medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Modinterv COVID-19: An online platform to monitor the evolution of epidemic curves

Arthur A. Brum[a], Giovani L. Vasconcelos[b], Gerson C. Duarte-Filho[c], Raydonal Ospina[d,*], Francisco A. G. Almeida[c], Antônio M. S. Macêdo[a]

[a]*Departamento de Física, Universidade Federal de Pernambuco, 50670-901 Recife, Pernambuco, Brazil*
[b]*Departamento de Física, Universidade Federal do Paraná, 81531-990 Curitiba, Paraná, Brazil*
[c]*Departamento de Física - Universidade Federal de Sergipe, 49100-000, São Cristóvão, Sergipe, Brazil*
[d]*Departamento de Estatística, CASTLab, Universidade Federal de Pernambuco, 50740–540, Recife, Pernambuco, Brazil*

Abstract

Background: The COVID-19 pandemic is one of the worst public health crises the world has ever faced. A major hindrance in making apt decisions by health control systems is the fact that protocols tested in other epidemics are no guarantee of success to control the COVID-19 epidemic, given its singular nature and complexity. The occurrence of two or more waves of infections all over the world poses an even greater challenge. An effective way to assist health authorities in adopting public policies to face the COVID-19 pandemic depends on smart analytics and visualization.
Purpose: We present the software Modinterv as a tool to monitor, in an automated and user-friendly manner, the evolution and trend of COVID-19 epidemic curves, both for cases and deaths.
Methods: The Modinterv software uses parametric generalized growth models, together with machine learning algorithms, to fit epidemic curves with one or two waves of infections of cases and deaths for countries around the world as well as for states and cities in Brazil and the USA. The richness of the implemented models lies in the possibility of detecting the distinct acceleration regimes of the disease in places where there are one or two waves of infections.
Results: We show how growth models can be combined with machine learning algorithms in an automated software that can identify the current stage of the COVID-19 epidemic curve in the selected place. We describe the backend structure of software as well as its practical use. The software helps the user not only to understand the current stage of the epidemic in the chosen location but also to make short term predictions as to how the curves may evolve.

Keywords: COVID-19, Epidemic curve, Growth models, Intervention strategies, Machine learning

1. Introduction

The complex evolution pattern of an epidemic outbreak, such as the ongoing COVID-19 pandemic, requires a dynamic response of public health policies based on reliable

*Corresponding author: raydonal@de.ufpe.br

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

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

factual evidence and accurate information analysis. Generally speaking, one can identify different stages of response to an outbreak, ranging from the detection of the first cases, involving surveillance systems and especially qualitative measures of risk assessment, followed by the assessment of the dynamics of transmission in the intervention phases, where more complex analyses are required to inform and guide the authorities as to the adoption of appropriate non-pharmacological interventions (e.g., intermittent lockdowns) and, if available, pharmacological measures as well (e.g., vaccination strategies), up to the registration of the last cases that recover or dies (assuming the disease does not become endemic). Finally, a post-intervention stage emerges, where lessons learned can help and improve protocols and preparation for a possible next epidemic.

An important feature of the modern response to epidemics is the increasing focus on exploiting all available data, including geo-referenced information of specific population groups, such as countries, cities, and even smaller administrative units (e.g., neighborhoods or census sectors). Taking into account this wealth of information can help in a macro view of the situation, enabling rapid responses and evidence-based decision making [5, 7]. The use of data and modeling techniques, together with information tools that range from data collection at service points to the generation of informative situational reports (apps, dashboards, tweets, etc.) [32, 4], allows for a better dissemination of reliable information and contributes to an accurate public perception of the epidemic situation.

From this perspective, we can obtain important insights about the evolution of an outbreak in a given population group by analyzing the corresponding epidemic curves, as represented by either the cumulative or the daily number of cases or deaths as a function of time [12, 11]. Epidemic curves are useful in many aspects because they provide a simple visual outline of demographic dynamics, which can be used to assess the growth or decline of an outbreak [16] as well as to assess the effect of intervention measures. In addition, epidemic curves also form the raw material used by a wide range of modeling techniques for monitoring and forecasting [28, 27, 15, 31]. Having good mathematical models to describe the empirical data is a necessary condition for this endeavor. In this context, phenomenological growth models are an important tool to analyze epidemic data [6, 28, 27], because contrary to other epidemiological models, such as compartment models [13, 9] and agent-based models [14, 22], growth models often admit an analytic solution—a fact that simplifies both the model analysis and its numerical fitting to the data.

In this paper, we present an automated software application, called Modinterv, that enables the user to monitor the COVID-19 epidemic curves of cases and deaths for different countries around the world as well as for states and cities in Brazil and the USA. The app implements a general class of growth models (to be discussed later) to fit the case and death curves of the selected location. The implemented models can be divided into two main subclasses, depending on whether the chosen data has one or two waves of infections. To fit single-wave curves, the app implements four specific models, as follows: i) the q-exponential model, which is suitable for the early rapid-growth phase of the epidemic; ii) the Richards model and iii) the generalized Richards model, which are used for epidemics that are in the intermediate stage; and iv) the beta logistic model (BLM), which is appropriate for the late stage of the epidemic when the cumulative curve is approaching a plateau. For locations where, after an initial leveling-off of the curve, there has been a resurgence of cases and deaths, indicating a second wave of infections, the app uses a generalized version of the BLM where the model parameters become time dependent, so as to capture the two-wave pattern. The current version of the Modinterv app uses

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

models with only up to two waves. Implementing automatic models with higher number of waves, although possible in principle, is a more complex computational undertaking that is beyond the scope of the present paper.

Once the user chooses the type of data (i.e. cases or deaths) and the desired location, the app automatically decides, based on a machine learning (M.L.) algorithm, whether the corresponding data has one or two waves and then fits the data with the appropriate best model selected among the ones mentioned above. From the best fitted model, the app provides the user with relevant information about the epidemic evolution in the chosen location. For example, the app presents an output plot with the theoretical curve, as obtained from by the best-fit model, superimposed with the empirical data, where special marks (colored vertical lines) are drawn to indicate the different acceleration regimes of the epidemic curve. Furthermore, by comparing the location of the last data point in relation to these special points, the app then informs the user the current stage of the epidemic, according to a refined classification scheme that considers not only the curve acceleration but also its second and third derivatives, known as the jerk (or jolt) and the jounce (or snap), respectively. The app also allows the user to perform additional analyses of the fitted results by clicking on extra check boxes or moving certain sliding bars.

It will be argued that the mathematical analysis provided by our application can be useful to public health authorities, not only because it indicates the current dynamical stage of the epidemic in a given place but also because it can help to predict its likely evolution in the near future. This type of information, combined with other analyses, can in turn help the authorities in their decision making process regarding, say, the adoption or relaxation of non-pharmacological interventions [25] and other containment measures. It should also be emphasized from the outset that the Modinterv app, by directly implementing mathematical models, provides information about the epidemic dynamics which cannot be obtained neither from a mere visual inspection of the raw empirical curve nor by using only its moving-average smoothed version.

The paper is organized as follows. In section 2 we discuss some general properties of epidemic curves, with emphasis on their different dynamical regimes for the cases of one and two waves of infections. In section 3, an overview of the mathematical models implemented in the Modinterv application is presented, while a more thorough discussion of the models is given in the Appendices. A detailed description of the backend structure of the app is presented in section 4, where the several `Python` modules used by the app are discussed. This section also contains a brief description of the machine learning algorithm implemented in the app to distinguish between curves with one and two waves. In section 5, we give an overview of the general features of the software and how the user can interact with these features. The user interface is divided into two sections, one for Countries and the other for States and Cities in Brazil and the USA, and the functioning of each section is discussed in detail. Illustrative examples for locations with one and two epidemic waves are given in section 6, where a detailed guide to generate and customize the output graphs is also presented. Finally, the main conclusions of the paper are summarized in section 7.

2. Epidemic curves and their dynamical stages

Here we discuss some general properties of typical epidemic curves and their different dynamical growth stages. First, we shall consider the case in which the epidemic has only one wave of infection and describe the main acceleration regimes of the corresponding

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

cumulative curve of cases or deaths. After that, we we will extend the discussion to the case of epidemic curves with two waves.

2.1. Single-wave epidemic curves

In the case of a single-wave epidemic, the cumulative curve, whether for the number of cases or deaths, typically has sigmoidal-like shape, as shown Fig. 1(a). Such a curve has three clear distinct regions, namely: i) an early period of rapid, accelerated growth; ii) an intermediate region where the curve grows approximately linearly in time; and iii) a late growth phase when the curve "bends away" from the linear profile and tends to a saturation plateau. One can give a better characterization of these three growth phases in terms of the corresponding acceleration regimes, as described next.

The early growth phase corresponds to a regime of *increasing acceleration*, when the acceleration grows from nearly zero at the onset of the epidemic and reaches a maximum value at some time, denote by t_1 and depicted in Fig. 1 by a a dashed orange vertical line. After this time, the epidemic enters its intermediate phase characterized by two acceleration regimes: i) first we have a regime of *decreasing acceleration*, where the acceleration decreases (from the maximum at t_3) and reaches zero at time $t_2 = t_c$, indicated by a yellow vertical line in Fig. 1, which represents the inflection point of the growth profile. After the inflection point, the acceleration becomes negative and increases in magnitude, thus starting the regime of *increasing deceleration*, which ends at the time, t_3, indicated by the green vertical line in Fig. 1, when the deceleration is maximum (the acceleration is minimum). After t_3, it begins the late growth phase where the deceleration starts to decrease and will approach zero towards the end of the epidemic.

The late growth phase is of particular interest because it indicates that the epidemic has well passed its "peak" (corresponding to the yellow line in Fig. 1) and is now entering its final phase (supposing there is no resurgence of infections). Because of its relevance for monitoring the possibly approaching end of the epidemic, it is convenient to divide this regime of decreasing deceleration into two distinct dynamical stages, according to whether the rate of change of the acceleration, known as jerk, is increasing or decreasing. In the first of such stages, the jerk increases from zero (at the time t_3) and reaches its maximum at some time which we denote by t_4 (indicated by the blue vertical line in Fig. 1). As the effect of the positive jerk is to start to bend the curve away from its near-linear profile seen in the intermediate phase [27], we shall refer to this regime of decreasing deceleration and increasing jerk as indicating a *transition to saturation*. The second stage of the late growth phase, which starts at t_4, corresponds to a regime of decreasing deceleration and decreasing jerk, which will be referred to as the *saturation* of the epidemic. In this final stage the first three derivatives of the growth profile are all decreasing functions of time (in absolute values), indicating that the epidemic curve is indeed approaching its saturation plateau.

In summary, the characteristic points that we shall use to classify the dynamical stages of an epidemic curve are as follows: i) the point t_1 of maximum acceleration; ii) the point $t_2 = t_c$ of zero acceleration (corresponding to the inflection point of the cumulative curve); iii) the point t_3 of minimum acceleration (or maximum deceleration); and iv) the point t_4 of maximum jerk in the deceleration phase. In order to classify the current stage of an ongoing epidemic, one first needs to fit the empirical data with an appropriate mathematical model from which these characteristic points can be computed; see Sec. 3 and the Appendices. Then, by comparing the position of the last data point of

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

the empirical curve, to be referred as the 'current time' t_f, with the characteristic points of the theoretical curve, one can estimate the current stage of the epidemic with more precision than, say, just by visual inspection of the data. Furthermore, this classification scheme has the advantage that it can be implemented automatically, i.e., without human assistance, as will be discussed later.

