidade Federal do Paraná, 81531-980 Curitiba, PR, Brazil

<sup>2</sup>Escola Normal Superior, Universidade do Estado do Amazonas, 69050-010 Manaus, AM, Brazil

<sup>3</sup>Departamento de Física, Universidade do Estado de Santa Catarina, 89219-710 Joinville, SC, Brazil

E-mail: \*rmarques@fisica.ufpr.br

E-mail: †cfabio.mendes@gmail.com

E-mail: ‡cesar.manchein@udesc.br

June 5, 2020

**Abstract.** After the spread of COVID-19 out of China, the evolution of the pandemic shows remarkable similarities and differences among countries across the world. Eventually, such characteristics are also observed between different regions of the same country. Herewith, we study the heterogeneous spreading of the confirmed infected cases and deaths by the COVID-19 until May 30<sup>th</sup>, 2020, in the Brazilian territory, which has been seen as the current epicenter of the pandemic in South America. Our first set of results is related to the similarities and it shows that: (i) a power-law growth of the cumulative number of infected people is observed for federative units of the five regions of Brazil; and (ii) the Distance Correlation (DC) calculated between the time series of the most affected federative units and the curve that describes the evolution of the pandemic in Brazil remains about 1 in most of the time, while such quantity calculated for the federative units with a low incidence of newly infected people remains about 0.95. In the second set of results, we focus on the heterogeneous distribution of the confirmed cases and deaths, which is demonstrated by the fact that only three regions concentrate 92% of the cases. By applying the epidemiological SIRD model we estimated the effective reproduction number  $\mathcal{R}_e$  during the pandemic evolution and found that: (i) the mean value of  $\mathcal{R}_e$  for the eight most affected federative units in Brazil is about 2; (ii) the current value of  $\mathcal{R}_e$  for Brazil is greater than 1, which indicates that the epidemic peak is far; and (iii) Ceará was the only federative unit for which the current  $\mathcal{R}_e < 1$ . Based on these findings, we projected the effects of increase or decrease the effective reproduction number and concluded that if the value of  $\mathcal{R}_e$  increases 20%, not only the peak might grow at least 40% but also its occurrence might be anticipated, which hastens the collapse of the public health care system. In all cases, to keep the effective reproduction number 20% below the current one can save thousands of people in the long term.

*Keywords:* Coronavirus, COVID-19, Distance Correlation, SIRD model

## 1. Introduction

According to the World Health Organization (WHO), the Coronavirus Disease 2019 (COVID-19) is provoked by a pathogen agent known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) [1]. Since the identification of SARS-CoV-2 in Wuhan, China, in December 2019, the pathogen kept spreading worldwide. One of the most remarkable characteristics of COVID-19 is its high infectivity, resulting in a global pandemic. Nowadays, tasks like protecting the people from the impacts of COVID-19 pandemic and the global economy are considered two major challenges. In this scenario, questions regarding the occurrence of the COVID-19 outbreak peak, how long it will persist, and how many people will be eventually infected are guiding a lot of studies.

Around the world, the battle against the COVID-19 spreading is based on several research approaches, which include treatments and preventive actions. However, to control the pandemic is found to be very challenging due to the following reasons: (i) it causes typical flu-like symptoms in human carriers; (ii) the human-to-human transmission via asymptomatic individuals; and (iii) the absence of proper clinical protocols (vaccine, drugs, and concrete ideas about the immunological response, for instance). In this complex scenario, an important research approach used to study the dynamics of the COVID-19 is the real time-series analysis, that has been extensively applied to obtain information about the evolution of the pandemic (infection, recovery, and death rates, for instance) [2, 3, 4, 5]. On the other hand, the uncertainty of the available official data, particularly related to the true number of infectious people, may lead to ambiguous results and inaccurate forecasts, as pointed out by other researchers [6] (see also references therein). For this reason, to explore the dynamics of COVID-19 by using combined approaches may be a necessary strategy to avoid misleading results.

Other important approaches used to study the evolution of COVID-19 are the epidemiological models like SIRD (Susceptible-Infected-Recovered-Dead) [3, 5, 7], SEIR (Susceptible-Exposed-Infected-Recovered) [8, 9, 10], and its modified versions [11, 12, 13, 14, 15, 16], which are fundamental tools to understand the evolution of the pandemic and to plan effective control strategies. One of the most important characteristics of these models is the quantity used to measure

the transmission potential of the disease: the basic reproduction number  $\mathcal{R}_0$ , which gives the number of secondary infected individuals generated by a primary infected in a population where all individuals are susceptible to infection. While for values  $\mathcal{R}_0 < 1$  the number of newly cases decreases exponentially, for  $1 < \mathcal{R}_0 < \infty$  it increases exponentially [17, 18].

In recent studies about the COVID-19 outbreak, it was found a power-law growth of the cumulative number of infected individuals [11, 12, 13, 19, 20], which might be typical of small-world networks [21], and possibly is related to fractal kinetics and graph theory [22]. Such behavior seems to be general since the real time-series of countries from four distinct continents were fitted by power-law curves until March 27<sup>th</sup> [11]. Besides that, by calculating the Distance Correlation (DC) [23, 24, 25, 26] between the time series of these countries, it was possible to demonstrate that such data are highly correlated, which suggests a universal characteristic of the virus spreading.

Several studies have focused on the contagion scenarios in China [4, 5], Europe [5, 11, 12, 16], and United States of America (USA) [11, 12], due to the trajectory followed by the COVID-19 outbreak. As far as we know, there are only a few studies analyzing the dynamics of the COVID-19 in Brazil [27, 28, 29], which has the largest cumulative numbers of infected cases and deaths of South America. For this reason, we decided to investigate the dynamics of the COVID-19 in Brazil, which is the home of more than 210 million people, the world's fifth-largest country by area, and currently the world's eighth-largest economy. Another reason to study the dynamics of the COVID-19 outbreak in Brazil is due to the remarkable heterogeneity in the distribution of the infected people across the Brazilian territory. Brazil is composed of 26 federative units (or states) and one Federal District, which are distributed in five large regions: South, Southeast, Central-West, Northeast, and North. Besides the intrinsic differences between these regions, as the weather conditions (temperature and the duration of the raining seasons, for instance), cultural behaviors, and demographic density of people, the Human Development Indexes (HDI) are completely different if compared between the federative units. Interestingly, until May 30<sup>th</sup>, 32% of the cumulative number of infected cases and 44% of the deaths reported in Brazil due to COVID-19 happened in São Paulo (SP) and Rio de Janeiro (RJ), which represent

the two largest economic units of the country and also concentrate a significant part of the population.

To scrutinize the heterogeneity of the COVID-19 spreading in Brazil, we propose to analyze the real time-series of all the federative units. Our main findings show that: (i) the cumulative number of infected people grows according to a power-law in all regions of Brazil; and (ii) the DC calculated between the time series of the most affected federative units and the time series of Brazil remains about 1 in most of the time, indicating a high similarity between the spread of the disease across such localities. Besides that, we used the SIRD model to estimate the effective reproduction number  $\mathcal{R}_e$  and to project some scenarios based on its current value. A better and a worse scenario for the cumulative and active number of infected people, as well as for the number of dead people, were simulated and our findings corroborate the relevance of keeping the value of  $\mathcal{R}_e$  the most next of 1 as possible. Additionally, the current value of the effective reproduction number reveals the epidemic stage and, in this way, may be used to indicate the right moment to act through containment measures.

The manuscript is organized as follows. Section 2 is designated to introduce an explanation about the real time-series collected and to define the methods applied here. Also, we explain the method to compute the DC between the time series and introduce the SIRD model formed by four Ordinary Differential Equations (ODEs). Section 3 discusses the numerical results obtained by fitting the real time-series, by calculating the DC between such data, and by applying the SIRD model to project different scenarios for the most affected Brazilian federative units. Finally, Section 4 summarizes our results and suggests some directions for future research studies.

## 2. Data and methods

This section is devoted to introducing the basic information about the real time-series and the methods used to explore the dynamics of COVID-19 spreading in the five Brazilian regions.

