A non-homogeneous Markov early epidemic growth dynamics model. Application to the SARS-CoV-2 pandemic

Néstor Ruben Barraza[1,*], Gabriel Pena, Verónica Moreno

Universidad Nacional de Tres de Febrero, Argentina

ARTICLE INFO

ABSTRACT

This work introduces a new markovian stochastic model that can be described as a non-homogeneous Pure Birth process. We propose a functional form of birth rate that depends on the number of individuals in the population and on the elapsed time, allowing us to model a contagion effect. Thus, we model the early stages of an epidemic. The number of individuals then becomes the infectious cases and the birth rate becomes the incidence rate. We obtain this way a process that depends on two competitive phenomena, infection and immunization. Variations in those rates allow us to monitor how effective the actions taken by government and health organizations are. From our model, three useful indicators for the epidemic evolution over time are obtained: the immunization rate, the infection/immunization ratio and the mean time between infections (MTBI). The proposed model allows either positive or negative concavities for the mean value curve, provided the infection/immunization ratio is either greater or less than one. We apply this model to the present SARS-CoV-2 pandemic still in its early growth stage in Latin American countries. As it is shown, the model accomplishes a good fit for the real number of both positive cases and deaths. We analyze the evolution of the three indicators for several countries and perform a comparative study between them. Important conclusions are obtained from this analysis.

1. Introduction

The coronavirus disease (COVID-19) transmitted by the SARS-CoV-2 appeared in China by the end of 2019. More than 17,000,000 of cases and 675,000 deaths were reported since then. The lack of a vaccine and an effective treatment for this disease forced the governments to adopt lockdown actions in order to protect the population and to slow down the spread of the outbreak. Despite those actions, health systems in several countries resulted overwhelmed. Many mathematical models were developed in order to predict the behavior of the outbreak in several ways, most of them based on the well known SIR (Susceptible-Infectious-Recovered) models. We present in this work a different approach having the advantage of a rapid and clear interpretation of its parameters. We obtain this way a quite useful new indicator, the mean time between infections (MTBI).

Modeling contagion has attracted the interest of statisticians for decades. These models are useful to model the spread of contagious diseases in a population or plantation. A well-known example is the Polya urn model, where balls of different colors are extracted from an urn in such a way that when a given ball is extracted, not only is it put back but also a certain number of balls of the same color is added. The probability of getting a ball of that color in the next drawing is thus increased, modeling a contagious process. Contagion modeling has been already applied in many areas of engineering, see for example [1–3]. The Polya stochastic process is obtained as the limit from the Polya urn model in a similar way as the Poisson process is obtained from a Bernoulli process.

As it happens with the homogeneous and non-homogeneous Poisson processes, the Polya stochastic process can be described by the Pure Birth equation (see chapter 17 of [4]). The main difference between these two types of Pure Birth processes is that, in the Polya process, the event rate is not just a function of time but also of the number of individuals in the population. There are also cases where the event rate depends only on the number of previous events but not on time, like the Yule process. However, since the probability of births rises due to an increase in the population while the probability of an individual birth remains constant, it does not describe a true contagion process.

The stochastic mean of the Polya process is a linear function of time, as will be shown. This characteristic makes the Polya process

---

* Corresponding author
E-mail address: nbarraza@untref.edu.ar (N.R. Barraza).
[1] The author is also with the School of Engineering of the University of Buenos Aires.

unsuitable for application in many real cases of disease growth dynamics and engineering. Hence, we propose a different model based on a Pure Birth process with an event rate that, like Polya's, depends on both the elapsed time and the number of previous events, but with a different functional form. We obtain this way a mean number of events that is a nonlinear function of time, with either an increasing or decreasing first derivative provided a given parameter is greater or less than one.

An important application of our proposed model is the study of the spreading of diseases. We obtain this way a process controlled by two competitive phenomena, infection and immunization, i.e. two opposing forces. Hence, we get a model with two parameters: the transmission and immunization rates. The mean value function shows different behavior according to whether the ratio between infection and immunization rates is greater or lower than one. In the former case, its second derivative is positive and the mean value curve is convex, whereas on the latter case its second derivative is negative and the resulting curve is concave. The limit case of infection/immunization ratio $\gamma/\rho = 1$ is also possible, resulting on a mean value function that grows linearly over time. The evolution of the parameters shows how well the pandemic is being controlled, either by lowering the infection rate via isolation and quarantine measures or by increasing the immunization rate with a vaccination campaign. We obtain this way three useful indicators: the infection/immunization ratio, the immunization rate and the mean time between infections (MTBI).

Our model is quite suitable to be applied to early epidemic growth dynamics, where exponential models fail (this matter is discussed in [5]). Since ours falls in the category of subexponential models, the results presented in this article support the idea developed in the cited reference. Another advantage of our approach is that, as a subset of the subexponential cases, we are able to model cases with linear and sublinear (concave) cumulative case incidence curves, as it will be shown.

As an interesting and quite important application, we apply our model to the recent SARS-CoV-2 pandemic as an alternative to the SIR models. Our theoretical values match real data, which allows us to assess the evolution of the outbreak and to predict its trend a few days ahead. The parameters can be easily estimated through simple methods, such as Least Squares, applied either to positive cases or deaths. We show applications to data from several countries with good agreement.

This paper is organized as follows: Motivation and related work are exposed in Section 2, the proposed model is presented in Section 3, applications to the SARS-CoV-2 pandemic are developed in Section 4 and some discussions on the results we obtained are presented in Section 5. The final conclusions are exposed in Section 6.

2. Motivation and related work

2.1. Motivation

Our main motivation is to obtain a model that describes an epidemic outbreak at its first stage, before it reaches the inflection point in the case incidence curve, which is useful to monitor how contagion is spreading out. This first stage corresponds to an *R* naught greater than one. Our model is inspired in the Polya-Lundberg process, which comes from the contagious urn model as explained in the previous section. Since the mean value function of the Polya-Lundberg process is a linear function of time (see Appendix B), we introduce a modification in the event rate in order to get a mean value function that grows subexponentially with either positive or negative concavity as we observe in the early epidemic growth curves usually reported.

2.2. Related work

There exists a wide variety of models that can be used to describe the evolution of an epidemic. Main standard is held by the so called compartmental models, i.e., the family of SIR based models (SIR, SEIR, SIRS, etc. See for instance [6–9]). These models consist of a set of differential equations that take into account transitions between different compartments. The first classical SIR model was porposed by Kermack and McKendrick [10].

Due to the recent COVID-19 coronavirus pandemic, most of the scientific community is dedicated to study its behavior, both by using SIR based models and introducing new ones. In [11], a stochastic term is introduced in the system of differential equations to simulate noise in the detection process. In [12] the authors use an ARIMA based method to correct errors arising from the delay between the day the sample is taken and the day a diagnosis is made, which contaminate the time series. The impact of statistical assumptions within the parameter inference for the phenomenological models is analyzed in [13].