In summary, the ModInterv software (to be discussed later) classifies a given single-wave epidemic curve according to the following five epidemic stages:

1. Increasing acceleration: $t_f < t_1$.
2. Decreasing acceleration: $t_1 < t_f < t_2$.
3. Increasing deceleration: $t_2 < t_f < t_3$.
4. Transition to saturation: $t_3 < t < t_4$.
5. Saturation: $t_f > t_4$.

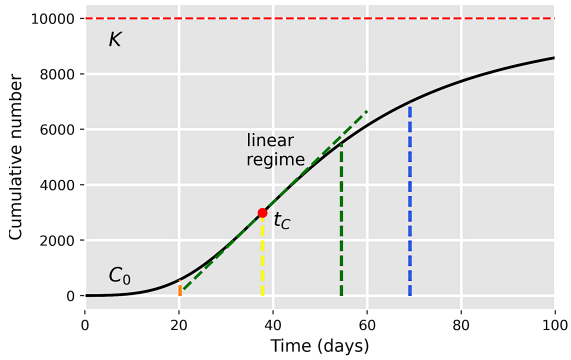
In order to describe epidemic data with only one wave of infections, the Modinterv implements four mathematical growth models to fit the empirical curve of case and death curves for the selected location. The choice of the model that best fits the respective chosen data depends on the respective current stage of the empirical curve, as discussed in more detail below. In the case of a single-wave epidemic curve all four models have analytic solution, from which the characteristic points t_i (vertical lines in Fig. 1) can be computed. Each particular model is more appropriate for a given curve depending on its current dynamical stage, as will be discussed later.

2.2. Multiple-wave epidemic curves

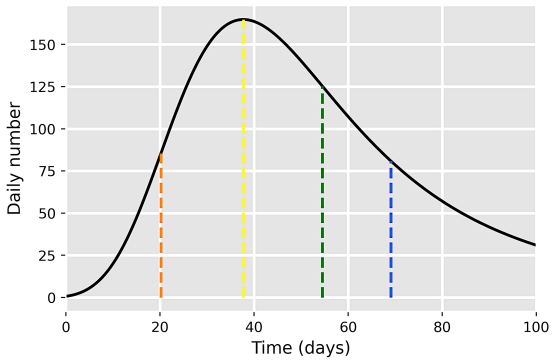
The above classification scheme can be naturally extended to the case where the epidemic curve has more than one wave. An example of a cumulative curve with two waves is shown in Fig. 2(a). In this case, each wave will in general undergo the five dynamical stages described in the preceding section. In other words, at some point after the first wave enters the saturation regime, the empirical curve reverses trend and starts to accelerate again, reflecting a resurgence of infections, and a new sequence of acceleration regimes ensues, until the epidemic enter its final saturation stage and can be said to be definitely under control (assuming there is no subsequent wave). Thus, in this case we now have two sets of characteristic points $\{t_i^{(I)}, t_i^{(II)}\}$, $i = 1, .., 5$, for the first and second waves, respectively The characteristic points $t_i^{(II)}$ of the acceleration regime for the second wave are indicated in Fig. 2(a) by colored vertical solid lines, whereas the characteristic points $t_i^{(I)}$ of the first wave are represented by dashed vertical lines. In this figure, the parameters K_1 represents the plateau of the first wave, which is an estimate of the number of cases/deaths if the second wave had not happened; whereas K_2 is the actual final plateau after the second wave. Also indicated in Fig. 2(a) is the location (the red circle from which descends a dashed black line) of the beginning of the second wave. The daily curve corresponding to the cumulative of Fig. 2(a) is shown in Figure 2(b), where the peak of each wave is indicated by inverted red triangles.

As will be discussed below, the two-wave model used by the app Modinterv does not have an exact solution, so that the calculation of the characteristic points t_i has to be performed numerically. More specifically, in the case of empirical epidemic curves with two waves, after the numerical fit is done, an interpolation of the corresponding theoretical curve generated by the fitted model is made using splines. The location of the maxima and minima of the second and third derivatives of the spline interpolation are computed,

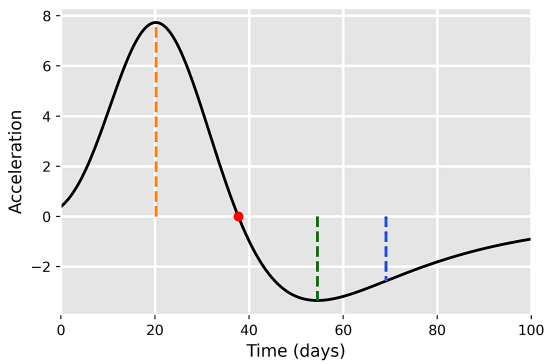
medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



(a)



(b)

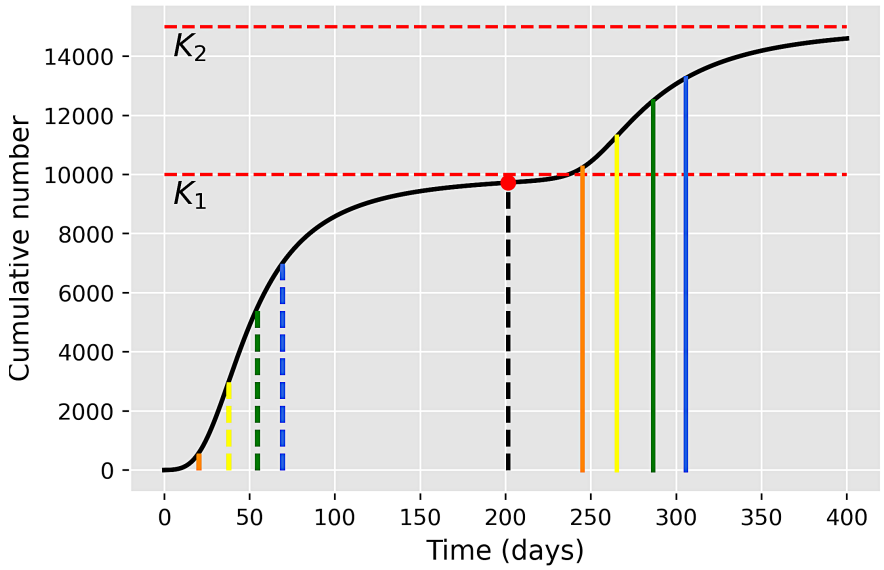


(c)

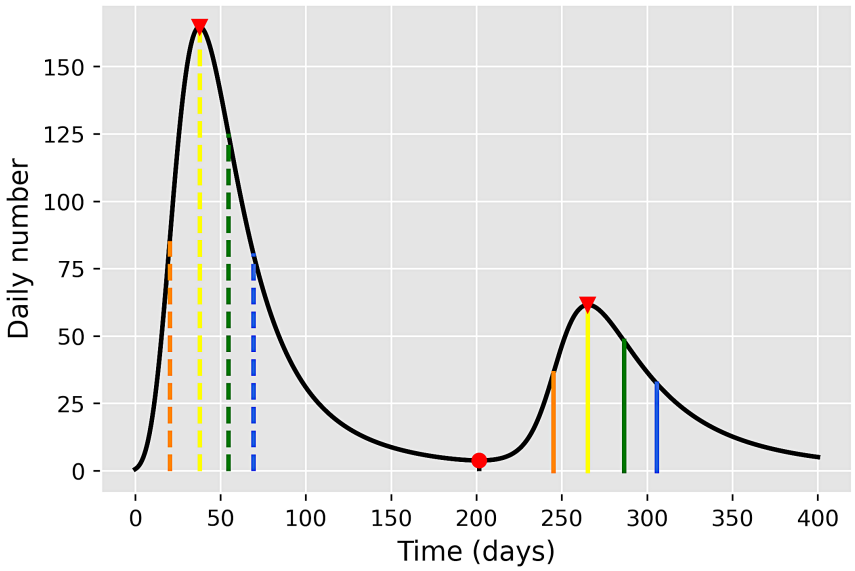
Figure 1: (a) Schematic of a cumulative epidemic curve with one wave of infection, illustrating its four main acceleration regimes, as separated by the dashed colored vertical lines; see main text for a description of the corresponding acceleration regimes. The point t_c indicates the inflection point of the cumulative curve. (b) Velocity and (c) acceleration curves, corresponding to the first and second derivatives, respectively, of the cumulative curve shown in (a).

thus determining the position of the four vertical lines for each of the waves, as illustrated in Fig. 2. As in the case of a single-wave epidemic discussed above, the current stage of the second wave in a given location is determined by comparing the last data point (or current time) with the respective characteristic points of the theoretical curve that

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



(a)



(b)

Figure 2: Same as Fig. 1 for (a) cumulative and (b) daily epidemic curve for the case of two waves of infections. The dashed (solid) colored vertical lines indicate the acceleration regimes during the first (second) wave. The red dot indicates the beginning of the second wave, while inverted triangles in (b) indicate the respective peaks of the first and second waves.

best fits the data. The classification of the dynamical stages for epidemic curves with two waves discussed above, see Fig. 2, naturally extends to three or more waves, where each wave would have its corresponding sequence of acceleration regimes. However, as will be argued later, devising an automated software to classify epidemic data with more

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

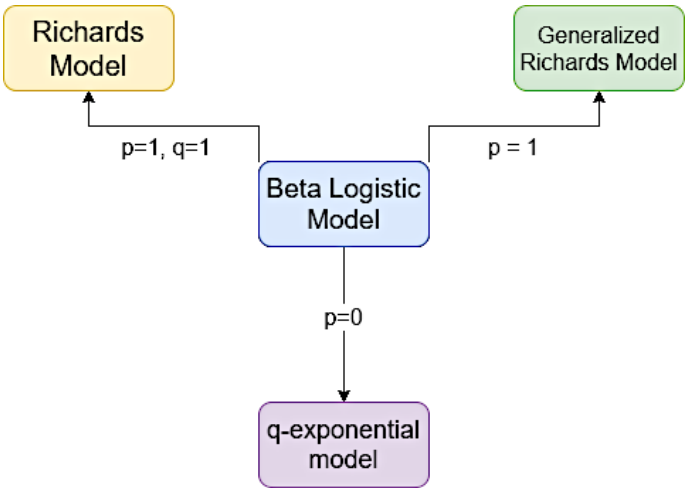


Figure 3: Schematic figure indicating the q-exponential, Richards and generalized Richards models as particular cases of the beta logistic model.

than two waves is a more challenging and numerically demanding task that will be left for future work.

3. Mathematical models implemented in the app: an overview

As already mentioned, in order to classify the dynamical stage of a given epidemic curve, it is first necessary to adjust a mathematical model to the empirical data. The ModInterv app implements two general classes of deterministic growth models, depending on whether the chosen data has been identified as presenting one or two waves of infections. (A machine learning algorithm is used to automatically classify epidemic curves according to the number of waves, as will be described in Sec. 4.4, after which the relevant class of models is applied.) Here we give a brief overview of the models implemented by ModInterv, while referring the interested reader to Appendices A and B for their mathematical aspects.