### 2.1. Real time-series

The real time-series refers to the cumulative number of confirmed cases of COVID-19 and deaths until May 30<sup>th</sup>, 2020, reported in the 27 federative units and also for the compiled results that form the time series of Brazil. Fundamental information about federative units, like their geographical position in Brazilian territory, the total population, the date of the first case, the cumulative number of confirmed cases, and deaths caused by the COVID-19 are detailed in Table 1. The time series were collected from the official website [30]

that retrieves the daily information about COVID-19 cases from all 27 Brazilian State Health Offices, gather them, and make it publicly available. Among the 27 federative units, we focused our numerical investigations on 10 representative cases (two of each region) with a high incidence of COVID-19 cases: São Paulo (SP), Rio de Janeiro (RJ), Ceará (CE), Pernambuco (PE), Amazonas (AM), Pará (PA), Santa Catarina (SC), Paraná (PR), Federal District (FD) and Goiás (GO), which concentrate  $\sim 69\%$  of the cumulative number of confirmed cases and  $\sim 82\%$  of deaths. The time-series period of each state varies, once each state counts since the day of its first documented case until the day of the last report.

### 2.2. Distance Correlation (DC)

The Distance Correlation (DC) is a statistical measure of the dependence between random vectors based on Euclidean distances [23] and is derived from quantities such as *variance* and *covariance of distance*. DC assumes null values if, and only if, the random vectors are completely independent, and can be defined from random vectors in arbitrary dimensions. For example, assuming that  $l$  and  $m$  are positive integers, the vectors can be defined as  $X \in \mathbb{R}^l$  and  $Y \in \mathbb{R}^m$ . In this case, the *distance covariance* between random vectors  $X$  and  $Y$  is defined by

$$\begin{aligned} \sigma^2(X, Y) &= \| \varphi_{X,Y}(t, v) - \varphi_X(t)\varphi_Y(v) \|^2, \\ &= \int_{\mathbb{R}^l} \int_{\mathbb{R}^m} \frac{|\varphi_{X,Y}(t, v) - \varphi_X(t)\varphi_Y(v)|^2}{c_l c_m |t|^{l+1} |v|^{m+1}} dt dv, \end{aligned} \quad (1)$$

where  $c_l$  and  $c_m$  are constants,  $\| \cdot \|$  is the Euclidean norm,  $\varphi_X(t)$  and  $\varphi_Y(v)$  are characteristic functions of  $X$  and  $Y$ , respectively, and  $\varphi_{X,Y}(t, v)$  is the joint characteristic function. The joint characteristic function under the independence of two random vectors leads to the property  $\sigma^2(X, Y) = 0$  if, and only if,  $X$  and  $Y$  are independent, that is,  $\varphi_{X,Y}(t, v) = \varphi_X(t)\varphi_Y(v)$  for all  $t \in \mathbb{R}^l$  and  $v \in \mathbb{R}^m$ . We can obtain similarly the *distance variance*, given by

$$\sigma^2(X, X) = \sigma^2(X) = \| \varphi_{X,X}(t, v) - \varphi_X(t)\varphi_X(v) \|^2, \quad (2)$$

which is analogous for  $\sigma^2(Y)$ . Therefore, the DC coefficient between random vectors is defined by

$$\mathcal{C}(X, Y) = \frac{\sigma(X, Y)}{\sqrt{\sigma(X)}\sqrt{\sigma(Y)}}, \quad (3)$$

and such statistical measure is contained in the interval  $[0, 1]$ .

*2.2.1. Computational procedure to compute the DC:* The computational method defined here will be used to analyze statistically the real data of

**Table 1.** Data for the 27 federative units and for Brazil related to COVID-19 outbreak until May 30<sup>th</sup> [30]. Inside the block of each region, the federative units are listed in decreasing order of the number of confirmed infected cases.

Region	Locality	Acronyms	Population	First case	Confirmed cases	Deaths
North	Amazonas	AM	4144597	03/15/2020	40560	2047
	Pará	PA	8602865	03/19/2020	37296	2900
	Amapá	AP	845731	03/20/2020	9313	215
	Acre	AC	881935	03/19/2020	6072	142
	Rondônia	RO	1777225	03/20/2020	4743	151
	Tocantins	TO	1572866	03/19/2020	3981	71
Northeast	Roraima	RR	605761	03/22/2020	3299	110
	Ceará	CE	9132078	03/17/2020	46506	2956
	Pernambuco	PE	9557071	03/12/2020	33427	2740
	Maranhão	MA	7075181	03/21/2020	32620	932
	Bahia	BA	14873064	03/06/2020	17626	638
	Paraíba	PB	4018127	03/19/2020	12862	347
	Alagoas	AL	3337357	03/08/2020	9223	424
	Rio Grande do Norte	RN	3506853	03/13/2020	7402	305
Cent.-West	Sergipe	SE	2298696	03/15/2020	6805	149
	Piauí	PI	3273227	03/20/2020	4745	157
	Federal District	FD	3015268	03/07/2020	9474	162
	Goias	GO	7018354	03/13/2020	3585	122
Southeast	Mato Grosso	MT	3484466	03/20/2020	2361	57
	Mato Grosso do Sul	MS	2778986	03/16/2020	1418	19
	São Paulo	SP	45919049	02/26/2020	107142	7532
	Rio de Janeiro	RJ	17264943	03/05/2020	52420	5277
South	Espírito Santo	ES	4018650	03/06/2020	13437	583
	Minas Gerais	MG	21168791	03/08/2020	9630	263
	Rio Grande do Sul	RS	11377239	03/10/2020	9242	218
South	Santa Catarina	SC	7164788	03/13/2020	8778	136
	Paraná	PR	11433957	03/12/2020	4473	181
<b>Brazil</b>			<b>210147125</b>	<b>02/26/2020</b>	<b>498440</b>	<b>28834</b>

the Brazilian federative units regarding the COVID-19 outbreak. Consider a joint sample  $(\mathbf{X}, \mathbf{Y}) = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)\}$  defined by  $i, j = 1, \dots, N$ , with  $X \in \mathbb{R}^l$ ,  $Y \in \mathbb{R}^m$  and  $N \geq 2$ . Thus, the *empirical DC coefficient* for a joint sample  $(\mathbf{X}, \mathbf{Y})$  is given by

$$\mathcal{C}_N(\mathbf{X}, \mathbf{Y}) = \frac{\sigma_N(\mathbf{X}, \mathbf{Y})}{\sqrt{\sigma_N(\mathbf{X})}\sqrt{\sigma_N(\mathbf{Y})}}. \quad (4)$$

The *empirical distance covariance* for  $(\mathbf{X}, \mathbf{Y})$  is given by

$$\sigma_N^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{N^2} \sum_{i,j=1}^N A_{ij} B_{ij}. \quad (5)$$

To calculate numerically the distance covariance by using Eq. (5) is much simpler than by using Eq. (1), although both equations describe the same quantity. For details about the equivalence between such definitions, we suggest the Ref. [23].

For a sample  $\mathbf{X}$ , the *empirical distance variance* is given by

$$\sigma_N^2(\mathbf{X}) = \frac{1}{N^2} \sum_{i,j=1}^N A_{ij}^2, \quad (6)$$

and for a sample  $\mathbf{Y}$ ,

$$\sigma_N^2(\mathbf{Y}) = \frac{1}{N^2} \sum_{i,j=1}^N B_{ij}^2. \quad (7)$$

The matrix  $A_{ij}$  is obtained from  $\mathbf{X}$ :

$$A_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \quad (8)$$

where  $a_{ij} = |X_i - X_j|_l$  is the Euclidean norm of the distance between the pairs of elements of the sample,  $\bar{a}_{i.} = \frac{1}{N} \sum_{j=1}^N a_{ij}$  is the arithmetic mean of the rows,  $\bar{a}_{.j} = \frac{1}{N} \sum_{i=1}^N a_{ij}$  is the arithmetic mean of the columns, and  $\bar{a}_{..} = \frac{1}{N^2} \sum_{i,j=1}^N a_{ij}$  is the general mean. Similarly, the matrix  $B_{ij}$  is obtained from  $\mathbf{Y}$ .