More recent approaches to the problem of assessing the behavior of an epidemic include computational simulation and machine learning. Agent-Based Models (ABM) perform simulations of a population made by a large number of individual agents who can take decisions based on some criteria. In [14] the authors propose a SEIR ABM which simulates the behavior of a society under different scenarios in order to forecast both the economic and epidemic impact of possible government measures. A comparison between ABM and classical compartmental models can be found in [15]. In [16,17] some well-known machine learning and time series algorithms (like ARIMA and SVR) are used to learn from a present dataset and forecast the case incidence curve up to the following six days. Combinations of SIR models embedded with some random components was proposed in very recent articles, see for instance [18–20]. A theoretical framework to model epidemics by a non homogeneous birth-and-death process was proposed in [21].

An extensive analysis of epidemics at their early stage was performed in [5]. In this study, the authors propose a phenomenological model (see also [22,23]) and analyze the application of SIR based models with either exponential and subexponential behavior. Our approach is quite different. On one hand, our model is stochastic whereas theirs is deterministic. On the other hand, we
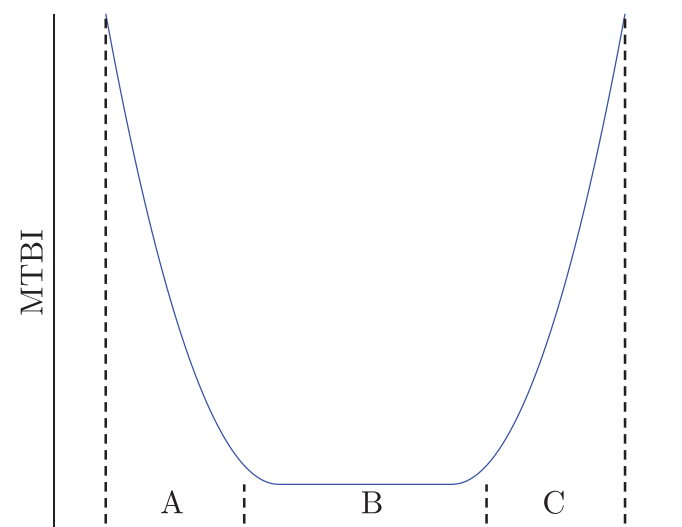
Fig. 1. Expected shape of the MTBI indicator. Different portions show the following: A-The epidemic is spreading out, MTBI decreases since infections occur more often. B- Infections occur at a constant rate. C- MTBI increases since infections occur less often.
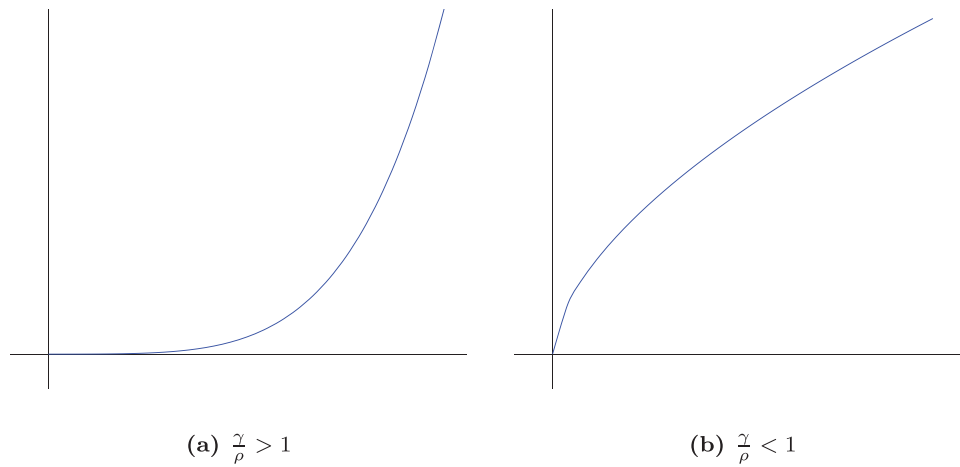
(a) $\frac{\gamma}{\rho} > 1$ (b) $\frac{\gamma}{\rho} < 1$

Fig. 2. Different behaviors of Eq. (11) depending on values of $\frac{\gamma}{\rho}$.

obtain our contagion model as a special case of a very general and well-known probabilistic model, the Pure Birth process. It is interesting to compare the differential equations the cited authors arrive with ours, which will be done in a future publication. Furthermore, and unlike them, we are also able to model cumulative case incidence curves that grow with negative concavity.

3. The proposed model

The model we propose is based on a Pure Birth process. These processes describe the evolution of a population where individuals can only be born. We are then modeling the epidemic spread as births in a population, where every birth corresponds to a new infection case. Our novel approach consists of finding a proper incidence rate that describes the contagion phenomenon. We develop next the fundamentals of Pure Birth processes and our proposal.

3.1. Pure birth processes

We propose to get the probability of having r infections in a given time t from a Pure Birth stochastic differential equation. Pure Birth processes describe the behavior of a population where individuals can only be born and are not allowed to die. All the individuals are assumed to be identical. The probability of having r individuals in the population at a given time t, $P_r(t)$, is given by the solution of the following differential equation, see [4,24]:

$$P_r'(t) = -\lambda_r(t)P_r(t) + \lambda_{r-1}(t)P_{r-1}(t), \quad r \geq 1,$$
$$P_0'(t) = -\lambda_0(t)P_0(t). \tag{1}$$

Assuming we have 0 individuals at time 0, we impose the following initial conditions on Eq. (1): $P_0(0) = 1$ and $P_r(0) = 0$ for $r \geq 1$. To see that $P_r(t)$ is well defined see Appendix A.

The process given by Eq. (1) is also markovian, where the number of individuals corresponds to the state of the system, $S_r(t)$. The only transitions allowed[2] are $S_r(t) \rightarrow S_r(t + dt)$ and $S_r(t) \rightarrow S_{r+1}(t + dt)$ with probabilities:

$$P(S_r(t) \rightarrow S_{r+1}(t + dt)) = \lambda_r(t)\,dt,$$
$$P(S_r(t) \rightarrow S_r(t + dt)) = 1 - \lambda_r(t)\,dt. \tag{2}$$

The dependence of $\lambda_r(t)$ on t defines the type of process. If $\lambda_r(t)$ is a function of t only or a constant, then the process is non-homogeneous or homogeneous respectively, with independent increments in both cases. If $\lambda_r(t)$ is also a function of r, the process has dependent increments.

---

[2] Actually, other transitions have probabilities that are higher order infinitesimals.