To fit single-wave curves, the app implements a generalized logistic model with *constant parameters*, known as the beta logistic model (BLM), which is one of the most general mathematical growth models and includes several well-known growth models as particular cases [23, 27]. As the full BLM is more suitable to epidemic curves that are in the late growth phase, the app also implements separately three particular cases, which are in general applicable to curves in less advanced stages. The relevant particular cases of the BLM are as follows (in decreasing degree of complexity): i) the generalized Richards model (GRM); ii) the Richards model (RM); and iii) the q-exponential model, as illustrated in Fig. 3. The latter model describes monotonically increasing curves and so it is only suitable for curves in the early growth phase (i.e., with increasing acceleration); whilst the three other models (BLM, GRM, and RM) describe sigmoidal curves with different degrees of mathematical complexity; see Appendix A for a detailed description of the four single-wave models above.

For a given empirical curve with only one wave, the Modinterv seeks to determine the corresponding best model by successively applying the four single-wave models above, from the most complex to the simplest. More specifically, first the app tries to fit the data with the BLM. If the BLM does not converge, indicating that the epidemic curve is

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

probably not yet in the saturation phase, then the app tries to fit the data with the next less complex model, namely the GRM. Similarly, if the GRM does not converge either, then the RM is applied. If the RM also fails, this is taken as an indication that the curve is in the early growth phase, when the acceleration is still growing (and hence no inflection point is present), in which case the q-exponential model is applied.

For epidemic curves with two waves, the app implements a generalized version of the BLM where the model parameters are now taken to be time dependent. More specifically, each of the five parameters of the BLM are assumed to vary in time, according to a logistic-like function, between two plateaus, corresponding to the parameter values for the first and second waves, respectively; see Appendix B. Thus, if the machine learning algorithm classifies a given epidemic curve as having two waves, the app fits the data with the two-wave BLM.

The two-wave model mentioned above can be naturally extended to include an arbitrary number N of waves, by assuming that the time dependence of the parameters is a generalized logistic-like function with $N \geq 2$ plateaus [24]. It should be noted, however, that implementing a fitting algorithm that operates without human assistance for a generic N-wave models poses two main technical challenges. First, one needs an "intelligent" algorithm for deciding how many waves there are in a given empirical curve. As will be discussed in Sec. 4.4, in our app Modinterv we have implemented a machine learning algorithm that can distinguish between epidemic curves with one and two waves, but an extension to three or more waves, although possible in principle, would render the app too slow for the online user due to the extensive amount of data required to train the machine learning model. The second technical difficult is the number of free parameters which rapidly increases with the number N of waves; see Appendix B. This poses an additional challenge for devising an automated fitting procedure that can handle such a large number of parameters without incurring in excessive overfitting. For these reasons, the current version of the Modinterv is limited to up to only two waves. (Current work is being carried out to extend the app to process data with a large number of waves, but this is beyond the scope of the present paper.)

One important point to note here is that once a mathematical model (between the one and two-wave models) is adjusted to the data for a given location, several important information concerning the epidemic evolution in that location can be extracted from the model. For instance, if one of the three single-wave models (i.e., BLM, GRM, or RM) that describe sigmoidal curves is selected, the characteristic points t_i that define the various acceleration regimes of the curve (see vertical lines in Fig. 1) are given by analytic expressions in terms of the model parameters; see Appendix A. (If the q-exponential model is selected, then the curve is necessarily in the regime of increasing acceleration, as discussed above.) For the two-wave model, the set of characteristic points $\{t_i^{(I)}, t_i^{(II)}\}$ for the two waves have to be determined numerically, but this is not a difficult task, as discussed in the preceding section. Thus, in either case (i.e., for one- and two-wave curves), once a mathematical model is fitted to the data, the current stage of the epidemic curve can be readily determined, according to the classification scheme discussed in Sec. 2. In all cases, the name of the best fitted model and the corresponding epidemic stage of the curve are shown in the legend box of the output plot. Additional relevant information, such as the dates of the peaks of the daily curve as well as the starting date of the second wave (when there is one), is also given in the output plot, as will be described in Sec. 5. But before going into that, we would like to present the general backend structure of the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

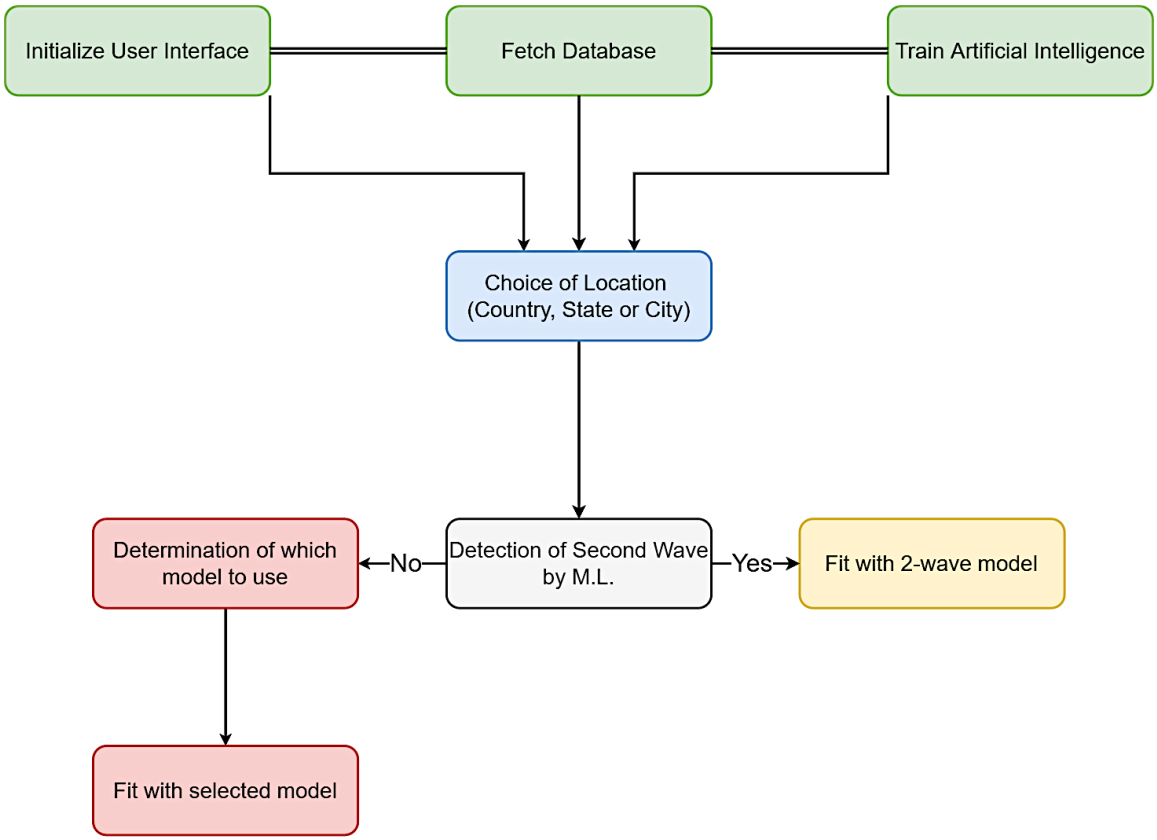


Figure 4: Simplified schematic description of the workflow of the software behind the app Modinterv.

ModInterv software.

4. Software system: backend structure

In this section we will explain in detail the workflow of the software behind the app Modinterv and all the tools used in its implementation. A simplified schematics of the functioning of the app is shown in figure 4. A detailed description of the different app components is given below.

4.1. Initialization

When the app is accessed (via browser or Android App), the user interface starts to load and the square buttons that appear along the page will read "*Loading Widgets...*"; after a few seconds the text in the square buttons will change to "*Show Widgets*". When the user clicks any of these buttons, the text in all of them changes to "*Initializing Widgets...*"; and after a short while a brief introductory text about the app appears, along with a loading bar. While this loading bar appears, the machine learning model is being trained (this procedure will be discussed in detail in a later section). Finally, after a few more seconds, the rest of the app will be loaded and ready to be used, as explained below.

4.2. User interface

The user interface of the App Modinterv was designed using objects called *Widgets*, which are implemented in the Python module `ipywidgets`. This module offers a vast

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

```
widgets.ToggleButton(
    value=False,
    description='Click me',
    disabled=False,
    button_style='', # 'success', 'info', 'warning', 'danger' or ''
    tooltip='Description',
    icon='check' # (FontAwesome names without the `fa-` prefix)
)
```

✔ Click me

(a) Button Widget.

```
widgets.FloatSlider(
    value=7.5,
    min=0,
    max=10.0,
    step=0.1,
    description='Test:',
    disabled=False,
    continuous_update=False,
    orientation='horizontal',
    readout=True,
    readout_format='.1f',
)
```

Test: 7.5

(b) Slider Widget.

```
widgets.Checkbox(
    value=False,
    description='Check me',
    disabled=False,
    indent=False
)
```

☐ Check me

(c) Checkbox Widget.

```
widgets.Dropdown(
    options=['1', '2', '3'],
    value='2',
    description='Number:',
    disabled=False,
)
```

Number: 2

(d) Dropdown Widget.

Figure 5: Illustrations of some of the Widgets from the `iPywidgets` module used in the app

miscellaneous of *Widgets* to accomplish the most diverse tasks. In the development of the app Modinterv we have used only five types of *Widgets*, namely: HTML, Button, IntSlider, Dropdown and Checkbox.

The HTML *Widgets* is used to display dynamic text throughout the app, avoiding the whole page to be reloaded when the user changes the language (from Portuguese to English or vice-versa). The other four *Widgets* are used to receive inputs from the user, such as the type of the epidemic curve to be fitted (i.e., whether Cases or Deaths), the Country, State or City chosen, the number of days in the epidemic curve to be considered in the fit, among others. Illustrations of each widget are shown in figure 5.

4.3. Data acquisition

The COVID-19 data for Countries used in the Modinterv are obtained from the database made publicly available by the Johns Hopkins University [10], which lists in automated fashion the number of the confirmed cases and deaths for each country in their database. The data used for States and Cities in Brazil were obtained from the GitHub database maintained by Wesley Cota [8], which is also automated and updated daily. Each time the app is initialized, both databases mentioned are accessed and the data for all Countries, States and Cities become available for the user.

4.4. Machine learning algorithm for detecting epidemic waves

The classification of curves that feature one or two epidemic waves is quite a challenge to be implemented automatically in an algorithm using purely mathematical and computational techniques. Although a trained human eye can easily identify when there are multiple waves of infections in a given empirical epidemic curve, this is non-trivial task to

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

be performed using purely mathematical quantities, such as derivatives, averages, etc. To solve this issue, we have implemented a machine learning algorithm that, after trained, classifies epidemic curves as having one or two waves.

The algorithm is based on the Python module `scikit-learn`, which implements several ML models. Among these models, we have selected three of them, namely: K-Nearest Neighbors [18], Decision Tree [21] and Random forest [3]. These are all supervised learning models, which in this specific problem are used for binary classification. Each set of entries related to an epidemic curve in the training dataset is labeled according to the presence (label 'yes') or absence (label 'no') of a second wave of infections or deaths.