It is important to mention that, to obtain the DC by applying the procedure described above, we need only the samples  $\mathbf{X}$  and  $\mathbf{Y}$ . This can be quite relevant when it comes to experimental data. In Eq. (4) it is possible to verify that  $\mathcal{C}_N(\mathbf{X}, \mathbf{Y})$  is independent of the scale since the samples  $\mathbf{X}$  and  $\mathbf{Y}$  can be multiplied by a real number and  $\mathcal{C}_N(\mathbf{X}, \mathbf{Y})$  remains unchanged. This is an important feature of the DC coefficient which allows us to obtain relevant information even if data with different magnitudes are compared. Furthermore, it is important to mention that, if all elements of a sample

$\mathbf{X}$  or  $\mathbf{Y}$  are identical, then we have  $\sigma_N(\mathbf{X})\sigma_N(\mathbf{Y}) = 0$  and therefore  $C_N(\mathbf{X}, \mathbf{Y}) = 0$  as shown in [23].

### 2.3. The SIRD model

Since the pioneering work of Kermack and McKendrick [31], mathematical models have been developed to describe infectious disease dynamics and to help us to analyze the efficiency of preventive measures that could be adopted to control infections spread. Containment strategies such as social distance, which includes mask-wearing, quarantine, and other actions, are essential to avoid the uncontrolled increase in the number of infected people and the collapse of the health care systems.

In this work, we used the well-known SIRD model to reproduce the realistic data, to estimate some important epidemiological parameters, and to study the effects of containment measures around the different regions of Brazil. In this model, the total population  $N$  is divided into subgroups of Susceptibles ( $S$ ), Infected ( $I$ ), Recovered ( $R$ ), and Dead ( $D$ ) people. For all time  $t$ ,  $N = S + I + R + D$ . The ODEs that describe the dynamics of such groups are the following [31, 32]:

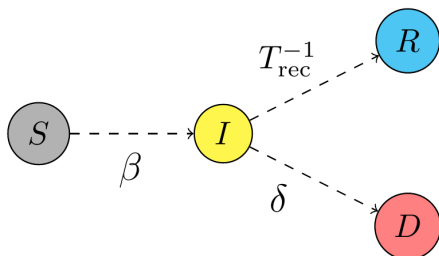
$$\frac{dS}{dt} = -\beta \frac{SI}{N}, \quad (9)$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \left( \frac{1}{T_{\text{rec}}} + \delta \right) I, \quad (10)$$

$$\frac{dR}{dt} = \frac{1}{T_{\text{rec}}} I, \quad (11)$$

$$\frac{dD}{dt} = \delta I, \quad (12)$$

where  $\beta$ ,  $T_{\text{rec}}$ , and  $\delta$  are ‘‘effective’’ parameters, once they are based on the reported confirmed cases and deaths. Fig. 1 shows a schematic representation of the SIRD model with the relationship between variables and parameters. The parameter  $\beta$  is the infection rate, which reflects the probability per unit time that a susceptible individual becomes infectious when entering in contact with an infected person. The term  $T_{\text{rec}}^{-1}$  is the recovery rate, with  $T_{\text{rec}}$  being the



**Figure 1.** Schematic representation of the SIRD model highlighting its variables and the role of the parameters connecting the variables.

infection duration or the recovery time. Finally,  $\delta$  is the death rate, which means the portion of people that die per unit time in comparison with the number of active infected people. The initial conditions are  $[S(t_0), I(t_0), R(t_0), D(t_0)]$ , and the population at each time is given by  $N(t) = N(t_0) - D(t)$ , being  $N(t_0)$  the initial population described in Table 1. To update the population is important for future projections, once the number of deaths  $D(t)$  could be significant when compared to  $N(t)$ . In our simulations, we set  $R(t_0) = 0$  and  $I(t_0), D(t_0)$  were obtained from the real time-series. We considered that the initial number of susceptible people is given by  $S(t_0) = N(t_0) - I(t_0) - R(t_0) - D(t_0)$ . In addition to these equations, we compute the cumulative number of confirmed cases  $C(t)$  of COVID-19 by the relation  $C(t) = I(t) + R(t) + D(t)$ .

The values of  $\beta$ ,  $T_{\text{rec}}$ , and  $\delta$  depend on several circumstances and must differ significantly in different regions and different periods of the pandemic. Besides, to obtain reliable values for these parameters in each locality, we propose to adjust the parametric set  $(\beta, T_{\text{rec}}, \delta)$  using the evolution of the real data. To start the fitting, first of all, we divided the time series in epidemiological weeks (EPI weeks), which are, by international convention, count from Sunday to the next Saturday. With this split, we intend to find the best combination of  $\beta(k)$ ,  $T_{\text{rec}}(k)$ , and  $\delta(k)$  that describes the evolution of the pandemic in each EPI week number  $k$ . Moreover, the number of data inside each EPI week satisfies the minimal number of  $2r + 1$  experiments with real data needed to adjust trusty  $r$  unknown parameters [33] (in our case,  $r = 3$ ).

To define the parametric set in each EPI week, we performed simulations for all possible combinations obtained by varying  $\beta \in [0.0, 0.5]$  with a step of  $10^{-2}$ ,  $T_{\text{rec}} \in [10, 21]$  with a step of  $10^{-1}$ , and  $\delta \in [10^{-4}, 3 \times 10^{-2}]$  with a step of  $10^{-4}$ . For each set  $(\beta, T_{\text{rec}}, \delta)$ , we (i) calculated the Root Mean Square Error (RMSE)  $\Delta_C$  between the cumulative number of infected people  $C(t)$  obtained by the model in each day and the real data; and (ii) calculated the RMSE  $\Delta_D$  between the number of deaths  $D(t)$  obtained by the model in each day and the real data. The best parametric set is that for which the quantity  $\sqrt{\Delta_C^2 + \Delta_D^2}$  is minimized. We picked up the range 10-21 days for the recovery time because it covers a reliable estimate around the world [3, 7, 34]. On the other hand, in Brazil, a variation from 0.01% to 3% for the daily death rate  $\delta$  is a satisfactory range [30, 35]. To obtain better results, the first day considered on the fitting is the next Sunday after the first death in each time series analyzed. For example, the first death in Brazil was confirmed on March 17<sup>th</sup>, the Tuesday belonging to the EPI week number 12. Then, we started adjusting the parametric set from

March 22<sup>nd</sup>, the beginning of the EPI week number 13.

### 3. Results

#### 3.1. Power-law grows

Figure 2 displays the data of the cumulative number of infected people by COVID-19 of ten representative federative units and of Brazil as a function of the days, using the linear scale. The analyzed federative units were chosen to represent all the five regions that compose the Brazilian territory (see Table 1 and the caption of Fig. 2). Each block that composes the Table 1 represents one Brazilian region, where the federative units are listed in decreasing order of the number of infected people. In Fig. 2, the black and brown continuous curves are the corresponding fitting curves  $\alpha + \gamma t^\mu$ , where  $t$  is the time given in days after the reported hundredth case,  $\alpha$ ,  $\gamma$  and  $\mu$  are parameters. We choose to discard the initial data, regarding the days with less than 100 infected individuals, since in this way it was possible to capture the real tendency of the fitted curves and avoid misunderstandings due to poor statistics. The straight lines in each inset, plotted in logarithm scale, indicate power-law growth as previously observed for different countries in four continents [11, 12].

It is fundamental to understand the stage of the pandemic to take the right decision. This knowledge may be obtained through the contamination's curves since they provide essential information about what is expected in the future and about what can be done to flatten the curves. Such information is encoded in the exponent  $\mu$  that changes for distinct federative units. The complete fitting parameters are given in Table 2. Results in Table 2 are presented in the same order as plotted in Fig. 2. Through this figure, it is clear that the cumulative number of confirmed cases is still growing according to power-laws with different exponents  $\mu$  in the most federative units as in the country. These differences may be caused by several factors as, for example, due to the size of the federative units, the heterogeneous concentrations of poor people, or the complex social and political organizations of them, that have applied different strategies in different stages of the contagious process. Nowadays, the administrative organization of Brazil gives to the federative units and municipalities some independence to adopt mitigation approaches, but these are limited by economic dependence on superior spheres. So, the tendency is to have similar approaches within federative units, although someones have obtained much better results than others. Interestingly, while PA and FD have exponents  $\mu$  larger than Brazil, the other federative units have smaller exponents.

**Table 2.** Details about the parameters of the fitting curves for the power-law behavior  $\alpha + \gamma t^\mu$  shown in Fig. 2.