From Eq. (1), the probability of having no births in a certain time interval greater than $t - s$ given the population has r individuals by the time s is given by the well-known exponential waiting time:

$$P(\text{no births in } T \geq t - s) = e^{-\int_s^t \lambda_r(t)dt} \quad t \geq s. \tag{3}$$

The mean number of infections in a given time is

$$M(t) = \sum_r r\, P_r(t). \tag{4}$$

Under certain conditions (see Appendix B to verify those are satisfied by our model) we have the following differential equation for the mean value:

$$M'(t) = \sum_{r=0}^{\infty} \lambda_r(t)P_r(t). \tag{5}$$

Depending on the proposed function for $\lambda_r(t)$, the mean number of failures may or may not be easily obtained.

Our formulation allows us to calculate the mean time between infections (MTBI), this is, the mean time between births in a Pure

Table 1
Parameter values and R^2 coefficient for the incidence curves.

Country	ρ	$\frac{\gamma}{\rho}$	R^2
Argentina	0.110403	4.72391	0.954157
Brazil	0.3629236	4.0432285	0.98094
Chile	0.3139428	3.8070482	0.975592
China	0.310895	4.57184	0.944164
Germany	0.181481	5.0317	0.945185
Italy	0.573406	3.52942	0.932463
Spain	0.206731	4.93668	0.943004
United Kingdom	0.0274503	12.5826	0.975626
USA	0.0494053	9.53863	0.967517

Table 2
Parameter values and R^2 goodness of fit for the curve of deaths.

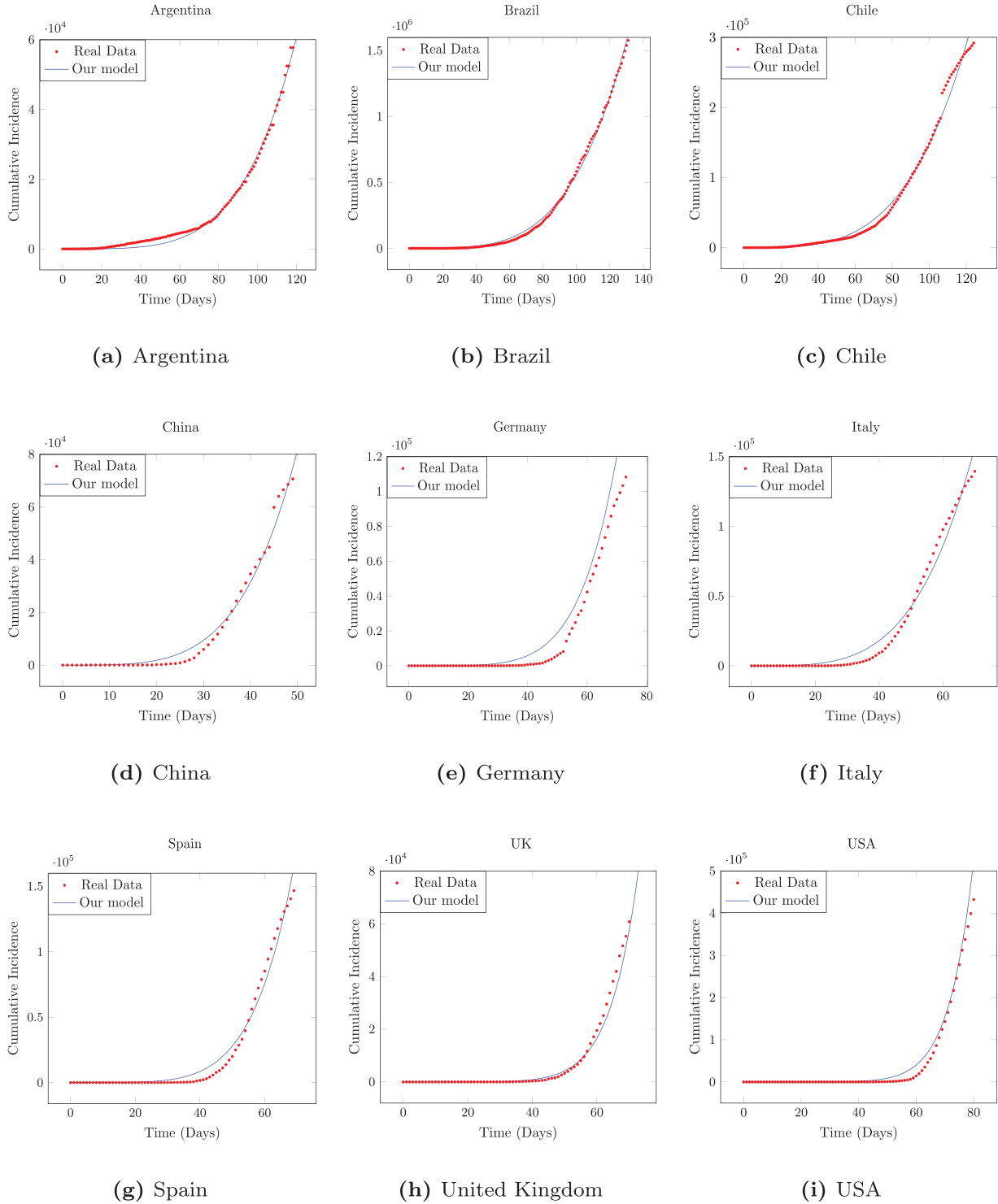| Country	ρ	$\frac{\gamma}{\rho}$	R^2
Argentina	0.166995	2.69049	0.962536
Brazil	1.26084	2.43167	0.9591606
Chile	0.07983	4.62764	0.9798674
Germany	0.407652	3.36218	0.958036
Italy	1.75157	2.40184	0.843577
South Korea	1.25081	1.35826	0.97169
Spain	2.17936	2.37205	0.916289
UK	0.258118	4.46389	0.960576
USA	0.153287	5.74097	0.972502

Fig. 3. Cumulative case incidence.

Birth process. From Eq. (3), we can predict the mean time between infections (MTBI) after r infected individuals were detected by the time s using a model with incidence rate $\lambda_r(t)$ as:

$$MTBI(s, r) = \frac{1}{Z} \int_s^\infty \lambda_r(t) \, t \, e^{-\int_s^t \lambda_r(t)dt} \, dt - s. \tag{6}$$

Here, Z is a normalizing constant to consider cases where the probability of having no infections in an infinite time interval is greater than zero and is given by Eq. (7) (see Appendix C for a full

deduction):

$$Z = 1 - e^{-\int_s^\infty \lambda_r(t)dt}. \tag{7}$$

Eq. (6) gives the expected time to the next infection given that by the time s the infected population consists on r individuals, this is, the mean time between two consecutive detections.