The database used to train the ML models is composed of cumulative curves generated by the one- and two-wave beta logistic models described in Sec. 3 and the Appendices. In order to achieve a diverse scenario of epidemic curves to train the ML algorithm, we have used random parameters within their range of definition for each model. From each generated epidemic curve, we select a set of representative points (as explained below) to be passed to the ML models as features, together with a label, which is selected as follows: if the corresponding curve was generated by a one-wave model, then the label is *no*; whereas when the epidemic curve is generated by a two-wave model, the label chosen in *yes*.

Note that this label provides an answer to the following question: *Does this curve exhibit a second wave?* Each time the user selects a type of data (cases or deaths) and a location, the corresponding set of points selected from the chosen empirical curve is passed to the already trained ML algorithm and the same question is asked. The training database can be found online at `https://gist.github.com/ArthurAraujoBrum/12d867cc198f1ac2e3730ad067eed46d`.

Before passing the points to the machine learning models, the curves (both the ones from the training database and the empirical ones) are normalized, so that the last point of each curve is assigned the value 1, instead of the actual number of cases or deaths. This procedure helps the ML models to learn and, after they have learned, to compare new curves to the ones already trained, by eliminating the great difference between the number of cases or deaths in distinct locations.

In order to speed up the process, for each epidemic curve (either in the training or classifying stages), we feed into the ML algorithm only a set of 20 points equally spaced along the curve. As an example, we show in Figs. 6a and 6b the empirical epidemic curve and the set of selected points (without rescaling), respectively, for the cumulative number of COVID-19 deaths in Brazil up to 08/03/21 .

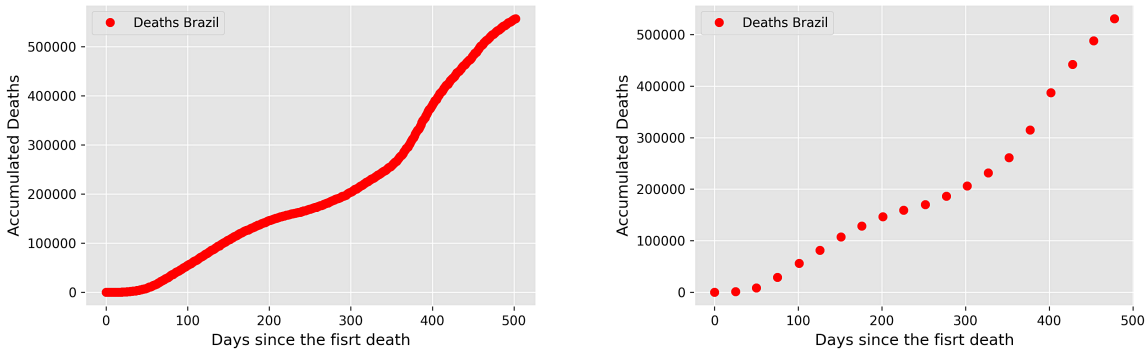
It is important to note, however, that these selected points are used only by the ML algorithm to identify whether the corresponding curve (from which the points were extracted) has one or two waves of infections. Once that decision is made, the subsequent numerical fits are performed using the entire curve (without rescaling), as discussed next.

4.5. Numerical fitting

If the ML algorithm identifies the chosen curve as having two waves, then the two-wave growth model is applied. If instead the curve is classified as having only one wave, then the four one-wave models are tested, according to the sequence described in Sec. 3, so as to decide which one is the most suitable for the chosen dataset.

In all numerical fits, for both the one-wave and two-wave models, the app employs the Levenberg-Marquardt (LM) algorithm to solve the non-linear least square optimization problem, as implemented in the `lmfit` package for the `Python` language, which has

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.



(a) Cumulative Deaths curve for Brazil up to 08/03/21.

(b) Selection of 20 evenly spaced points.

Figure 6: Selection of points from epidemic curvers to be passed as input for the machine learning models.

a built-in routine for estimating the errors of the fitted parameters via the covariance matrix [17]. The results of the fitting procedure are deemed acceptable when the errors in the parameters are smaller than the estimated values themselves; otherwise, a warning message is displayed to alert the user that the corresponding numerical fit should be viewed with some caution because of large errors in one or more parameters.

In the case of one-wave models, all free parameters are constant in time, so that a straightforward fitting procedure is performed: the empirical data and the specific model to be fitted are passed onto the LM routine, which then returns the estimated parameters and their errors. The number of parameters to be fitted depends on the chosen model. As discussed in Appendix A, the BLM has five parameters, $\{r, q, \alpha, p, K\}$; the GRM requires four parameters, $\{r, q, \alpha, K\}$; the GM contains three parameters, $\{r, \alpha, K\}$; whereas the q-exponential has only two parameters, $\{r, q\}$. The epidemiological meaning of these parameters are discussed in Appendix A. Because there are several parameters to be fitted, special care must be taken concerning the issue of over-fitting [27]. To minimize this risk, the parameters of the one-wave models are restricted to certain allowed ranges. More specifically, we have found that the restrictions $p \geq 1$, $0 < q \leq 1$, $0 < \alpha \leq 1$, and $0 < r < 1$ are useful criteria to reduce over-fitting; see discussion in Appendix A.5.

The numerical fitting of the two-wave model is more challenging because now all parameters $\{r, q, \alpha, p, K\}$ become functions of time. As briefly mentioned in Sec. 3 and discussed in more detail in Appendix B, each model parameter varies in time as a logistic-like function, going from an initial value corresponding to the first wave to a final value representing the second wave. In addition, the logistic function contains two additional parameters: the transition time, t_1, and the rate of transition, ρ_1, between the first and second waves. The fitting procedure in this case is performed in two steps, as follows. In the first step, we give an initial educated guess for the possible location of the transition time, t_1, between the first and second waves. We then fit the data up to this time with the one-wave BLM. The parameters found in this first step are then used as initial guesses for the parameters $\{r_1, q_1, \alpha_1, p_1, K_1\}$ relative to the first wave of the full two-wave model. Initial guesses for the respective parameters corresponding to the second wave are chosen arbitrarily within their respective range. With these initial guesses, we then carry out the LM numerical fit of the entire empirical data using the complete two-wave model, thus obtaining the two sets of estimated parameters, $\{r_i, q_i, \alpha_i, p_i, K_i\}$, for $i = 1, 2$, as well the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

time, t_1, and rate, ρ, of transition between waves. The LM routine also returns the error for all estimated parameters.

Notice that for the two-wave model one would have in principle a total of 12 free parameters to be determined. With such a large number of parameters, extra steps must be taken to minimize over-fitting issues. First, we impose the same range restrictions on the parameters r_i, q_i, α_i, and p_i, for $i = 1, 2$, as in the one-wave case; see above. Furthermore, the parameters α_2 and p_2 are kept fixed at unity, since we observed that letting them free tends to cause over-fitting. This can be explained by the fact that the parameter α is connected with the asymmetry of the daily curve around a peak, while p governs the decay rate after the peak [27]. However, there are often less points in the second wave [24], so that estimating these parameters for the second wave is arguably less reliable. Hence we prefer to set $\alpha_2 = p_2 = 1$ to reduce overfitting.

4.6. Classification of the epidemic dynamical stage

Once the numerical fit has been performed for a chosen empirical curve and the parameters of the corresponding growth model are determined, the app then computes the points t_i that characterize the distinct acceleration regimes of the curve; see Sec. 2. As shown in Appendix A, in the case of single-wave models the characteristic points t_i are all given by analytic expressions in terms of the model parameters; whereas in the case of the two-wave model, the two sets of characteristic points $\{t_i^{(I)}, t_i^{(II)}\}$, one for each wave, are computed numerically, after interpolating the theoretical curve with a spline and determining the corresponding zeros of the second, third, and fourth derivatives; see Sec. 2.2. Recall that each set of characteristic points divides the respective epidemic wave into five acceleration regimes, namely: i) increasing acceleration; ii) decreasing acceleration; iii) increasing deceleration; iv) transition to saturation; and v) saturation. As discussed in Sec. 2, by comparing the final time, t_f, of the last point of the empirical data (assumed to be the 'current time') with the computed characteristic points t_i that define the above regimes, the app then determines the current stage of the epidemic in that location. The corresponding stage for the chosen epidemic curve is then informed by the app in the legend box of the output graph, which shows the empirical data and the fitted curve, together with additional relevant information obtained from the fit, as will be discussed in Sec. 5.

4.7. Creating an interactive user interface

The code for the Modinterv app was written using iPython notebooks, a versatile cloud computing environment that allows for developing codes whithout the need to install all dependencies, as one would usually do when compiling codes directly from one's own computer. Unfortunately, iPython notebooks are not well suited for sharing with public users, because it does not have a user-friendly interface. Furthermore, in order for the users to be able to compile the code and access the app, every one of them would need to have editing permission, which could lead to undesirable results.

Thus, in order to generate an interactive webpage that could be accessed by the users with a customizable, user-friendly interface and with the code running in a backend separate from the interface, we have used `nbinteract`, a Python module that creates an static HTML that allows widgets to remains interactive by using `Binder` servers as the computational backend. The HTML file produced then has to be hosted in a server to be publicly accessible on the internet. The HTML page thus produced for the Modinterv app

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

can be currently accessed via the address `http://fisica.ufpr.br/modinterv`. Below we shall explain the contents of the app, as accessed via this homepage.

5. Software system: output graphs and interaction with user

As already mentioned, the Modinterv app allows the user to select the type of epidemic curves (i.e., either Cases or Deaths) to be analyzed as well the location of interest. To help the user choose the location of interest, the app is structured into two main sections: i) Countries and ii) States and Cities in Brazil and the US. The app also allows the user to choose how the output graphs are displayed on screen as well as to generate figures to be downloaded. A brief explanation of the app features is given below.

5.1. Countries section

In the first section of the app, the user can analyze the curves of cases and deaths for countries. After the user selects from corresponding dropdown menus the type of `Data` (Deaths or Cases) and the chosen `Country`, a preview with the respective accumulated curve (red circles) is generated for the total number of deaths/cases as a function of time, measured in days from the first death/case.

After clicking the button `Perform fit`, a text box with the message `Computing` will appear, and soon after a new graph with the model fit (black curve) superimposed on the data (red circles) will be displayed below the original preview plot. The name of the selected country is shown in the plot title, together with the date up to which the empirical data was considered, while the model that best fits the data is indicated in the legend box. The plot also shows colored vertical lines corresponding to the characteristic points t_i that define the five acceleration regimes of an epidemic wave, as explained in Sec. 2. The number of such vertical lines displayed will depend, of course, on the number of waves and the specific evolution stage of the empirical curve under analysis. In particular, for the case of two waves, the beginning of the second wave is also shown as a black dot on the fitted curve, from which it is drawn a dashed black vertical line to mark the separation between the first and second waves. In the legend box of the plot, the app also shows complementary information, such as i) the calendar date of the first case/death and ii) the dynamical stage of the epidemic, according to the classification scheme presented in Sec. 2. Furthermore, if the corresponding fit was performed with the two-wave model, i.e., if the empirical data was classified as having two waves, then the starting date of the second wave is also shown on the legend box.