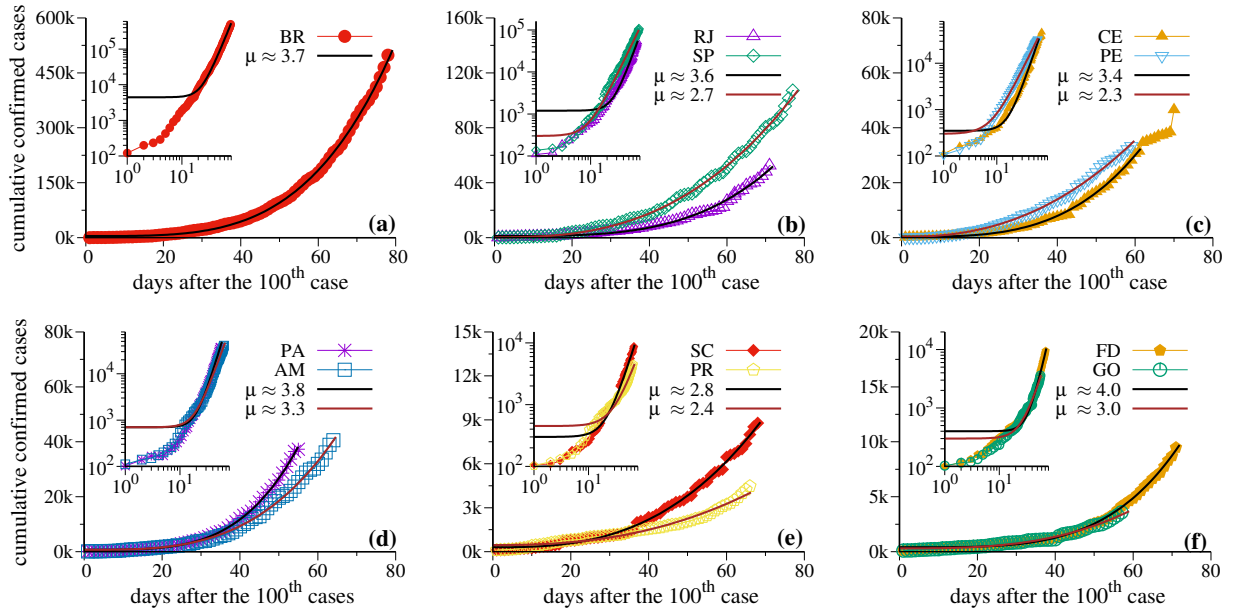
Locality	$\alpha$	$\gamma$	$\mu$
Brazil	4450	0.040	$3.743 \pm 0.022$
RJ	1200	0.011	$3.600 \pm 0.064$
SP	300	0.052	$2.699 \pm 0.022$
CE	350	0.008	$3.367 \pm 0.037$
PE	300	2.428	$2.340 \pm 0.023$
PA	700	0.008	$3.849 \pm 0.032$
AM	700	0.052	$3.259 \pm 0.033$
SC	300	0.066	$2.783 \pm 0.035$
PR	450	0.129	$2.441 \pm 0.080$
FD	400	$2.892 \times 10^{-4}$	$4.044 \pm 0.035$
GO	300	0.019	$2.971 \pm 0.068$

Additionally, SP and Brazil present quite similar  $\mu$  values, where  $\mu \sim 3.6$ . This is not a coincidence, as discussed in detail in the next subsection.

The most desired behavior is that the exponent  $\mu$  becomes smaller, leading to the flattening of the curves. Although, to achieve this scenario is not that easy. Besides CE, which seems to be stabilizing the epidemic spread, in all other federative units the growth remains strictly on the power-law fitting curve and  $\mu$  essentially does not change in time. In Subsection 3.4.2, we discuss some possibilities to flatten the power-laws.

#### 3.2. The DC between Brazil and its federative units

Figure 3 presents the DC coefficient  $\mathcal{C}_N(\mathbf{X}, \mathbf{Y})$  (plotted in colors) computed between the real time-series of Brazil (sample  $\{\mathbf{X}\}$ ) and each one of its 27 federative units (sample  $\{\mathbf{Y}\}$ ), considering the data for the cumulative number of cases in Fig. 3(a) and the number of deaths in Fig. 3(b). For simplicity, we now will use only the notation  $\mathcal{C}$  for the DC coefficient. As mentioned in Subsection 2.2, the DC is a measure that analyzes how similar the shapes of the curves are to each other independently of the magnitude of such data. With this statistical analysis, we intend to find the federative units for which the time series is more correlated with the data of Brazil. Since  $\mathcal{C}$  measures the “similarity” of such curves, with this information it is possible to know the states that are still following a power-law growth together with the country or are flattening the curve. The different colors in Fig. 3 correspond to distinct values for  $\mathcal{C}$  between the real time-series. According to the palette, the black region corresponds to the absence of data (ND), *i.e.*, without reported infected people and deaths for the federative units, except for SP. The white and the gray regions mean a low value for DC, while the red, yellow, cyan and navy blue colors are related to high correlations.



**Figure 2.** The cumulative number of confirmed cases of COVID-19 as a function of time for (a) Brazil, (b) SP and RJ, (c) CE and PE, (d) AM and PA, (e) SC and PR, (f) FD and GO. The same quantities are also plotted in logarithm scale in the insets. The black and brown continuous curves represent the function  $\alpha + \gamma t^\mu$  that fit the time-series, and the parameters  $\alpha$ ,  $\gamma$ , and  $\mu$  for each locality are described in Table 2.

In Fig. 3(a), all the federative units are shown from the left to the right in alphabetical order. On the vertical axis, the date is displayed since February 26<sup>th</sup>, the day when the first case of COVID-19 was reported in SP. In this figure, we can see that the time series of SP is highly correlated with the time series of Brazil since the first reported case, which is indicated by the long vertical navy blue stripe ( $C \sim 1$ ). This result was expected since the power-law growth of the cumulative number of infected people for both Brazil and SP shows a very similar exponent, with  $\mu \sim 3.6$  (see Table 2 and Fig. 2). Similarly, we observe that the time series of BA, CE, and RJ were also highly correlated in most of the time with the data of Brazil, which means that such federative units also showed a power-law increasing of the cumulative number of infected people, how can be observed in Fig. 2.

On the other hand, it is interesting to note that for PR, in the period from April 22<sup>nd</sup> to May 20<sup>th</sup>, the coefficient  $C$  suffers a slight decay, remaining around 0.95. This value reflects the oscillations in the time series of PR, which can be seen in Fig. 2(e). SE and TO, two federative units with a relatively low incidence of infected people, have presented an evident similarity in the behavior of the DC coefficient in time. For both states, the value of  $C$  presented oscillations around 0.90 and 0.98 during the period considered.

Another important achievement of the DC coefficient is to identify a possible change in the trend of one time series. Fig. 2(c) shows that CE is apparently on the way of flattening the curve, which

no longer follows a power-law growth. Since this new behavior is recent, the change in the value of  $C$  is slight and can not be depicted by the color scale in Fig. 3. For this reason, we plotted the time series of Brazil and CE in Fig. 4(a) and the value of  $C$  calculated between such data in Fig. 4(b). In this figure, it is possible to see that, while both time series follow a power-law growth, the DC coefficient increases until 1. However, once that CE starts flattening its curve and Brazil continues growing up,  $C$  starts to decrease.

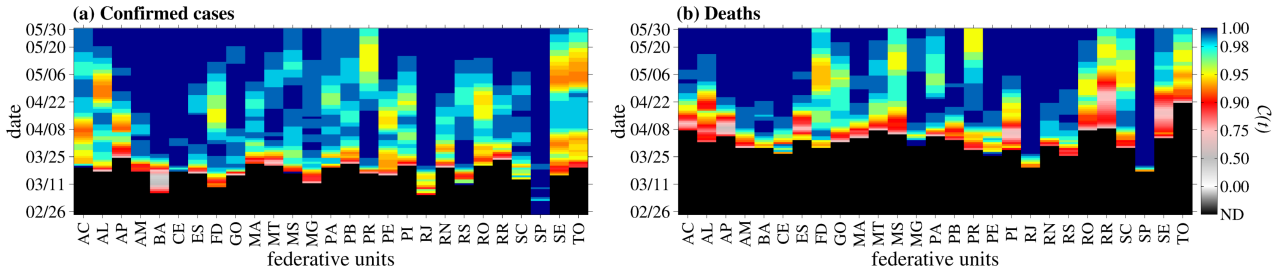
In Fig. 3(b) we plotted the value of  $C$  for the real time-series of the number of deaths caused by COVID-19. In this case, it is remarkable that the times series of SP and RJ, the two federative units with the greatest numbers of deaths, are highly correlated with the data of Brazil in most of the time. Besides that, it is also possible to note that AM, BA, and CE reached high values of  $C$  from April 22<sup>nd</sup>.