This indicator is, in most cases, U-shaped (see Fig. 1). An exceptional case is also shown in Fig. 4, and will be discussed later). It is expected to decrease at first, showing the acceleration of the spread (infections occur more often). When the cumulative case
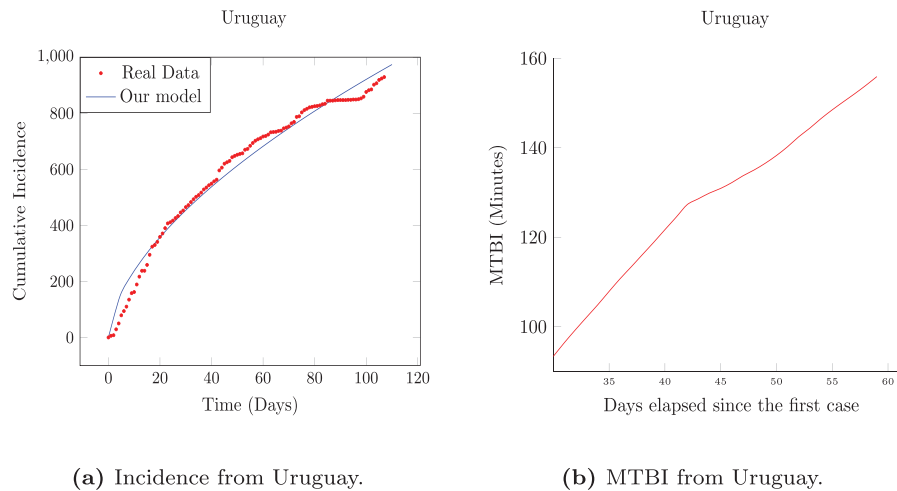
(a) Incidence from Uruguay.

(b) MTBI from Uruguay.

Fig. 4. The Uruguayan case.

incidence curve reaches the inflection point and the epidemic starts to mitigate, MTBI shows a flattening portion (infections occur at a constant rate). Then, due to the deceleration of the epidemic (infections occur less often) MTBI tends to get larger again. This curve resembles the bathtub curve, well known in Reliability Engineering, which is also obtained when this class of models is applied to reliability studies, see [25].

3.2. Proposed incidence rate

We propose a two-parameter incidence rate given by the following expression:

$$\lambda_r(t) = \rho \frac{1 + \frac{\gamma}{\rho} r}{1 + \rho t}. \tag{8}$$

The parameters involved in Eq. (8), indicate how fast the outbreak is spreading. We can see that the γ parameter is related to the strength of the contagion effect, since it is a factor of the number of infections, and the ρ parameter is related to the outbreak mitigation, either by natural immunization or due to external measures. That is the way how these two parameters, once estimated, are useful to monitor the disease progress and the impact of actions taken by health institutions. These actions can affect either the γ parameter via the population solation and quarantine, or the ρ parameter by, for instance, vaccination campaigns. As we will show, the ratio between both rates is the exponent of time t on the mean value function (Eq. (11)), and it determines the curve's concavity, provided it is greater or less than one. We can compare our proposed incidence rate with other formulations by rewriting Eq. (8) as follows:

$$\lambda_r(t) = \frac{\rho}{1 + \rho t} + \frac{\gamma r}{1 + \rho t}. \tag{9}$$

The numerator of the second term in Eq. (9) can be compared with the numerator of the incidence rate analyzed in [26] (Eq. 1: $\lambda(t) = \beta c \frac{I(t)}{N(t)}$, where $I(t)$ is the number of infections, β the transmission probability per contact and c the contact rate) by replacing $I(t)$ with r and βc with γ; our γ parameter can be interpreted as the transmission rate in the same way. In that proposal, the $N(t)$ denominator corresponds to the total population, since the authors consider the per capita incidence rate. As an interesting remark, it can be seen that it also coincides with that of the Polya-Lundberg process. The first term in Eq. (9) is equal to $\lambda_0(t)$, hence it can be considered as the initial incidence rate. On the other hand, following the usual behavior of ρ, this term rapidly vanishes with time.

Being the ratio $\frac{\gamma}{\rho}$ the coefficient of r in the contagion rate (Eq. (8)), it takes into account both the infection and immunization rates, so it can be directly related to the well-known effective reproduction number R_t, an indicative that is often used to qualify the spreading and magnitude of an epidemic. The denominator in Eq. (8) shows how the outbreak is getting weak as a function of time and is an indicator of how this Effective-reproduction-number-like parameter goes down.

The evolution of these two parameters, the ratio $\frac{\gamma}{\rho}$ and ρ as a function of time, is indicative of how well the epidemic is being controlled by actions taken from health institutions. We expect a strong decrease in $\frac{\gamma}{\rho}$, and in consequence, a strong increase in ρ.

Our definition of the outbreak parameters deserves a detailed analysis. The well-known definition of the basic reproduction number R_0 assumes all the population is susceptible and no immunization is present. This would imply ρ to be zero; in that case, Eq. (8) reduces to

$$\lambda_r(t) = \lambda r. \tag{10}$$

which is the event rate of another classic stochastic model, called the Yule process. In that case, the event rate given by Eq. (10) grows linearly with the population size, which results in an exponential expression for the mean value function (see [27] for more details about the Yule process). This means that, if we allow our ρ to eventually be zero, the Yule process becomes a special case of ours, hence exponential growth can also be modeled. However, as it can be seen in Table 1, even the worst case scenarios (very sharp incidence curves) differ from exponential growth, being this another empyrical confirmation of Chowell's thesis presented in [5].

Table 3
MTBI. Minimum values reached. Countries marked with * have not reached a minimum by the time this article is being written; the present value is then indicated.

Country	Minimum MTBI [Minutes]
Argentina*	0.7
Brazil*	0.032
Chile	0.14
China	0.19
Germany	0.14
Italy	0.15
South Korea	1.21
Spain	0.10
United Kingdom	0.17
USA	0.024