Besides the information displayed on the plot of the fitted cumulative curve, the user has a few additional options to further analyze the results of the fitting process. First, checking the checkbox `Check to display/hide the daily curve` shows the empirical daily curve (red circles) together with the theoretical daily curve (black curve), where the latter corresponds simply to the time derivative of the mathematical fit for the cumulative curve. Second, checking the checkbox `Check to display the parameters of the fit` will produce a second plot of the cumulative curves (both empirical and theoretical), but now showing the estimated parameters of the best fitted model together with their errors. Third, moving the slider `Time Range` gives a short term prediction, ranging from 7 up to 28 days after the last data point, for the number of extra cases/deaths during the selected period ahead. The total cumulative number of cases/deaths at the end of this period is also shown. Fourth, the user has the option to choose the scale on both the x-Axis (horizontal axis) and y-Axis (vertical axis) between the linear and logarithmic scales.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Finally, clicking on the buttons Generate Figure File will generate the link for downloading the files with the figures, where the formats `png` and `eps` are supported.

To see a new type of data or choose a different country, the user only needs to select the desired options from the corresponding dropdown menus. After a new selection is made, the button `Perform fit` has to be clicked again to produce the fitted plot for the newly chosen empirical data.

5.2. Section of States and Cities in Brazil and the US

In the second section of the app, the user can analyze the death and case curves for states and cities in Brazil and states and counties in the United States (US). First, the user must select the chosen Country (Brazil or US), after which the Type of data (Deaths or Cases) has to be selected. The user then needs to select the type of Region (States or Cities) and the desired state or, if the option Cities was selected for Region, the desired city (in Brazil) or county (in the US). Again, a preview graph will be generated with the selected data for the chosen location. The rest of the procedure is as explained in the Country section.

6. Illustrative examples

In this section we demonstrate the Modinterv application for locations with one and two waves of infections and discuss each step of the process.

6.1. Example with two waves

As of this writing, a significant number of countries have experienced three (or more) waves of COVID-19, but we can still find some countries that present only two waves of infections. For this example, the user can go to the first section of the app and select from the `Data` menu the type of epidemic curve to be analyzed (in this case we chose Deaths) and the desired `Country`. Once the country is selected (Slovakia, in our case), a preview of the chosen epidemic curve is displayed, as shown in Fig. 7. After that, the user can proceed to click the button *Perform Fit* shown at the bottom of the preview plot, see Fig. 7, after which a small bar written *Computing* will appear and, soon after, the result of the numerical fit will be shown, as illustrated in Fig. 8.

Just above the output plot of the numerical fit, there are two checkboxes written `Check to display/hide the daily curve` and `Check to display the parameters of the fit`, respectively; see Fig. 8. Checking the first checkbox will produce the corresponding daily epidemic curves (both empirical and theoretical), as shown in Fig. 9. In this figure, the red dots correspond to the daily number of deaths and the black line is the time derivative of the theoretical cumulative curve shown in Fig. 8. Also shown in the plot of the daily curves are the locations (black dots) of the peaks of the first and second waves, with the corresponding calendar dates for the peaks being given in the legend box; see Fig. 9.

Checking the second checkbox yields a more detailed output of the numerical fit, showing the values of the fitted parameters and their respective errors, as illustrated in Fig. 10. The values K_1 and K_2 of the first and second plateaus, respectively, are also shown in the legend box of the plot, as determined from the fitted two-wave model; see Sec. Appendix B. The interpretation of these two parameters must, however, be done with care. More specifically, when the computed errors for the parameters K_1 and K_2 are reasonably small, as in the example shown in Fig. 10, one can then tentatively use the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

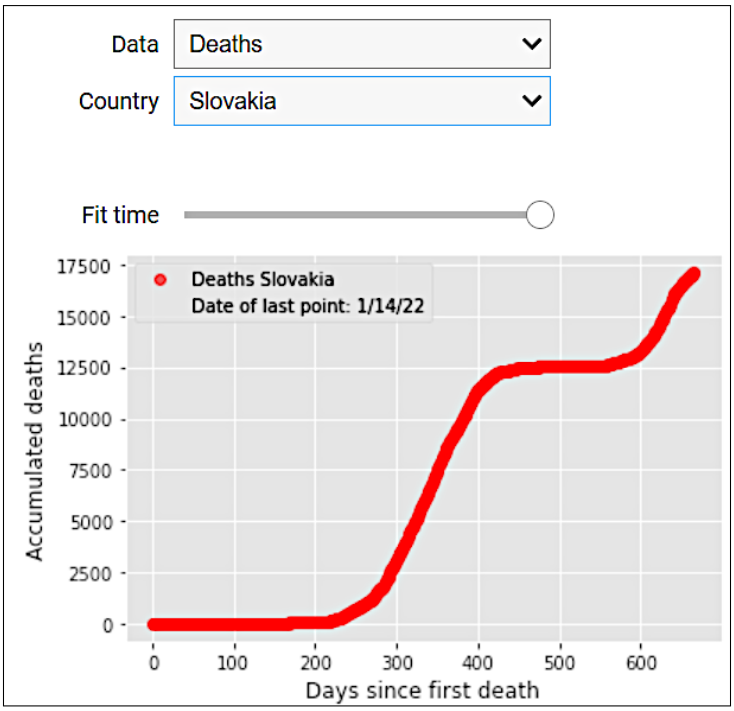


Figure 7: Preview output graph of the app Modinterv after selection of the type of Data (Deaths) and Country (Slovakia).

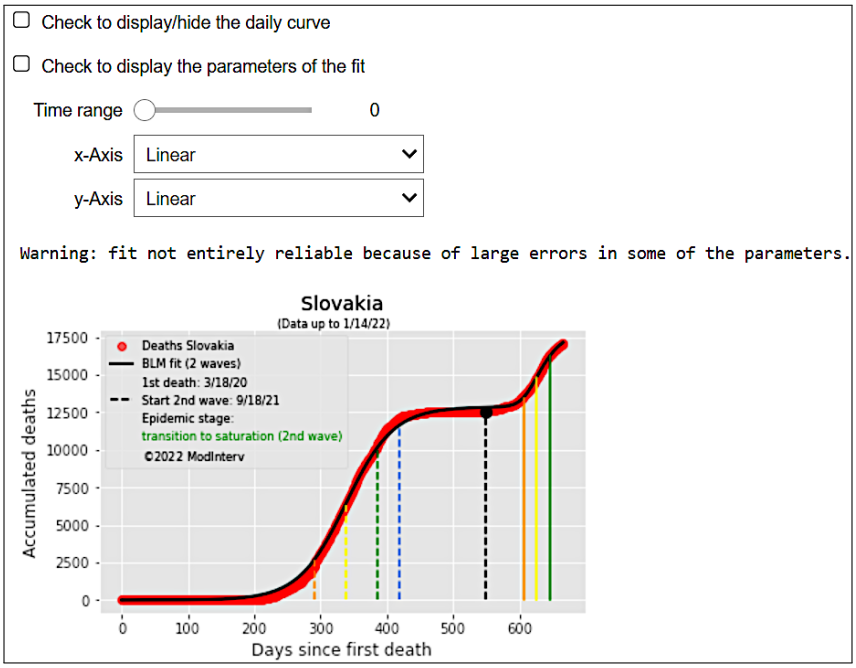


Figure 8: Output of the numerical fit of the epidemic curve (Deaths) for the selected country (Slovakia).

difference $\Delta K = K_2 - K_1$ as a rough estimate of the excess of cases/deaths owing to the second wave of infections [24].

6.2. One-wave example

To this day, practically every country has experienced at least a second wave of COVID-19 infections. This means that a single-wave growth model cannot describe the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

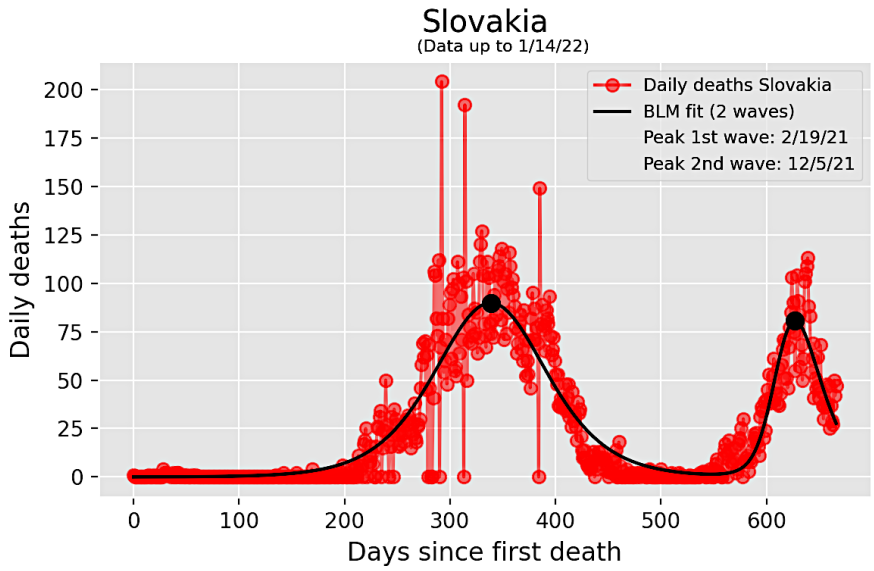


Figure 9: Daily curves (empirical and theoretical) corresponding to the cumulative curves shown in Fig. 8.

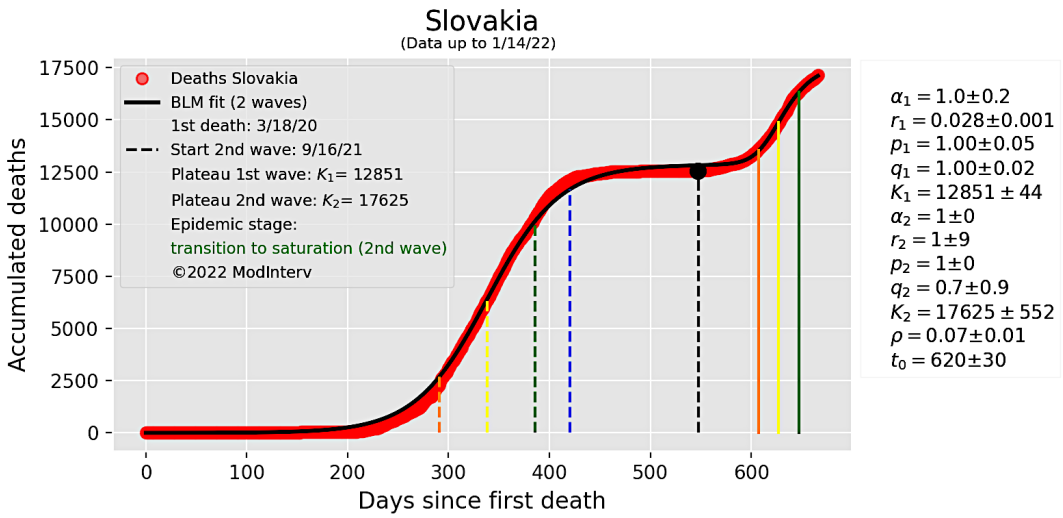


Figure 10: Detailed output of the numerical fit shown in Fig. 8, indicating the values of the fitted parameters with their respective errors.

COVID-19 epidemic curves for these countries up to the present time. In order to demonstrate the Modinterv for the case of one wave only, we can use a feature implemented in the app that allows the user to truncate the epidemic curve at a past date and perform the corresponding numerical fit only up to that date.

Let us consider, as an example, the case of Brazil. In order to go back to a time where we had only one epidemic wave, we can slide backwards the slider written *Time fit*. The preview plot is then automatically updated showing the empirical data only up to the selected date, which is shown in the corresponding legend box of the plot as the `Date of last point`.