In general, the oscillations presented by  $C$ , which are consequences of the different behaviors of the data along the time, testify the heterogeneity of the COVID-19 spreading in the Brazilian territory. Such differences might be related to some reasons as, for example, the different containment measures adopted in each region, the lack of testing, underreported data, and others.

### 3.3. An estimate of the effective reproduction number

As explained in Section 1, the basic reproduction number  $\mathcal{R}_0$  is the average number of secondary infections originated from an infectious person considering a pop-

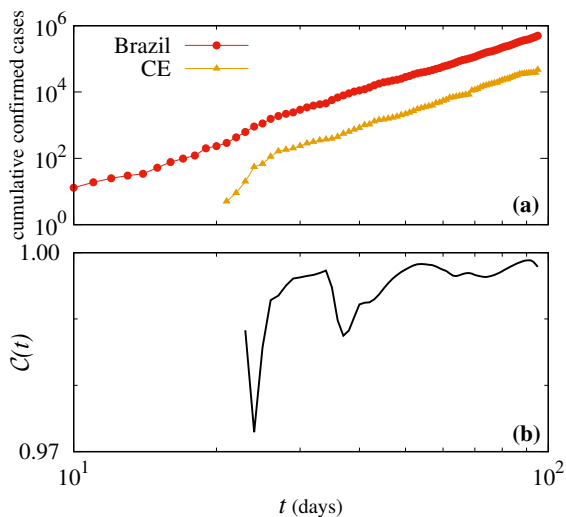




**Figure 3.** DC coefficient ( $C$ ) plotted in colors for the cumulative number of (a) confirmed cases and (b) deaths. Analyzing these panels, it is possible to see that BA, CE, RJ, and SP showed a behavior similar to Brazil in most of the time, which is indicated by the long vertical navy blue stripes. On the other hand, the value of  $C$  for states that show a different behavior is lower (about 0.95), as is possible to see for PR, SE, and TO, for instance.

ulation where everyone is susceptible. However, this condition is strictly valid only when the pathogen is in the early days. With the evolution of time, the susceptible fraction of the population decreases, and the fundamental epidemiological quantity that determines the dynamical evolution of the pandemic is the *replacement number* or the *effective reproduction number*  $\mathcal{R}_e(t)$ , which is defined as the average number of newly infected individuals produced by a typical infective during the period of infectiousness [18]. This quantity captures changes in the pandemic evolution and the key aim of the government containment measures is to reduce  $\mathcal{R}_e(t)$ . If  $\mathcal{R}_e(t) < 1$ , the incidence of new infections decreases, and the epidemic is controlled. On the other hand, if  $\mathcal{R}_e(t) > 1$ , the number of newly infected people will continue to grow until the peak and, eventually, can decline due to the acquisition of immunity.

The effective reproduction number  $\mathcal{R}_e(t)$  is a



**Figure 4.** (a) The time series of the cumulative number of confirmed cases of Brazil and CE and (b) the DC coefficient ( $C$ ) calculated between these data. On the horizontal axis, the days are count since the first case in Brazil.

dynamical quantity, and by using the parameters obtained by the procedure described in Subsection 2.3 we can estimate the value of  $\mathcal{R}_e(k)$  for each EPI week  $k$  using the relation

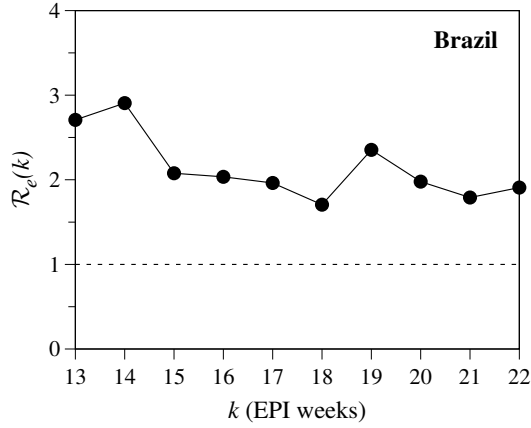
$$\mathcal{R}_e(k) = \frac{\beta(k)}{1/T_{\text{rec}}(k) + \delta(k)}, \quad (13)$$

where  $\beta(k)$ ,  $T_{\text{rec}}(k)$ , and  $\delta(k)$  is the best parametric combination for each EPI week. The effective reproduction number calculated for the data of Brazil is shown in Fig. 5.

The first case of COVID-19 in Brazil was confirmed during the EPI week 9, and the value of the basic reproduction number  $\mathcal{R}_0$  is estimated in the range of 3-4, in line with estimates of transmissibility derived from Europe [35]. We can observe in Fig. 5 that  $\mathcal{R}_e(13) = 2.7$  and  $\mathcal{R}_e(14) = 2.9$ , results obtained after some federative units had adopted the first containment measures during the EPI week 12. After these weeks, the effective reproduction number continued decreasing, reaching the value 1.7 in the EPI week 18. In early May, a relaxing on the social distance measures was observed in Brazil and the number of newly infected cases exploded. Our simulations demonstrate this increase in the value of  $\mathcal{R}_e(k)$  in the EPI week number 19, when  $\mathcal{R}_e > 2$  again. After this increase, the value of  $\mathcal{R}_e$  remained below 2 in the last EPI weeks.

It is important to mention that there are other more precise methods to estimate the value of  $\mathcal{R}_e$  (see, for example, the Refs. [35, 36, 37]). However, even though the SIRD model is rather crude, our simulations provide reliable values for the reproduction numbers and, more importantly, capture trends for such quantity. Besides the mean value of  $\mathcal{R}_e(k)$ , we display in Table 3 the mean values for the infection rate  $\beta(k)$ , recovery time  $T_{\text{rec}}(k)$ , and death rate  $\delta(k)$  for Brazil and for the eight federative units with the higher incidence of infected people, which are, in decreasing order: SP, RJ, CE, AM, PA, PE, MA, and BA. These mean values were calculated over all EPI weeks considered in our analysis. The value of  $\mathcal{R}_e(22)$





**Figure 5.** The effective reproduction number  $\mathcal{R}_e$  estimated for each EPI week  $k$  for Brazil.

obtained in the EPI week number 22 is also displayed since such quantity will be used in the next subsection to analyze the trends for  $C(t)$  and  $D(t)$  and to simulate effects of increase or decrease the effective reproduction number on the pandemic evolution in Brazil and its federative units.

### 3.4. Simulating the effects of increase or decrease the effective reproduction number

In this section, we describe the results obtained by using the SIRD model to fit the real data and to project some scenarios. First, we analyzed the trends for the next days assuming that the effective reproduction number  $\mathcal{R}_e(22)$  of the EPI week number 22 remains unchanged. Thereon, we study the effects of changing the effective reproduction number on quantities as  $C(t)$ ,  $D(t)$ , and  $I(t)$ . Such variables are certainly affected by the recovery rate  $T_{\text{rec}}$ , and also by the death rate  $\delta$ . However, since these parameters are mostly changed by pharmacological interventions, in

**Table 3.** Table of mean values of the epidemiological parameters obtained by fitting the real time-series. The values of  $\langle\beta\rangle$  and  $\langle\delta\rangle$  are given in  $\text{days}^{-1}$ ,  $\langle T_{\text{rec}}\rangle$  is given in days, and  $\mathcal{R}_e$  is a dimensionless parameter.