(a) Argentina

(b) Brazil

(c) Chile

(d) Germany

(e) Italy

(f) South Korea

(g) Spain

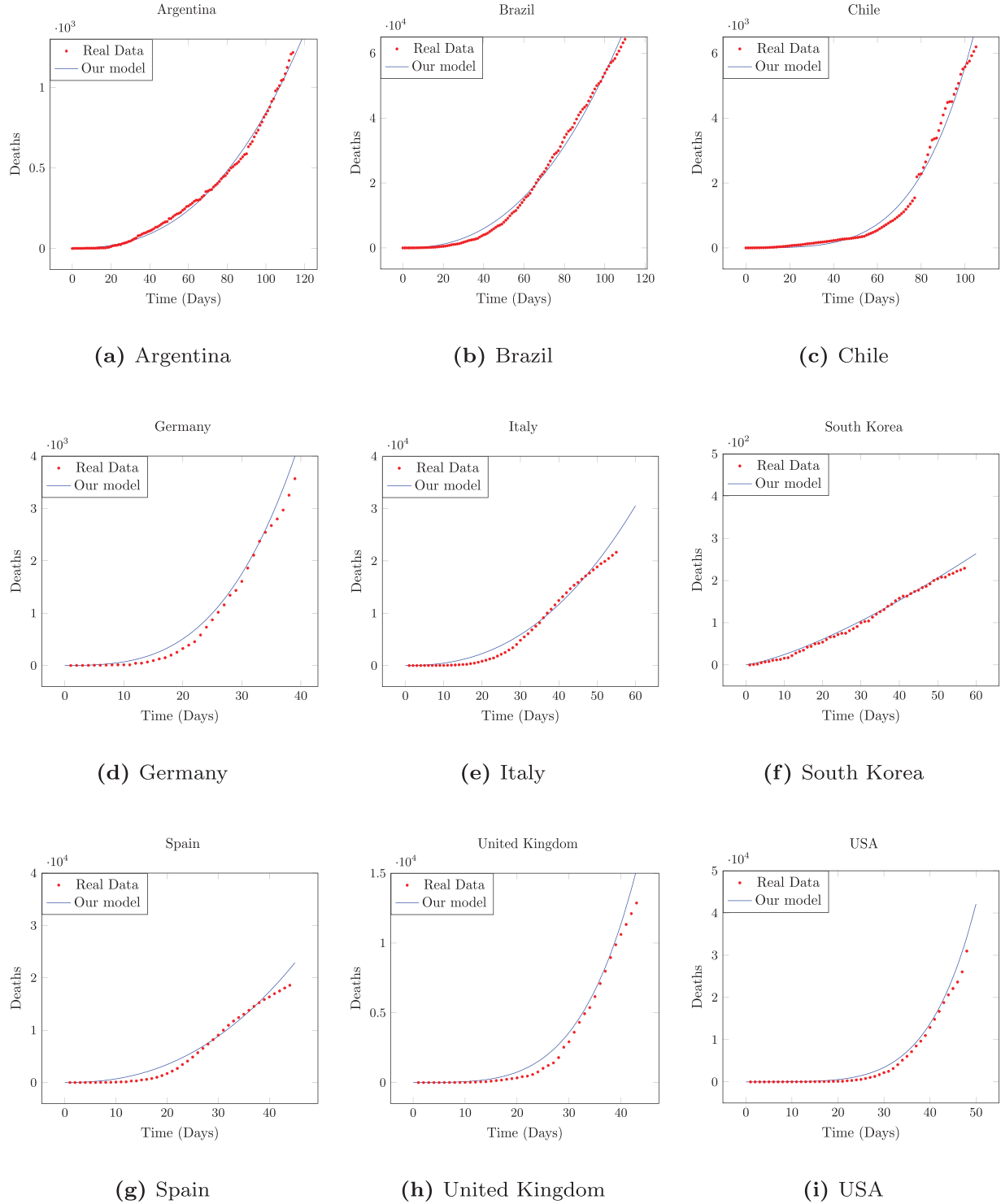**(h)** United Kingdom

(i) USA

Fig. 5. Cumulative number of deaths.

It should be remarked that our proposed rate (Eq. (8)) is different from that of the Polya-Lundberg process: $\lambda_r(t) = \rho \frac{\frac{\gamma}{\rho} + r}{1 + \rho t}$, (see for example [1]), which leads to a mean value function that depends linearly on time when replaced in Eq. (5), as demonstrated in Appendix B.

3.3. The case incidence curve

From the incidence rate (Eq. (9)), we are able to get the exact solution of $P_r(t)$ (see Appendix A) and therefore we can show

that $M(t)$ is finite and Eq. (5) is valid (see Appendix B). Solving Eq. (5) for our model we get the following expression for the mean (see Appendix B):

$$M(t) = \frac{\rho}{\gamma}((1 + \rho\, t)^{\frac{\gamma}{\rho}} - 1). \tag{11}$$

As seen from Eq. (11), the mean value obtained from our model is a non-linear function of time with a positive or negative concavity whether $\frac{\gamma}{\rho}$ is greater or lower than one, as shown in Fig. 2. This expression is the functional form that describes the disease

spreading at the early stage and fits well the COVID-19 data for both total infections and total deaths, as it will be shown next.

3.4. The mean time between infections

For our proposed model, it is straightforward to obtain an expression for the MTBI by inserting Eq. (8) into Eq. (6). This leads to Eq. (12):

$$MTBI(r, t) = \frac{(1 + \rho\, t)}{\gamma\, r}. \tag{12}$$

Replacing r by the mean number of infections given by Eq. (11) we get Eq. (13)

$$MTBI(t) = \frac{1}{\rho} \frac{(1 + \rho\, t)}{(1 + \rho\, t)^{\frac{\gamma}{\rho}} - 1}, \tag{13}$$

which is a useful conditional expression, as shown in Appendix C.

4. Applications to the SARS-CoV-2 disease

In this section, we show the indicators obtained from several countries' reports. Data were taken up to early July 2020 from https://ourworldindata.org/coronavirus-source-data. For Asian and European countries, and also the USA, only data from the early epidemic stages were considered for analysis, this is, up to the inflection point. Real data and fitted curves are shown in Figs. 3 and 5.

We fit Eq. (11) to the incidence and death curves of the recent CoV-2 pandemic in order to obtain the three mentioned indicators.

4.1. Infection cases

The case incidence curves are shown in Fig. 3, while Table 1 shows the parameters. The coefficient of determination R^2 for measuring the goodnes of fit is also shown there. The model parameters were estimated by the Least Squares method, the resulting nonlinear equations being solved by the Levenberg-Marquardt algorithm.

In order to show that the proposed model fits well not just positive concavities as shown before, we consider the data from Uruguay, which presents a concave cumulative case incidence curve from the very beginning of the epidemic. Real data and the fitted curve are shown in Fig. 4a. Corresponding parameter values are $\rho = 459.363858194007$ and $\frac{\gamma}{\rho} = 0.586214262477143$. Another interesting case can be seen in Fig. 5f, which shows the number of deaths in South Korea. The curve is almost a straight line, and the corresponding $\frac{\gamma}{\rho} \approx 1$ as expected. Linear behavior is another limit case considered within our model.

4.2. Fatality cases

Cumulative number of deaths curves are shown in Fig. 5. The parameters and the coefficient of determination are shown in Table 2.
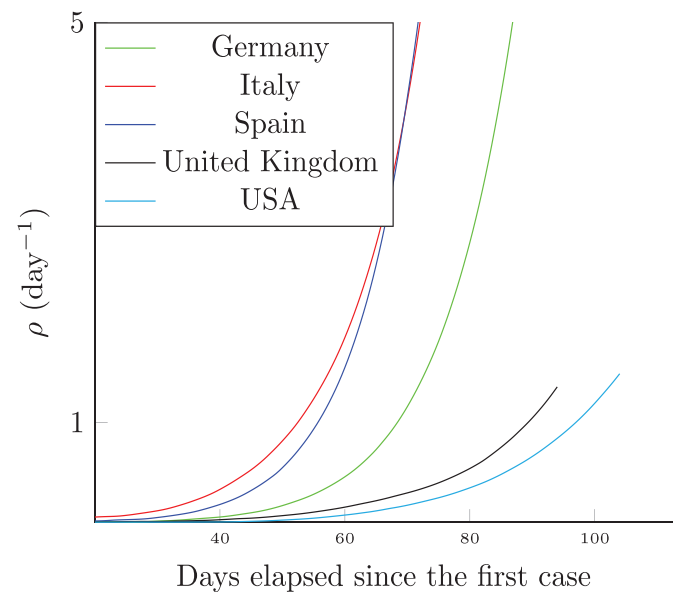
4.3. Infection/immunization ratio over time

In this section we show the evolution of the $\frac{\gamma}{\rho}$ parameter obtained from our model as a function of time, for the analyzed countries. In order to obtain the evolution of this parameter over time, we perform several estimation runs with successive subsets of the dataset and record each estimate.