After we choose an appropriate date up to which there is only one wave, we can click the `Perform Fit` button, which will lead to the same steps discussed in the two-wave

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

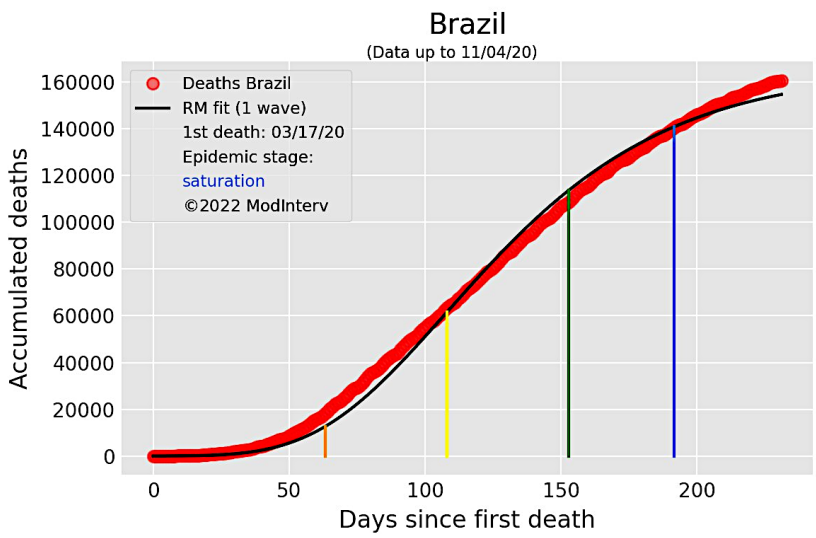


Figure 11: Numerical fit for the cumulative number of deaths in Brazil up to a time (11/04/2020) when there was only one wave of infection.

example above, but now the numerical fit will be performed with one of the four single-wave growth models discussed in Sec. 3. An example of this procedure is shown in Fig. 11, where the final date was chosen to be 11/04/20. The epidemic curve up to that date does indeed contain only one wave, which is best fitted with the Richards model, as shown in Fig. 11.

7. Conclusion

In this paper we have described an automated software application, , called Modinterv, that enables the user to analyze COVID-19 epidemic curves of cases and deaths for different countries around the world as well as for states and cities in Brazil and the USA. The application uses epidemic data available in several public databases. Once a location and the type of data (i.e., either cases or deaths) are selected, the app automatically fits the empirical data with a general class of mathematically growth models for curves with one or two waves of infections. From the best fitted mathematical model, relevant information about the progress of the COVID-19 epidemic in that location can be inferred.

One important information that can be obtained from the fitted models, as implemented in the Modinterv, is the dynamical stage of the epidemic in the chosen location. Our methodology allows for a finer classification scheme of the different growth regimes of an epidemic curve by considering five main acceleration regimes, as explained in Sec. 2. This should be compared, in contrast, with the common way to track epidemics evolution in terms of the effective reproduction number R_t, where we recall that $R_t > 1$ ($R_t < 1$) implies that the epidemic is accelerating (decelerating). Although R_t is widely used by epidemiologists and public health authorities, this quantity has nonetheless some drawbacks [2, 1, 25]. For example, as R_t essentially represents a measure of the epidemic acceleration, it cannot by itself distinguish between regimes of increasing or decreasing acceleration (for the same value of the acceleration). In this context, it is therefore useful to have additional tools to obtain a fuller description of the epidemic evolution, say by also analyzing its different acceleration regimes, as implemented in the Modinterv.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Furthermore, the classification scheme implemented in our application can provide relevant information to health authorities not only in regard to the current stage of the epidemic in the place of interest but also to its likely evolution in the near future. More specifically, by analyzing how recently the epidemic curve entered its current dynamical regime, one can also make some prediction as to when it is likely to progress to the next stage (if it has not entered the final one). Such an information can help the authorities in their decision making process regarding, say, the implementation or relaxation of non-pharmacological interventions [25]. Moreover, by relying on a general class of flexible growth models, from which the points separating the different acceleration regimes can be easily computed, the classification of a given epidemic curve can be performed automatically (i.e., without human assistance) by the software, thus making the method easily accessible to any interested person and without requiring specific mathematical knowledge.

The application also provides additional relevant information about the course of the epidemics which cannot be easily obtained from a visual inspection of neither the raw empirical data nor its moving-average smoothed version. For example, for epidemic curves with two waves the app provides from the fitted mathematical model a more precise estimate not only for the starting date of the second wave but also for the dates when the two peaks occurred. This information can, in turn, be useful for government and health authorities, as they can compare these relevant points in the epidemic evolution with the corresponding containment measures in place at those particular times. We believe this type of analysis can help one to understand the underlying reasons for the pattern changes of the epidemic curve.

The current version of the Modinterv app is limited by practical reasons to epidemic data that contain only up to two infection waves. Work is currently underway to extend the app to process data with more than two waves. Our software was primarily designed to process COVID-19 data, but given its general structure it could be readily applied to data from other infectious diseases. In this regard, and as an interesting research perspective, one can thus envisage a general purpose platform to analyze data for different diseases. Technically, this would be rather simple to implement. The main requirement/difficulty is that the relevant empirical data be available in public repositories.

References

[1] Adam, D. (2020). A guide to R-the pandemic's misunderstood metric. *Nature*, 583(7816):346–349.

[2] Anderson, R., Donnelly, C., Hollingsworth, D., Keeling, M., Vegvari, C., Baggaley, R., and Maddren, R. (2020). Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. *The Royal Society*, 2020.

[3] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[4] Brum, A. A., Duarte-Filho, G. C., and Vasconcelos, G. L. (2020). Application Modinterv COVID-19. Available at: `https://fisica.ufpr.br/modinterv/`.

[5] Cauchemez, S., Fraser, C., Van Kerkhove, M. D., Donnelly, C. A., Riley, S., Rambaut, A., Enouf, V., van der Werf, S., and Ferguson, N. M. (2014). Middle east respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *The Lancet infectious diseases*, 14(1):50–56.

[6] Chowell, G., Hincapie-Palacio, D., Ospina, J., Pell, B., Tariq, A., Dahal, S., Moghadas, S., Smirnova, A., Simonsen, L., and Viboud, C. (2016). Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics. *PLOS Currents Outbreaks*.

[7] Cori, A., Donnelly, C. A., Dorigatti, I., Ferguson, N. M., Fraser, C., Garske, T., Jombart, T., Nedjati-Gilani, G., Nouvellet, P., Riley, S., et al. (2017). Key data for outbreak evaluation: build-

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

ing on the Ebola experience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721):20160371.

[8] Cota, W. (2020). Monitoring the number of COVID-19 cases and deaths in brazil at municipal and federative units level. *SciELOPreprints:362*.

[9] Hattaf, K. and Dutta, H. (2020). *Mathematical Modelling and Analysis of Infectious Diseases*. Springer.

[10] JHU (2020). Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) . Accessed: 2020-03-29 `https://coronavirus.jhu.edu/map.html`.

[11] Jombart, T., Kamvar, Z. N., Cai, J., Pulliam, J., Chisholm, S., FitzJohn, R., Schumacher, J., and Bhatia, S. (2019). reconhub/incidence: Incidence version 1.6.0.

[12] Kamvar, Z. N., Cai, J., Pulliam, J. R., Schumacher, J., and Jombart, T. (2019). Epidemic curves made easy using the R package incidence. *F1000Research*, 8.

[13] Kermack, W. and McKendrick, A. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721.

[14] Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., et al. (2021). Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):e1009149.

[15] Macêdo, A. M. S., Brum, A. A., Duarte-Filho, G. C., Almeida, F. A. G., Ospina, R., and Vasconcelos, G. L. (2020). A comparative analysis between a SIRD compartmental model and the Richards growth model. *medRxiv*.

[16] Matrajt, L. and Leung, T. (2020). Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerging infectious diseases*, 26(8):1740.

[17] Newville, M., Stensitzki, T., Allen, D., and Ingargiola, A. (2015). Non-linear least-squares minimization and curve-fitting for Python. *Chicago, IL*.

[18] Peterson, L. (2009). K-nearest neighbor. *Scholarpedia*, 4(2):1883.

[19] Picoli Jr, S., Mendes, R., Malacarne, L., and Santos, R. (2009). q-distributions in complex systems: A brief review. *Brazilian Journal of Physics*, 39(2A):468–474.

[20] Richards, F. (1959). A flexible growth function for empirical use. *Journal of experimental Botany*, 10(2):290–301.

[21] Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

[22] Shamil, M. S., Farheen, F., Ibtehaz, N., Khan, I. M., and Rahman, M. S. (2021). An agent-based modeling of COVID-19: validation, analysis, and recommendations. *Cognitive Computation*, pages 1–12.

[23] Tsoularis, A. and Wallace, J. (2002). Analysis of logistic growth models. *Math. Biosciences*, 179(1):21–55.

[24] Vasconcelos, G. L., Brum, A. A., Almeida, F. A. G., Macêdo, A. M. S., Duarte-Filho, G. C., and Ospina, R. (2021a). Standard and anomalous waves of COVID-19: A multiple-wave growth model for epidemics. *Brazilian Journal of Physics*, 51(6):1867–1883.

[25] Vasconcelos, G. L., Duarte-Filho, G. C., Brum, A. A., Ospina, R., Almeida, F. A., and Macêdo, A. (2022). Situation of COVID-19 in Brazil in August 2020: An Analysis via Growth Models as Implemented in the ModInterv System for Monitoring the Pandemic. *Journal of Control, Automation and Electrical Systems*, pages 1–19.

[26] Vasconcelos, G. L., Macêdo, A. M., Duarte-Filho, G. C., Brum, A. A., Ospina, R., and Almeida, F. A. (2021b). Power law behaviour in the saturation regime of fatality curves of the COVID-19 pandemic. *Scientific Reports*, 11(1):1–12.

[27] Vasconcelos, G. L., Macêdo, A. M. S., Duarte-Filho, G. C., Brum, A. A., Ospina, R., and Almeida, F. A. G. (2021c). Power law behaviour in the saturation regime of fatality curves of the COVID-19 pandemic. *Scientific Reports*, 11(1):4619.

[28] Vasconcelos, G. L., Macêdo, A. M., Ospina, R., Almeida, F. A., Duarte-Filho, G. C., Brum, A. A., and Souza, I. L. (2020). Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ*, 8:e9421.

[29] Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.*, 10:113–126.

[30] Wang, X.-S., Wu, J., and Yang, Y. (2012). Richards model revisited: Validation by and application

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

to infection dynamics. *Journal of Theoretical Biology*, 313:12–19.

[31] Wu, K., Darcet, D., Wang, Q., and Sornette, D. (2020). Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dynamics*, pages 1–21.

[32] Yozwiak, N. L., Schaffner, S. F., and Sabeti, P. C. (2015). Data sharing: Make outbreak research open access. *Nature*, 518(7540):477–479.