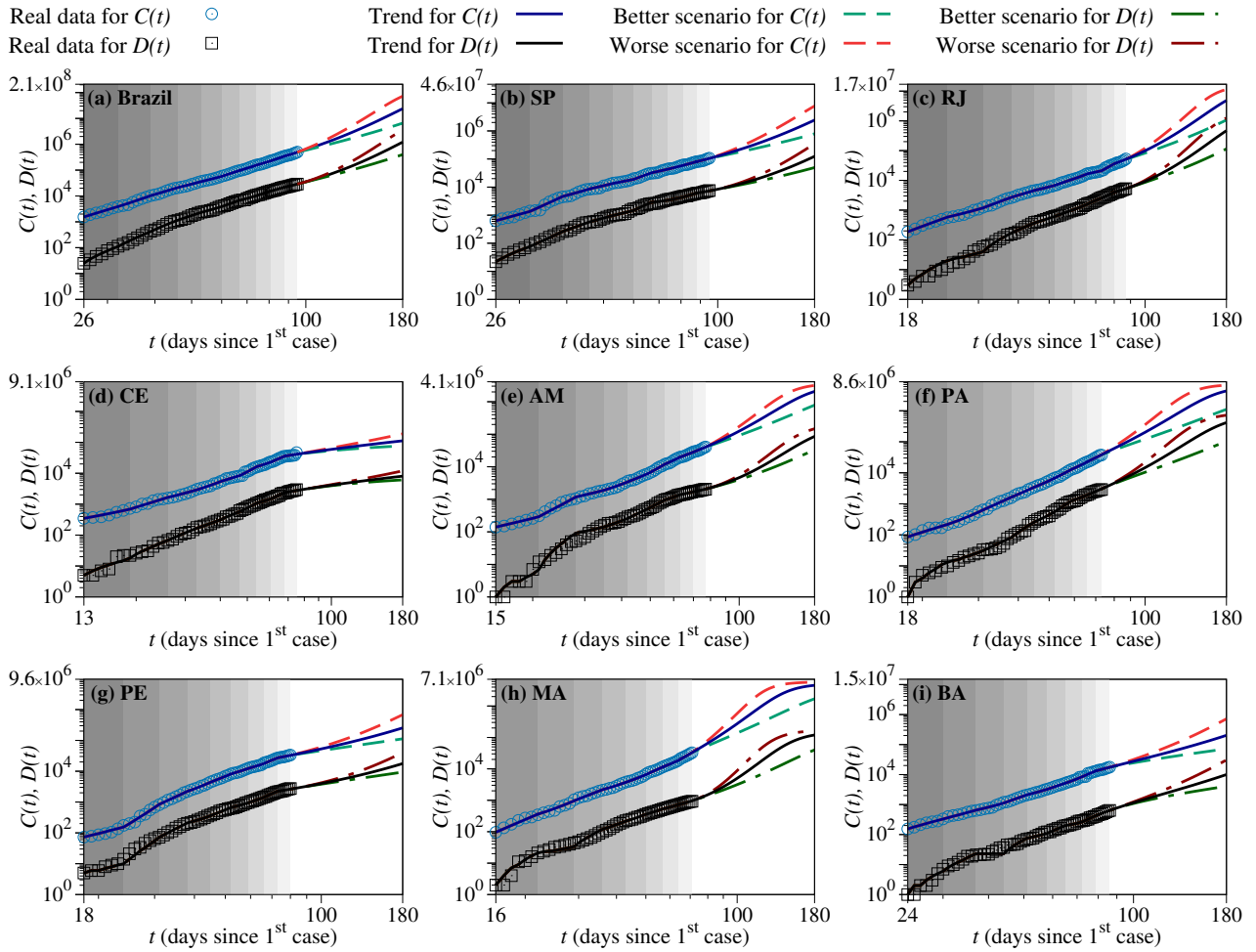
Locality	$\langle\beta\rangle$	$\langle T_{\text{rec}}\rangle$	$\langle\delta\rangle$	$\langle\mathcal{R}_e\rangle$	$\mathcal{R}_e(22)$
Brazil	0.14	18.0	0.83%	2.14	1.91
SP	0.13	18.5	0.95%	1.93	1.73
RJ	0.16	15.0	1.40%	1.84	1.86
CE	0.14	16.1	0.90%	1.84	0.97
AM	0.15	17.4	1.06%	2.13	1.97
PA	0.18	15.2	1.41%	2.24	1.67
PE	0.16	16.9	1.33%	2.07	1.34
MA	0.18	14.8	0.82%	2.27	1.73
BA	0.14	14.7	0.49%	1.91	1.31

our simulations they were kept fixed and we increase or decrease the value of  $\beta(22)$ , which is the quantity that reflects non-pharmacological containment measures as, for example, the social distance actions. With these simulations, we do not intend to preview an exact number of infected or dead people. Actually, we know that the epidemiological parameters are dynamical and change recurrently, which turns long-term projections a tough task. The results presented in the next subsections aim to demonstrate as a simple mathematical model can be used to describe the real data and to convince about the relevance of taking preventive measures as soon as possible to contain the virus spreading and to save lives.

#### 3.4.1. The cumulative number of cases and deaths:

Figure 6 shows the real times-series for the cumulative number of confirmed cases of COVID-19 (blue circles) and deaths (black squares) for Brazil and for the eight federative units most affected by the pandemic. In this Figure, we also display the curves of  $C(t)$  and  $D(t)$  obtained by integrating the equations of the model, represented by the dark blue and black continuous lines, respectively. We observe that, by adjusting the parameters  $\beta$ ,  $T_{\text{rec}}$ , and  $\delta$  in each EPI week (represented by the gray rectangles in the background), it is possible to describe precisely the behavior of the real data for all localities analyzed. After to find the best parameter combination that fits the real data for each EPI week, the dark blue and black continuous curves describe the trend for  $C(t)$  and  $D(t)$ , respectively, assuming that the values  $\beta(22)$ ,  $T_{\text{rec}}(22)$ , and  $\delta(22)$  of the EPI week number 22 will remain unchanged until the day 180 after the first case. Therefore, by Eq. (13), this assumption means to keep fixed the same effective reproduction number obtained in the EPI week 22, value that is shown in Table 3 for the localities studied.

In Figs. 6(a), 6(b), and 6(c), we see the trend for Brazil, SP, and RJ, in this order. SP and RJ are the most populous federative units of Brazil and together they cover 32% of the total number of confirmed cases of the country. In these three cases, the effective reproduction number  $\mathcal{R}_e(22)$  is greater than 1.7, and it is possible to note that the trend for  $C(t)$  and  $D(t)$  is to keep increasing for several weeks. Besides these cases, we also highlight the examples of AM [Fig. 6(e)], PA [Fig. 6(f)], and MA [Fig. 6(h)], for which  $\mathcal{R}_e(22) > 1.6$ . Our simulations show that, if the effective reproduction number does not change, about half of the initial population  $N(t_0)$  of such federative units can be infected until the day 180 after the first case. It is possible for these three states since their population is considerably smaller than SP and RJ, for example (see Table 1). Besides that, according to



**Figure 6.** The cumulative number  $C(t)$  of confirmed cases of COVID-19 (blue circles) and deaths  $D(t)$  (black squares) for (a) Brazil, (b) SP, (c) RJ, (d) CE, (e) AM, (f) PA, (g) PE, (h) MA, and (i) BA. The dark blue and black continuous curves show the trends for  $C(t)$  and  $D(t)$ , respectively, assuming that the effective reproduction number of the EPI week number 22 is kept fixed. A better scenario means to reduce in 20% the  $\mathcal{R}_e(22)$ , and this case is represented for  $C(t)$  by the light green dashed lines and for  $D(t)$  by the dark green dash-dotted lines. A worse scenario means to increase in 20% the  $\mathcal{R}_e(22)$ , and this case is represented for  $C(t)$  by the light red dashed lines and for  $D(t)$  by the dark red dash-dotted lines.