This picture is indicative of how effective the measures taken by different countries have been, or how fast the outbreak is reaching the mitigation stage. As the transmission rate approaches the im-
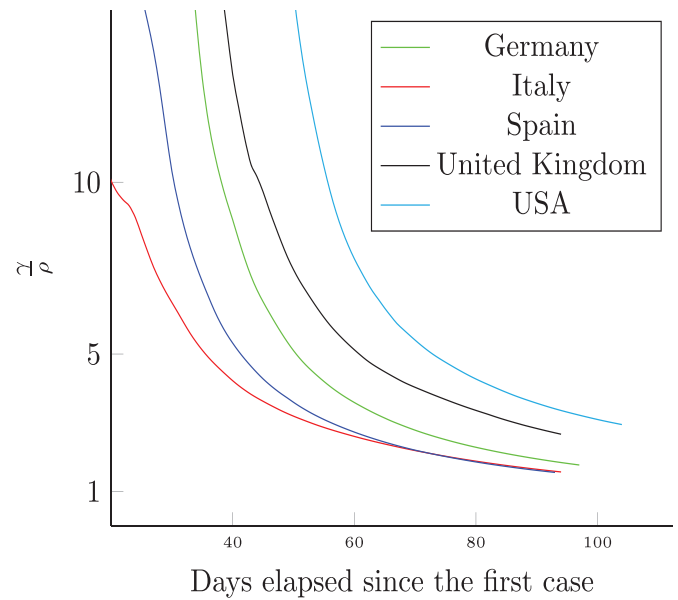
(a) Infection/immunization ratio $\frac{\gamma}{\rho}$ as a function of time.

(b) Immunization rate ρ as a function of time.

Fig. 6. Variation of the model parameters over time.

munization rate, this parameter reaches the value of 1 as a limiting case. Curves are depicted in Fig. 6a.

4.4. Immunization rate over time

As previously discussed, the strength force that contains the outbreak is determined by the immunization rate ρ parameter, and we expect it to increase due to actions taken by governments. Then, we analyze the evolution of this parameter in the same way
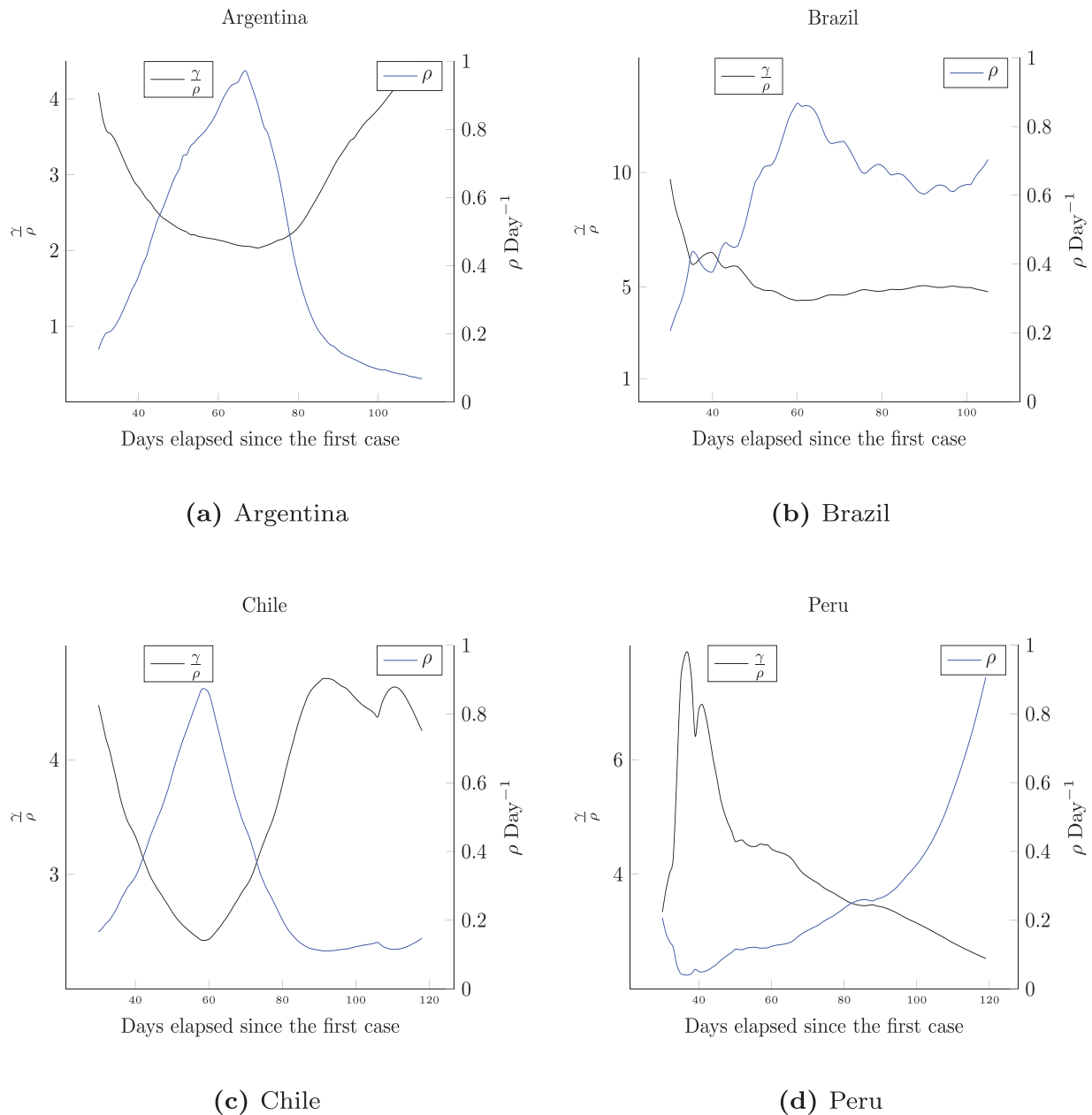
(a) Argentina

(b) Brazil

(c) Chile

(d) Peru

Fig. 7. Parameters as a function of time.

as it was done for $\frac{\gamma}{\rho}$ in the previous section. Values of ρ as a function of time are depicted in Fig. 6b.