Acknowledgments

This work was supported in part by the following Brazilian agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), under Grants No. 303772/2017-4 (GLV), No. 312612/2019-2 (AMSM), No. 305305/2019-0 (RO) and No. 167348/2018-3 (AAB), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Appendix A. Single-wave growth models

As mentioned in the main text, if the machine learning module of the Modinterv decides that the selected empirical curve exhibits only one wave, meaning that it can be described by a sigmoidal curve as shown in Fig. 1(a), then the app selects among four mathematical growth models which one best fits the data. These models and their main characteristics are described below.

Appendix A.1. The beta logistic model

For the case of an epidemic with one fully developed wave of infection, we model the time evolution of the cumulative quantity (cases or deaths) by means of the beta logistic model (BLM), which defined by the following ordinary differential equation (ODE) [27, 23]:

$$\frac{dC}{dt} = r\,[C(t)]^q \left[1 - \left(\frac{C(t)}{K}\right)^{\alpha}\right]^p, \tag{A.1}$$

where $C(t)$ is the cumulative quantity at time t. Here we assume that the model parameters $\{r, q, \alpha, p, K\}$ are all constant in time, in which case they can be interpreted as follows: r is the growth rate at the early stage; q controls the initial growth profile and allows to interpolate from linear growth ($q = 0$) to sub-exponential growth ($q < 1$) to purely exponential growth ($q = 1$); the exponent p controls the late-time growth rate, with $p > 1$ implying a slow-decaying polynomial rate, whereas $p = 1$ yields a fast exponential decay; the exponent α controls the degree of asymmetry with respect to the symmetric S-shape of the standard logistic curve; and, finally, K is the final size of the epidemic, meaning that $C(t) = K$, for $t \to \infty$. Equation (A.1) must be supplemented with the initial condition

$$C(0) = C_0, \tag{A.2}$$

for some given value of C_0.

The BLM admits an analytic solution [27] in implicit form given by

$$t = f(C) - f(C_0), \tag{A.3}$$

where

$$f(C) = \frac{C^{1-q}}{r(1-q)}\,{}_2F_1\left(p, \frac{1-q}{\alpha}; 1 + \frac{1-q}{\alpha}; \left(\frac{C}{K}\right)^{\alpha}\right), \tag{A.4}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

with ${}_2F_1(a, b; c; x)$ being the Gauss hypergeometric function. Equation (A.3) describes a sigmoidal curve, whose inflection point is located at the time $t_2 = t_c$ given by

$$t_c = f\left(\frac{Kq}{(q+\alpha p)^{1/\alpha}}\right). \tag{A.5}$$

For completeness, we also quote here the characteristic points t_1 and t_3, corresponding to the points of zero jerk, $\dddot{C}(t) = 0$, of the BLM, which are given by $t_{1,3} = f(Kx_{1,3})$, where [25]:

$$x_{1,3} = \left[\frac{\alpha^2 p + \theta - \alpha\left(p - 4pq \pm \sqrt{\Delta}\right)}{4\alpha^2 p^2 + \theta + 2\alpha p(-1+4q)}\right]^{1/\alpha}, \tag{A.6}$$

with θ and Δ being given by $\theta = 2q(-1+2q)$ and $\Delta = 4pq(-1+2q)+p^2(1-2\alpha+\alpha^2+8\alpha q)$, respectively. One can also compute the point t_4 of maximum jerk, i.e., $\ddddot{C}(t) = 0$, but in this case the expression is rather long and so it is given separately in Appendix C.

The BLM described above is one of the most general growth models, from which many other known models emerge as special cases [23, 27]. For instance, for $q = p = \alpha = 1$ the BLM recovers the Verhulst's logistic model [29], which yields a symmetric sigmoidal curve. However, as most epidemic curves (especially for COVID-19) are not symmetrical, the standard logistic model turns out to be too simple to capture the complexity of a human epidemic dynamics, hence this particular case will not be considered further here. Three relevant particular cases of the BLM are as follows (from the more complex to the simpler): i) for $p = 1$ but $q \neq 1$ and $\alpha \neq 1$, the BLM reproduces the so-called generalized Richards model [6]; ii) if in addition to $p = 1$ one sets $q = 1$ but keep $\alpha \neq 1$, one gets the Richards growth model [20]; and iii) the case $p = 0$ yields the q-exponential model. For epidemic curves with only one wave, the Modinterv app implements separately the BLM and these three particular cases, as illustrated in Fig. 3.

Among the four one-wave growth models implemented in the Modinterv, the BLM is the most complete one, in the sense that it is capable of describing the entire epidemic curve from beginning to end in a rather flexible way. This model is therefore applicable to epidemic curves that are already in the late growth phase, where the growth profile approaches a leveling plateau, whose value is represented by the parameter K (which we recall corresponds to the total number of cases or deaths at the end of the epidemic). This saturation regime is characterized by the parameter p, so that for $p > 1$ the curve approaches the plateau in a slow, subexponential way, whereas only for $p = 1$ does the curve approach the plateau exponentially fast [26]. The BLM is also the most demanding model in terms of the numerical fitting procedure, since we have to determine five parameters, namely: (r, q, α, p, K). In particular, only when the final 'tail' of the epidemic curve is relatively well formed does the BLM converge (i.e., one obtains a reliable estimate of the parameter p.)

For a given empirical curve with only one wave, the Modinterv first applies the BLM. In situations where the BLM does not converge, indicating that the epidemic is still probably in the intermediate phase or at most entering the transition to saturation regime, see Fig. 1(a), the Modinterv then tries to fit the data with the next less complex model, namely the generalized Richards model (GRM) [28], which is described below.

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Appendix A.2. The generalized Richards model

The generalized Richards (GRM) model is obtained from Eq. (A.1) by setting $p = 1$, so its defining ODE is given by

$$\frac{dC}{dt} = r\,[C(t)]^q \left[1 - \left(\frac{C(t)}{K}\right)^{\alpha}\right]. \tag{A.7}$$

The solution of this equation has the same implicit form of Eqs. (A.3) and (A.4), after putting $p = 1$. Hence the characteristics points, t_i, $i = 1, ..., 4$, for the GRM are obtained from the same formulas as for the BLM, only setting $p = 1$.

The GRM is in general suitable for epidemic curves that are well past the inflection point t_c, but which dot not yet display a well formed plateau. In this case, setting $p = 1$ reduces the number of fitting parameters in comparison with the BLM. Nonetheless, for the GRM one still needs to determine four free parameters, namely (r, q, α, K). In the Modinterv, if no good convergence is obtained for neither the BLM nor the GRM for a given empirical data, then the next simpler model, namely the Richards model, is employed, as discussed next.

Appendix A.3. The Richards model

The Richards model (RM) [20] can be obtained as a particular case of Eq. (A.1) after setting $p = q = 1$, which yields the following ODE:

$$\frac{dC}{dt} = rC(t)\left[1 - \left(\frac{C(t)}{K}\right)^{\alpha}\right]. \tag{A.8}$$

Historically, Richards [20] proposed the above model as a modification of Verhult's logistic model, where the new parameter α was introduced so as to allow for asymmetric growth profiles. We recall that the logistic curve, which is recovered after setting $\alpha = 1$ in (A.8), is symmetric with respect to the inflection point. The exponent α thus controls the asymmetry of the curve, i.e., how it deviates from the linear region (around the inflection t_c) and starts to bend towards the plateau. For epidemiological reasons, it is sensible to restrict the values of α to the range $0 < \alpha < 1$ [28], in which case the epidemic curve bends slower towards the plateau than the logistic curve. Note, however, that within the above allowed range of α, the higher the α the sharper the bending.

One major difference of the RM with respect the two previous models (BLM and GRM) is that the RM admits an *explicit solution* in the following form [28]:

$$C(t) = \frac{K}{\{1 + \alpha \exp\left[-\alpha r(t - t_c)\right]\}^{1/\alpha}}, \tag{A.9}$$

where the inflection point $t_2 = t_c$ can be obtained in terms of the initial condition C_0 via the relation: $C_0 = K/[1 + \alpha \exp(\alpha r t_c)]^{1/\alpha}$ or, alternatively,

$$t_c = \frac{1}{\alpha r} \ln\left[\frac{1}{\alpha}\left(\left(\frac{K}{C_0}\right)^{\alpha} - 1\right)\right]. \tag{A.10}$$

One can also obtain explicit expressions for the points of zero jerk for the RM [25]:

$$t_{1,3} = t_c - \frac{1}{r} \ln\left(\frac{\alpha + 3 \pm \sqrt{\alpha^2 + 6\alpha + 5}}{2}\right). \tag{A.11}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

The expression for the point t_4 of maximum jerk for the RM can be obtained from the corresponding result for the BLM, given in Appendix C, after setting $q = p = 1$.

The RM has three free parameters, namely (r, α, K), to be numerically determined from a fitting to the empirical data. It is therefore the simplest model that can describe an asymmetric sigmoidal curve, meaning that it has the least number of free parameters (among the ones implemented in the Modinterv). This particular feature renders the RM a rather robust model. For instance, the RM is mostly appropriate for curves that are still in the intermediate phase, that is, when the cumulative curve is in the near-linear regime, meaning that the curve has just past or is about to pass the inflection point, see Fig. 1(a). Nevertheless, this model can be also suitable for curves that are already in the late growth phase, but for which neither the BLM nor the GRM provide a good fit to the data.

It may happen, however, that the the RM does not provide a good convergence (meaning that the errors in the parameters are too large to be acceptable). This often happens for curves that are in the early growth phase, when the acceleration is still growing. In such cases, because of the paucity of data, the RM has a natural difficulty in estimating when the inflection will occur (in the future). Thus, for the RM to be acceptable we require the current time t_f (i.e., the final time of data) to be greater than the point t_1 of maximum jerk. If, however, we find that $t_f < t_1$, we conclude that the epidemic curve is still in the stage of increasing acceleration, in which case the q-exponential model is more appropriate, as described next.

Appendix A.4. The q-exponential model

As already anticipated, for curves that are in the early growth phase, i.e., when the acceleration is still increasing, the q-exponential model is used. This model is obtained by taking $p = 0$ in equation (A.1), which gives

$$\frac{dC}{dt} = r\,[C(t)]^q\,, \tag{A.12}$$

whose solution is function

$$C(t) = \left[C_0^{1-q} + (1-q)rt\right]^{1/(1-q)} = C_0\,\mathrm{e}_q\left(\frac{rt}{C_0^{1-q}}\right), \tag{A.13}$$

where the function $\mathrm{e}_q(x) = [1 + (1-q)x]^{1/(1-q)}$ is known in the physics literature as the the q-exponential function [19].

The q-exponential model thus has only two free parameters, namely (r, q). The parameter r is the (generalized) growth rate; whereas the parameter q characterizes the dynamical regime of the growth process. Here one has three distinct regimes, namely: i) for $q = 0$ one has a linear growth; ii) for $0 < q < 1$ the curve has a subexponential growth; and iii) if $q = 1$ the growth is purely exponential. In general, a subexponential regime is associated with mitigation measures, while an exponential growth is expected when no containment measure is adopted at the beginning of the epidemic [28].