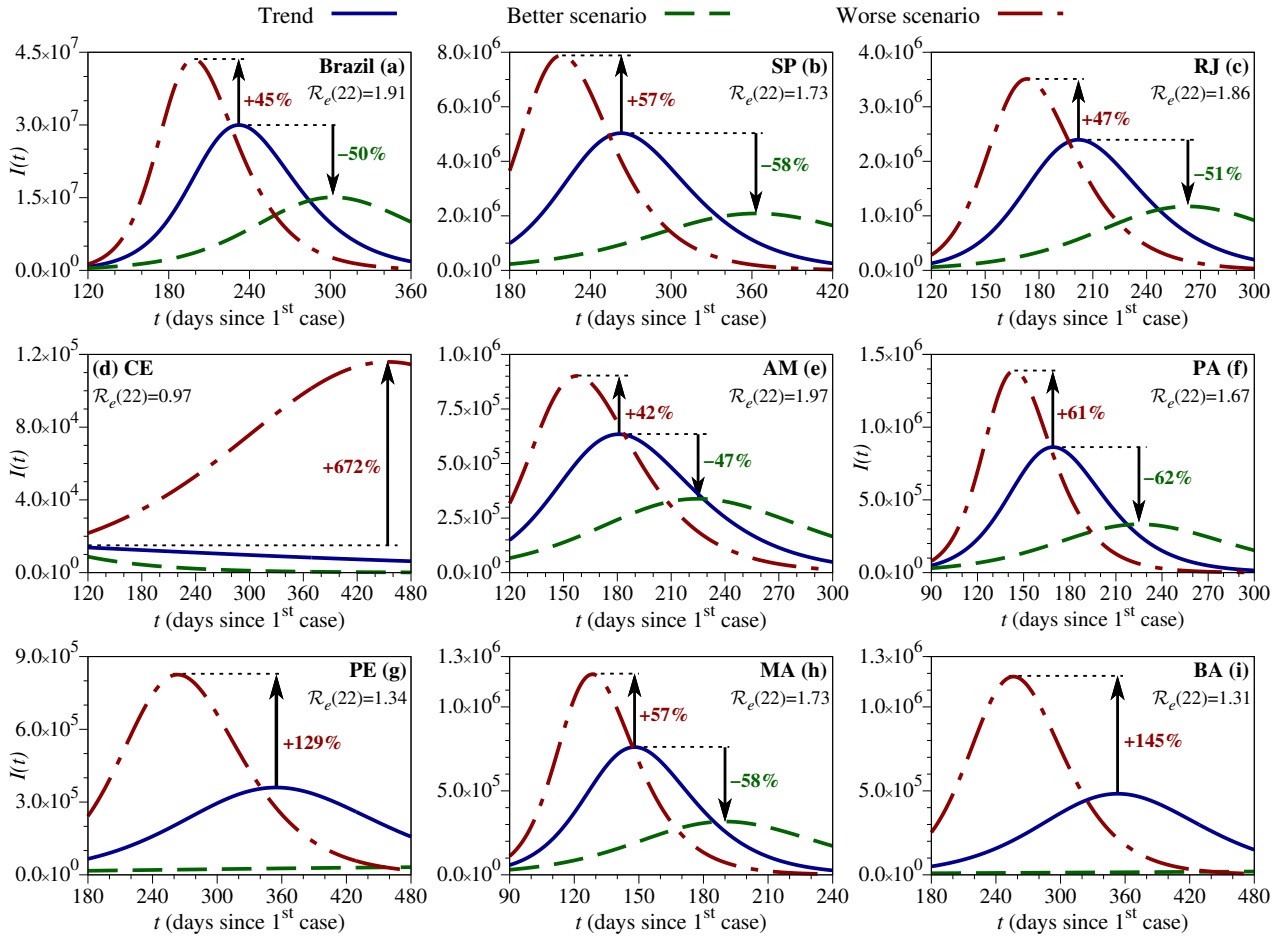
Table 3, these three federative units show the worst values of  $\langle \mathcal{R}_e \rangle$  over the EPI weeks considered.

Differently from the cases discussed above, PE [Fig. 6(g)] and BA [Fig. 6(i)] show a more attenuated increase for  $C(t)$  and  $D(t)$ , which reflects the fact that their effective reproduction number in the EPI week 22 is about 1.3. The best result was obtained for CE, the case displayed in Fig. 6(d), for which  $\mathcal{R}_e(22) < 1$ . In this case, the growth of the curves  $C(t)$  and  $D(t)$  is very slow, which means that the pandemic is under control if this scenario is kept.

Since the value of  $\mathcal{R}_e$  depends on several events and can change daily, then we project some scenarios considering the increasing and the decreasing of such quantity. For this, we kept fixed  $T_{\text{rec}}(22)$  and  $\delta(22)$ , and changed the value of  $\beta(22)$ , summing and reducing 20% of its value, which means to increase or decrease in 20% the effective reproduction number. A better

scenario was simulated by decreasing  $\mathcal{R}_e(22)$  in 20%, and this case for  $C(t)$  is represented by the light green dashed curves, and for  $D(t)$  by the dark green dash-dotted curves. In all cases, to keep the effective reproduction number 20% below the current one can save thousands of people in the long term. On the other hand, a worse scenario takes place when increasing in 20% the current value of  $\mathcal{R}_e$ , which is represented for  $C(t)$  by the light red dashed curves, and for  $D(t)$  by the dark red dash-dotted curves. This scenario increases both the number of confirmed cases and deaths for all localities.

**3.4.2. Flattening the curve of active infected individuals:** Now, we will describe the results obtained by analyzing the number of active infected people  $I(t)$ . This analysis is important since a huge number of simultaneously infected people certainly lead the health care sys-



**Figure 7.** Projections for the number of active infected individuals  $I(t)$  obtained by keeping the value of  $\mathcal{R}_e(22)$  (dark blue continuous curves), by decreasing in 20% the value of  $\mathcal{R}_e(22)$  (dark green dashed curves), and by increasing in 20% the value of  $\mathcal{R}_e(22)$  (dark red dash-dotted curves). The simulations were performed for (a) Brazil, (b) SP, (c) RJ, (d) CE, (e) AM, (f) PA, (g) PE, (h) MA, and (i) BA. The arrows and the percentage indicate the increase or decrease in the number of active infected people  $I(t)$  at the peak of each case when compared with the current trend.

tems to collapse. Figure 7 shows three curves for each locality: (i) the dark blue continuous curve describes the trend, obtained by keeping the value  $\mathcal{R}_e(22)$  for the next days; (ii) the dark green dashed line describes a better scenario, obtained by decreasing  $\mathcal{R}_e(22)$  in 20%; and (iii) the dark red dash-dotted line describes a worse scenario, obtained by increasing  $\mathcal{R}_e(22)$  in 20%. Analyzing Fig. 7, we observe that increasing the value of the effective reproduction number in 20%, the peak of  $I(t)$  not only grows but also occurs earlier. On the other hand, decreasing  $\mathcal{R}_e(22)$ , the desired flattening of the curves of infected people is reached, which means to shrink the peak and to delay its occurrence.

In Figs. 7(a) (Brazil), 7(b) (SP), 7(c) (RJ), 7(e) (AM), 7(f) (PA), and 7(h) (MA), localities for which  $\mathcal{R}_e > 1.6$ , we observe an enhancing in the peak of  $I(t)$  of at least 42% when increasing the effective reproduction number in 20%. Besides that, for PE and BA, for which  $\mathcal{R}_e(22) \sim 1.3$ , to increase in 20% such quantity can generate a peak with  $I(t)$  at least 120%

greater than the current trend. Another interesting case is CE, for which  $\mathcal{R}_e(22) = 0.97$ . For this federative unit, since the effective reproduction number is smaller than 1, the peak was already reached, and the curve  $I(t)$  is in a descendent way [see the blue continuous curve in Fig. 7(d)]. If  $\mathcal{R}_e(22)$  increases 20%, from 0.97 to 1.16, a new peak occurs and it can be 672% bigger than the first peak. This result demonstrates the relevance of keeping the effective reproduction number below 1. For all cases, decreasing in 20% the value of  $\mathcal{R}_e(22)$  generates a satisfactory reduction of the peak, at least 47%.

#### 4. Summary and conclusions

The present investigation has an apparent limitation, being performed with only data from Brazil (and its federative units), which is one country in the whole world. However, the results obtained from modeling the SARS-CoV-2 epidemic in large countries

like Brazil have interesting features that demand to be scrutinized due to the wide territorial, demographic, and infrastructural diversity. Brazil has more than 210 million citizens distributed in 27 federative units which belong to five large regions with completely different cultural and social habits, and economic situations. By scrutinizing the heterogeneous spreading of the COVID-19 outbreak in Brazilian territory, we can understand the epidemic current status in each federative unit and project the evolution of the outbreak in the country. This is important because each federative unit has been dealt with the epidemic in different ways when it comes to containment strategies.

The cumulative number of confirmed cases of COVID-19 until May 30<sup>th</sup>, 2020, is demonstrated for Brazil and its 27 federative units (see Table 1). In Fig. 2, for 10 representative states we observe that, after an initial period with a low incidence of newly infected people, a power-law ( $\alpha + \gamma t^\mu$ ) growth of such time series takes place. For each federative unit, we estimated a distinct growth exponent. Brazil is still in the power-law regime, as the federative units studied in Fig. 2, except CE which is achieving a plateau. FD, PA, RJ, and AM have the largest power-law exponents ( $\mu > 3.0$ ). Such projections are in complete agreement with the recent results regarding the power-law growth of the cumulative number of infected individuals by the SARS-CoV-2 around the world [11, 12].