4.5. Latin american countries

It is interesting to analyze the current behavior of the parameters for Latin American countries, since they are still at the first stage of the pandemic and where, with the exception of Brazil, the lockdown was relaxed and turned strict again more than once. Curves are depicted in Fig. 7.

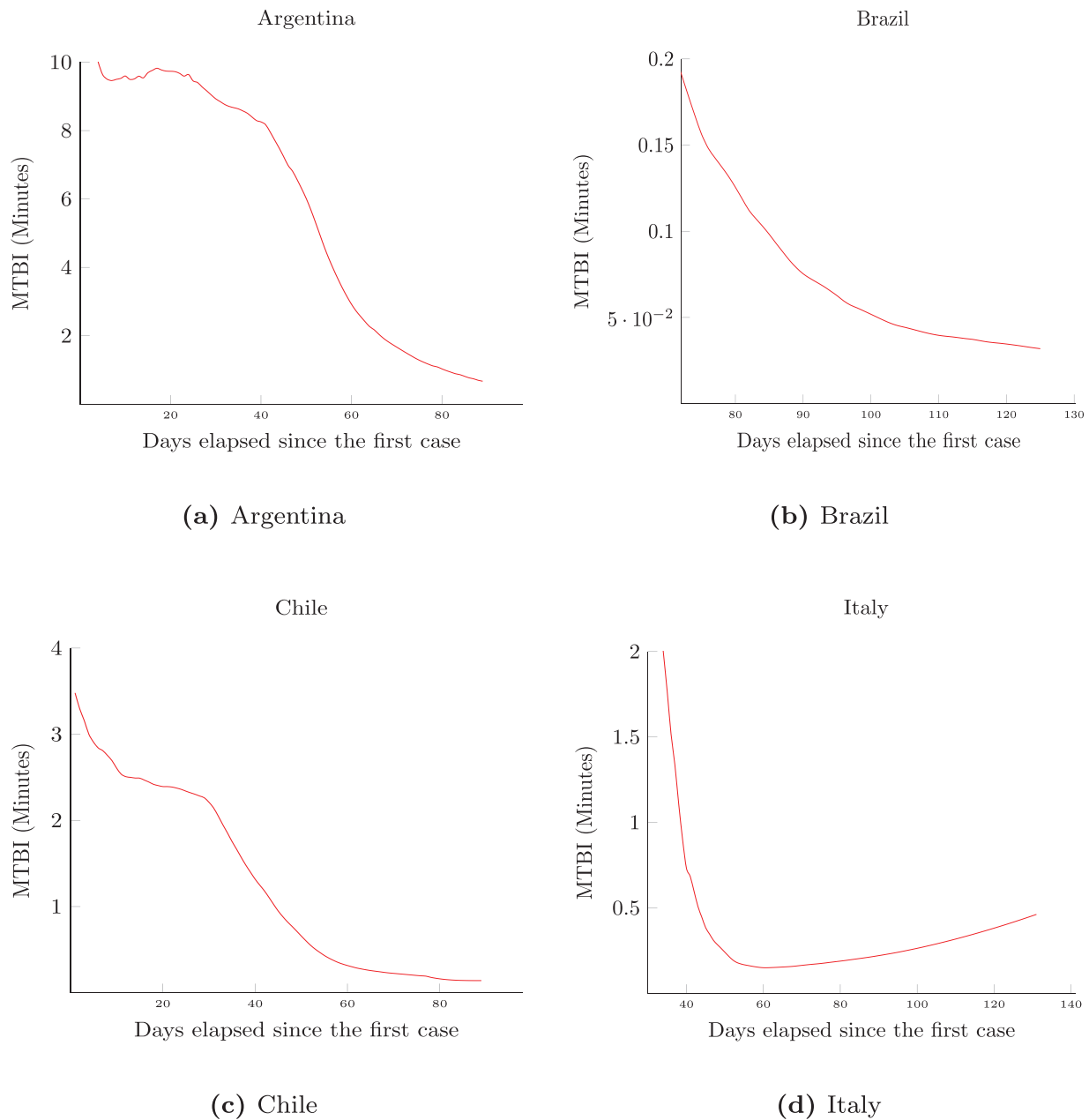### 4.6. Mean time between infections

Values of the MTBI obtained from Eq. (13) for Argentina and Italy are shown in Fig. 8. MTBI decreasing over time reflects the

pandemic is still spreading out and the inflection point has not been reached yet. This indicator is useful in order to estimate the moment when the curve will start to flatten.

In Fig. 9 we depict the MTBI for China in order to see the flattening section of this indicator as the cumulative cases of incidence reaches its inflection point.

By the time we are writing this report (early July 2020), contrarily to Europe and the USA, Latin American countries have not yet reached the inflection point in the cumulative incidence curve, which can also be seen through the MTBI characteristics. This inflection point in the cumulative incidence curve corresponds to a minimum in the MTBI indicator, as seen in Fig. 8 for Italy. It is interesting to observe the minimum values reached by several countries and compare those to the actual values of some Latin American ones. These are shown in Table 3.

(a) Argentina

(b) Brazil

(c) Chile

(d) Italy

Fig. 8. Mean time between infections (MTBI).

5. Discussion

The well goodness of fit obtained for several countries shows that the spreading is governed by the same law. Although with different parameter values, the cumulative case incidence curve follows the same mathematical expression. Those parameters indicate the level of contagion and the effectiveness of the actions taken by governments. The model fits rather well the number of infected and deaths in the population previous to the inflection point, at the early pandemic growth stage, though its prediction beyond four or five days ahead is generally too pessimistic and overestimates the actual data.

The MTBI minimum values indicated in Table 3 show an interesting result. Countries from Europe and China present similar values, between 0.1 and 0.2 minutes, i.e. between 5 and 10 infections per minute. It must be remarked that this value does not corre-

spond to the same vector or host: the Pure Birth process considers the population to be even. Hence, the MTBI value is the time elapsed between two consecutive infections, despite the agent that makes the infection. European countries and China have similar minimum values because those countries implemented quarantine actions rather late. Contrarily, Latin American countries like Argentina and Chile have implemented a lockdown at the beginning of the outbreak, reaching neither their minimum nor the values of the other regions yet. Brazil and USA deserve a more detailed analysis. Since these countries have not implemented a strict lockdown, the pandemic has strongly spreaded between their population and, in consecuence, the case incidence curve suddenly increased. This also results in a sudden decrease in the MTBI up to quite low values, between 26 and 41 infections per minute. We have seen then that our indicator shows similar values for countries that have implemented similar actions. An open source
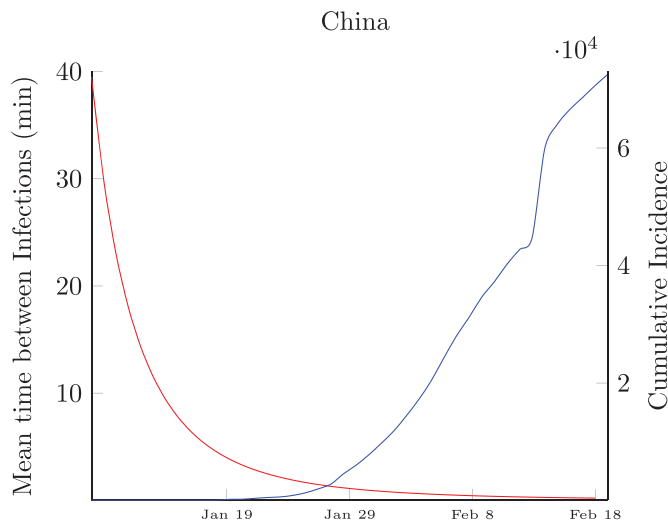
Fig. 9. Mean time between infections and cumulative cases incidence for China.

software application developed in R that implements our proposed model, and which was used to obtain the shown curves, is available at [28].