When the epidemic curve is in the q-exponential regime (i.e., increasing acceleration), it is not possible to make long-term forecasts. At best, one can make short term estimates, like the doubling time, T_d, which corresponds to the number of days (counted from the date of the last data point) that it will take until the number of cases or deaths reaches

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

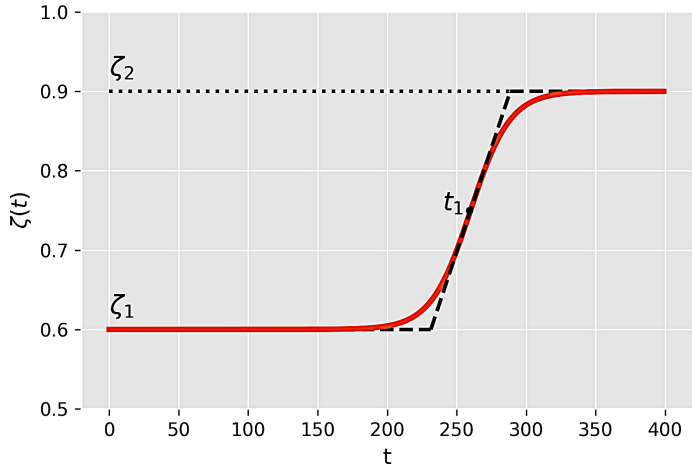


Figure A.12: Time dependence of the generic parameter $\zeta(t)$ of the two-wave model, as defined by the logistic function given in (B.2). The dashed line represents the linear approximation to the logistic function, where the inclined straight line meets the upper and lower horizontal lines at the points $t_1 \pm 2/\rho_1$, where t_1 and ρ_1 are respectively the transition time and rate between the two plateaus; see text.

the double of the present value, assuming that the curve will continue to follow the q-exponential trend. It is not difficult to show [25] that T_d is given by

$$T_d(t) = \left(2^{1-q} - 1\right)\left[t + \frac{C_0^{1-q}}{(1-q)r}\right]. \tag{A.14}$$

Thus, the doubling time in the q-exponential model grows linearly in time for $q < 1$; whereas it remains constant, i.e., $T_d = (\ln 2)/r$, for the purely exponential growth ($q = 1$). The fact that T_d increases linearly in time (for $q < 1$) is a direct manifestation of the subexponential growth. When the Modinterv chooses the q-exponential for a given epidemic curve, it also quotes the value of T_d, in addition to the fitted parameters (r, q), as will be discussed later.

Appendix A.5. Parameter ranges

As a final remark about the above single-wave models, it is important to note here that the parameters r, q, and α are restricted to certain allowed ranges. First we recall, that in the BLM one must have $p \geq 1$, to ensure a polynomial decay (for $p > 1$) of the daily curve after the peak [27], with an exponential decay occurring only in the limit $p = 1$. Similarly, the exponent q is limited to the range $0 \leq q \leq 1$, as $q > 1$ would imply a super-exponential growth which is not justified on epidemiological grounds. Furthermore, it is expected for biological reasons that the asymmetry parameter α should be restricted to the interval (0,1) [30, 28]. We also restrict the values of the growth rate r to the range (0,1), as we observed that values of r outside this interval tend to be an indication of possible over-fitting. In other words, in our numerical implementation of the single-wave models we impose the following range restrictions: $p \geq 1$, $0 < q \leq 1$, $0 < \alpha \leq 1$, and $0 < r < 1$.

Appendix B. Multiple-wave growth models

The two-wave model implemented in the Modinterv is described by the BLM equation (A.1), but where now we assume that *all parameters depend on time*, that is, $r = r(t)$,

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

$q = q(t)$, $\alpha = \alpha(t)$, $p = p(t)$, and $K = K(t)$. To capture the two distinct growth regimes (corresponding to the first and second waves, respectively), we propose that these parameters, here generically represented by the symbol $\zeta(t)$, obey the following logistic-like equation:

$$\frac{d\zeta}{dt} = \frac{\rho_1}{\zeta_2 - \zeta_1}(\zeta - \zeta_1)(\zeta_2 - \zeta), \tag{B.1}$$

whose solution, with the condition $\zeta(t_1) = (\zeta_1 + \zeta_2)/2$, is of the following form:

$$\zeta(t) = \zeta_1 + \frac{(\zeta_2 - \zeta_1)}{2}\left[1 + \tanh\left(\frac{\rho_1(t - t_1)}{2}\right)\right], \tag{B.2}$$

where ζ_1 and ζ_2 represent the corresponding parameter values for the first and second waves, respectively. A schematic of the generic parameter $\zeta(t)$, as defined in (B.2), is shown in Fig. A.12. The parameter t_1 in (B.2) determines the transition time between the first and second wave; whereas the parameter ρ_1 characterises how rapid this transition is, so that the larger the parameter ρ_1 the quicker the transition towards the second-wave regime. Note that the characteristic time scale t_1 and the corresponding transition rate ρ_1 are the same for all parameters. This is justified because an overall change in the epidemic dynamics, brought about, say, by a relaxation of control measures or by a change in the population behavior (or both), is expected to affect simultaneously all epidemiological parameters.

The two-wave model described above can be naturally extended to include an arbitrary number, N, of waves, by assuming that the time dependence of the parameters is as follows [24]:

$$\zeta(t) = \zeta_1 + \sum_{i=1}^{N-1} \frac{(\zeta_{i+1} - \zeta_i)}{2}\left[1 + \tanh\left(\frac{\rho_i(t - t_i)}{2}\right)\right]. \tag{B.3}$$

Considering that the standard BLM has a set of five parameters, i.e., $\zeta = \{q, r, \alpha, r, K, p\}$, it then follows that the BLM for two waves, i.e., with each parameter varying in time as in (B.3), has a total of $7N - 2$ free parameters for a given N. The large number of parameters (as N increases) makes it difficult to devise an automated fitting procedure for the N-wave model with an arbitrary N. For this reason, the current version of the Modinterv is limited to up to two waves only.

Appendix C. Points of maximum jerk

In order to compute the point of maximum jerk for the BLM, as given by Eq. (A.1), one needs to find the roots of the equation $\dddot{\dddot{C}}(t) = 0$. This equation has three roots, namely: i) a point of maximum jerk in the accelerating phase; ii) a point of minimum jerk; and iii) a point of maximum jerk in the decelerating phase. In the classification scheme presented in Sec. 2, we only need the third such root, which was denoted there by the symbol t_4. Using a software (such as Mathematica) for algebraic computation, the

medRxiv preprint doi: https://doi.org/10.1101/2022.01.31.22270192; this version posted February 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

desired root can be explicitly obtained, as follows:

$$
\begin{aligned}
t_4 = 12^{-1/\alpha} \Bigg(& \frac{1}{\text{denom}} \Bigg(-\frac{1}{\sqrt[3]{\text{term1}}} \Big(2i\sqrt[3]{2}\left(\sqrt{3}-i\right)\alpha^2 p \left(\alpha^2 p^3 \left(31\alpha^2 + 2\alpha(63q-22)+13\right)\right. \\
& + \alpha p^2 \left(-14\alpha^3 + \alpha\left(378q^2-289q+35\right) + \alpha^2(8q-7)+26q-14\right) \\
& + p\left(\alpha^4 + \alpha^2\left(-77q^2+56q-5\right) + 13q^2 + 2\alpha\left(189q^3-223q^2+70q-3\right)\right. \\
& \left.\left. + \alpha^3(6-14q) - 14q + 4\right) - 3q\left(6q^2-7q+2\right)\left(3\alpha-7q+3\right)\right)\Big) + 4\left(\alpha^3 p(7p-1)\right. \\
& + 4\alpha p\left(9q^2-7q+1\right) + \alpha^2 p(p(18q-7)+7q-3) + 3q\left(6q^2-7q+2\right)\Big) \\
& + 2^{2/3} i\left(\sqrt{3}+i\right)\sqrt[3]{\text{term1}}\Bigg)\Bigg)^{1/\alpha},
\end{aligned}
$$

where $\text{term1} = \text{term2} + \text{term3}$, and

$$
\begin{aligned}
\text{term2} = {} & 3\alpha^4 p^2 \Big(p^2\left(70q^2-92q+28\right) + p\left(360q^4-1015q^3+845q^2-238q+12\right) \\
& - 3q\left(678q^4-1295q^3+868q^2-231q+18\right)\Big) + \alpha^3 p\Big(2p^2\left(35q^3-69q^2+42q-8\right) \\
& + 9pq\left(30q^4-119q^3+144q^2-70q+12\right) - 54q^2\left(6q^2-7q+2\right)^2\Big) \\
& + 2\alpha^9 p^3\left(154p^3-120p^2+21p-1\right) + 3\alpha^7 p^3\Big(p^3(90q+56) + 3p^2\left(528q^2-399q+32\right) \\
& - 2p\left(602q^2-451q+56\right) + 28q^2-14q-4\Big) + 3\alpha^5 p^2\Big(p^3(70q-46) + p^2\left(540q^3-903q^2+394q-28\right) \\
& + p\left(-1656q^4+2401q^3-1170q^2+238q-26\right) + 12q\left(-42q^3+67q^2-35q+6\right)\Big) \\
& + 3\alpha^8 p^3\left(2p^3(279q-91) - 49p^2(4q-1) + p(49-106q) + 14q-6\right) \\
& + \alpha^6 p^2\Big(70p^4 + 3p^3\left(360q^2-189q-19\right) + 3p^2\left(756q^3-987q^2+216q+49\right) \\
& + p\left(-4228q^3+4971q^2-1596q+90\right) + 27q\left(6q^2-7q+2\right)\Big),
\end{aligned}
$$

$$
\begin{aligned}
\text{term3} = {} & 3\sqrt{3}\Big(-\alpha^6 p^2\Big(6\alpha^3 p^3 + \alpha^2 p^2(18q-7) + 2\alpha p\left(9q^2-7q+1\right) \\
& + q\left(6q^2-7q+2\right)\Big)^2\Big(p^4\Big(25\alpha^6 + \alpha^4\left(3193q^2-2534q+482\right) + \alpha^2\left(2473q^2-2114q+425\right) \\
& + \alpha^3\left(8232q^3-9554q^2+3990q-644\right) + 2\alpha^5(217q-88) + 4\alpha(56q-29) + 4\Big) \\
& - 2p^3\Big(7\alpha^6 + 7\alpha^4\left(77q^2-61q+12\right) - 4\left(28q^2-29q+7\right) - \alpha^3\left(1249q^3-1533q^2+365q+42\right) \\
& + \alpha\left(-2473q^3+3171q^2-1142q+84\right) + \alpha^2\left(-12348q^4+19108q^3-9667q^2+1715q-63\right) \\
& + \alpha^5(103q-42)\Big) + p^2\Big(\alpha^6 + 2473q^4 - 4228q^3 - 2\alpha^4\left(103q^2-84q+15\right) + 2284q^2 \\
& - 18\alpha^3\left(168q^3-201q^2+70q-6\right) + \alpha^2\left(-8471q^4+13076q^3-7154q^2+1680q-159\right) \\
& + 2\alpha\left(12348q^5-23885q^4+15344q^3-3213q^2-126q+54\right) - 336q - 28\Big) \\
& + 4pq\left(6q^2-7q+2\right)\Big(343q^3 - 54\alpha\left(9q^2-7q+1\right) - 396q^2 - 9\alpha^2(7q-3) + 63q + 27\Big) \\
& - 108q^2\left(6q^2-7q+2\right)^2\Big)\Big)^{1/2},
\end{aligned}
$$

and

$$
\text{denom} = q(2-7q+6q^2) + 2p(1-7q+9q^2)\alpha + p^2(-7+18q)\alpha^2 + 6p^3\alpha^3.
$$