Although the power-law growths with distinct exponents may look similar visually, it is interesting to quantify this similarity. This is performed through the computation of the Distance Correlation (DC) between the data of Brazil and each one of its federative units listed in Table 1, as plotted in colors in Fig. 3. In this figure, we show that the power-law growth for Brazil and for some federative units like SP, RJ, CE, and BA are strongly correlated, which is characterized by  $\mathcal{C} \sim 1$ . For other states as, for example, PR, SE, and TO, in most of the time the value of  $\mathcal{C} \sim 0.95$ , which lead us to conclude that such localities showed a different behavior during the pandemic evolution. In fact, these federative units were not so affected by the pandemic.

By using the SIRD model, it was possible to estimate the epidemiological parameters that best fit the real data in each epidemiological week  $k$  (EPI week): the infection rate  $\beta(k)$ , the recovery time  $T_{\text{rec}}(k)$ , and the death rate  $\delta(k)$ . Once these quantities are related by Eq. (13), we estimated the *effective reproduction number*  $\mathcal{R}_e(k)$  for each EPI week and projected some scenarios based on these values. Our findings suggest that the mean value of  $\mathcal{R}_e$  for the eight most affected federative units in Brazil is about 2, as shown in Table 3. Besides, our simulations showed that the trend for localities with  $\mathcal{R}_e \sim 2$  is to keep following

a power-law growth, while federative units with  $\mathcal{R}_e \sim 1$  are getting the flattening of the curve. Moreover, CE is nowadays the only federative unit for which  $\mathcal{R}_e < 1$ , and its curve of active infected people  $I(t)$  is in a descendent way [see Fig. 7(d)]. We also observed that, by increasing the value of the effective reproduction number in 20%, the peak of  $I(t)$  not only might be at least 40% greater but it also might occur earlier. Our results show that the only way to flatten the curve is to decrease  $\mathcal{R}_e$ . In further investigations, we intend to improve our findings considering stochastic perturbations to study how the uncertainty of the official data (due to several reasons) might change such projections. Beyond that, the transmissibility of asymptomatic individuals and the vulnerability conditions of some piece of the total populations also may be considered, mainly in developing countries like Brazil, and in least developed countries. Additionally, the airborne COVID-19 transmission also might be achieved in future studies.

### Acknowledgments

C.M. thanks CNPq (Brazil) for financial support (Grant Nos. 304918/2017-2, and 424803/2018-6). The authors also thank MSc. E.L. Brugnago and Dr. M.W. Beims for the fruitful discussions about the model implementation.

### Author contributions

R.M.S., C.F.O.M. and C.M. contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

### Competing interests

The authors declare no competing interests.

### Availability of materials and data

The data used in this study are openly available in Ref. [30].

### ORCID iDs

R.M. da Silva: <http://orcid.org/0000-0003-1372-8132>  
 C.F.O. Mendes: <http://orcid.org/0000-0002-3144-9287>  
 C. Manchein: <http://orcid.org/0000-0001-8786-3893>

### References

- [1] World Health Organization. Coronavirus disease (COVID-2019) situation reports.

- <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>, 2020.
- [2] J. T. Wu, K. Leung, and G. M. Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 395:689 – 697, 2020.
- [3] C. Anastassopoulou, L. Russo, A. Tsakris, and C. Siettos. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLOS ONE*, 15(3):1–21, 2020.
- [4] F. Ndaïrou, I. Area, J. J. Nieto, and D. F.M. Torres. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Soliton. Fract.*, 135:109846, 2020.
- [5] D. Fanelli and F. Piazza. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Soliton. Fract.*, 134:109761, 2020.
- [6] P. Wu et al. Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. *Eurosurveillance*, 25(3), 2020.
- [7] S. Chatterjee, A. Sarkar, S. Chatterjee, M. Karmakar, and R. Paul. Studying the progress of COVID-19 outbreak in India using SIRD model. *medRxiv*, 2020.
- [8] K. M. Khalil et al. An agent-based modeling for pandemic influenza in Egypt. In *Handbook on Decision Making*, pages 205–218. Springer-Verlag, Berlin, 2012.
- [9] B. J. Coburn, B. G. Wagner, and S. Blower. Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1). *BMC Med.*, 7:30, 2009.
- [10] M. Lipsitch et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–1970, 2003.
- [11] C. Manchein, E. L. Brugnago, R. M. da Silva, C. F. O. Mendes, and M. W. Beims. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies. *Chaos*, 30:041102, 2020.
- [12] E. L. Brugnago, R. M. da Silva, C. Manchein, and M. W. Beims. How relevant is the decision of containment measures against covid-19 applied ahead of time? *arXiv:2005.01473*, 2020.
- [13] B. F. Maier and D. Brockmann. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in Mainland China. *arXiv:2002.07572*, 2020.
- [14] Y. S. Long et al. Quantitative assessment of the role of undocumented infection in the 2019 novel coronavirus (COVID-19) pandemic. *arXiv:2003.12028*, 2020.
- [15] H. Hethcote, M. Zhién, and L. Shengbing. Effects of quarantine in six endemic models for infectious diseases. *Math. Biosci.*, 180:141 – 160, 2002.
- [16] E. L. Piccolomini and F. Zama. Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model. *arXiv:2003.09909*, 2020.
- [17] A. Vazquez. Polynomial growth in branching processes with diverging reproductive number. *Phys. Rev. Lett.*, 96:038702, 2006.
- [18] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.
- [19] H. M. Singer. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *arXiv:2004.03859v1*, 2020.
- [20] R. Marsland and P. Mehta. Data-driven modeling reveals a universal dynamic underlying the COVID-19 pandemic under social distancing. *medRxiv:10.1101/2020.04.21.20073890*, 2020.
- [21] T. Ray. Graph theory suggests COVID-19 might be a ‘small world’ after all. <https://www.zdnet.com/article/graph-theory-suggests-covid-19-might-be-a-small-world-after-all/>, 2020.
- [22] A. Ziff and R. Ziff. Fractal kinetics of COVID-19 pandemic, 2020.
- [23] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *Ann. Statist.*, 35:2769–2794, 2007.
- [24] G. J. Székely and M. L. Rizzo. The distance correlation t-test of independence in high dimension. *J. Multivariate Anal.*, 117:193–213, 2013.
- [25] C. F. O. Mendes and M. W. Beims. Distance correlation detecting Lyapunov instabilities, noise-induced escape times and mixing. *Physica A*, 512:721 – 730, 2018.
- [26] C. F. O. Mendes, R. M. da Silva, and M. W. Beims. Decay of the distance autocorrelation and Lyapunov exponents. *Phys. Rev. E*, 99:062206, 2019.
- [27] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. S. Coelho. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. *Chaos Soliton. Fract.*, 135:109853, 2020.
- [28] L. Tarrataca, C. M. Dias, D. B. Haddad, and E. F. Arruda. Flattening the curves: on-off lock-down strategies for COVID-19 with an application to Brazil. *arXiv:2004.06916v1*, 2020.
- [29] N. Crokidakis. COVID-19 spreading in Rio de Janeiro, Brazil: Do the policies of social isolation

- really work? *Chaos Soliton. Fract.*, 136:109930, 2020.
- [30] Ministério da Saúde. Painel coronavirus. <https://covid.saude.gov.br/>, 2020.
- [31] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A*, 115(772):700–721, 1927.
- [32] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. ii - the problem of endemicity. *Proc. R. Soc. Lond. A*, 138(834):55–83, 1932.
- [33] E. D. Sontag. For differential equations with  $r$  parameters,  $2r + 1$  experiments are enough for identification. *J. Nonlinear Sci.*, 12:553, 2002.
- [34] J. H. Beigel et al. Remdesivir for the treatment of covid – 19 — preliminary report. *N. Engl. J. Med.*
- [35] T. A. Mellan et al. Report 21: Estimating COVID-19 cases and reproduction number in Brazil. <http://hdl.handle.net/10044/1/78872>, 2020.
- [36] J. Wallinga and P. Teunis. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *Am. J. Epidemiol.*, 160(6):509–516, 2004.
- [37] R. N. Thompson et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.