6. Conclusion

A new early epidemic growth dynamics model was proposed. It is a slight variation of the Polya stochastic process, which in itself is a limiting case of the Polya urn model for contagion. Like the Polya process, the proposed model is a Pure Birth process with a different functional form for the event rate. In opposition to the Polya process, the new proposal results in a stochastic mean that depends non-linearly over time. Applications to the SARS-CoV-2 pandemic still in the first outbreak stage in several countries were shown. Our model is based on two competitive processes: transmission and immunization. Those are taken into account through the corresponding rates as factors. These parameters are easily estimated by a curve fitting procedure. The mathematical expression obtained from the proposed model fits well the incidence and deaths curves for several countries, even those with a concave shape as it is the case of Uruguay. Our model provides a quick and easy way to obtain indicators that can be used to monitor and evaluate the impact of measures taken in order to control the outbreak: the infection/immunization ratio, the immunization rate and the mean time between infections.

Declaration of Competing Interest

Authors declare no conflict of interest related to this article.

CRediT authorship contribution statement

Néstor Ruben Barraza: Conceptualization, Formal analysis, Investigation, Supervision, Writing - original draft. **Gabriel Pena:** Methodology, Investigation, Software, Writing - review & editing. **Verónica Moreno:** Formal analysis, Investigation, Methodology, Writing - review & editing.

Acknowledgments

Appendix A. Pure Birth processes

It can be easily proved that Eq. (1) has a unique solution and is given by (see also [29]):

$$P_r(t) = \int_0^t \left[e^{-\int_x^t \lambda_r(v)dv} \right] \lambda_{r-1}(x) P_{r-1}(x) dx, \quad r \geq 1,$$

$$P_0(t) = e^{-\int_0^t \lambda_0(v)dv}. \tag{A.1}$$

$P_r(t)$ is a probability mass function provided that $\sum_{r=0}^{\infty} P_r(t) = 1$; this is shown by following the same steps as in Section 4 of Chapter XVII from [4] (see also [30], section 4.3.2), where it is proved for the case $\lambda_r(t)$ depends on r but not on t. The authors are working on a future publication where this property will be properly generalized to functions that depend on t as well.

Recall our proposed functional form of $\lambda_r(t)$:

$$\lambda_r(t) = \rho \frac{1 + \frac{\gamma}{\rho} r}{1 + \rho t}. \tag{A.2}$$

Let the function $\mu_r(t)$ be:

$$\mu_r(t) = \int_0^t \lambda_r(v)dv = (1 + r \frac{\gamma}{\rho}) \ln(1 + \rho t). \tag{A.3}$$

It can be seen from Eq. (A.3) that $\mu_r'(t) = \lambda_r(t)$ and $\mu_r(0) = 0 \quad \forall r \geq 0$. We now define the auxiliary function

$$\mu_\Delta(t) = \mu_{r+1}(t) - \mu_r(t) = \frac{\gamma}{\rho} \ln(1 + \rho t), \tag{A.4}$$

which does not depend on r. Note that $\mu_\Delta(0) = 0$.

It is a fact that the following equality holds:

$$\int_0^t \lambda_r(x) e^{\mu_\Delta(x)} (e^{\mu_\Delta(x)} - 1)^r dx = \frac{(\frac{\rho}{\gamma} + r)}{r + 1} (e^{\mu_\Delta(t)} - 1)^{r+1}. \tag{A.5}$$

The proof can be done by differentiating the right side of Eq. (A.5) to show it's indeed a primitive of the left side's integrand. Since the functions are also equal on $t = 0$, fundamental theorem of calculus yields that the equalty is valid for every t.

Considering the initial condition $P_0(0) = 1$ (the beginning population is 0), the probability mass functions are given by:

$$P_r(t) = \frac{\Gamma(\frac{\rho}{\gamma} + r)}{r! \, \Gamma(\frac{\rho}{\gamma})} e^{-(\mu_r(t) - \mu_0(0))} (e^{\mu_\Delta(t)} - 1)^r. \tag{A.6}$$

This is demonstrated by induction over r, using Eq. (A.6) as inductive hypothesis and $P_0(t)$ from Eq. (A.1). In fact,

$$P_{r+1}(t) = \int_0^t e^{-(\mu_{r+1}(t) - \mu_{r+1}(x))} \lambda_r(x) P_r(x) dx$$

$$= \frac{\Gamma(\frac{\rho}{\gamma} + r)}{r! \, \Gamma(\frac{\rho}{\gamma})} e^{-(\mu_{r+1}(t) - \mu_0(0))} \int_0^t \lambda_r(x) e^{\mu_\Delta(x)} (e^{\mu_\Delta(x)} - 1)^r(x) dx$$

$$= \frac{\Gamma(\frac{\rho}{\gamma} + r)}{r! \, \Gamma(\frac{\rho}{\gamma})} e^{-(\mu_{r+1}(t) - \mu_0(0))} \frac{(\frac{\rho}{\gamma} + r)}{r + 1} (e^{\mu_\Delta(t)} - 1)^{r+1}.$$

We can rewrite Eq. (A.6) as:

$$P_r(t) = \frac{\Gamma(\frac{\rho}{\gamma} + r)}{r! \, \Gamma(\frac{\rho}{\gamma})} e^{-(\mu_r(t) - \mu_0(0) - r\mu_\Delta(t))} (1 - e^{-\mu_\Delta(t)})^r. \tag{A.7}$$

Using the fact hat $\mu_r(t) = \mu_0(t) + r\mu_\Delta(t)$ and replacing $\mu_0(t)$, $\mu_0(0)$ and $\mu_\Delta(t)$ by their expression yields:

$$P_r(t) = \frac{\Gamma(\frac{\rho}{\gamma} + r)}{r! \, \Gamma(\frac{\rho}{\gamma})} \left(\frac{1}{1 + \rho t} \right) \left(1 - \left(\frac{1}{1 + \rho t} \right)^{\frac{\gamma}{\rho}} \right)^r. \tag{A.8}$$

It should be noted that, since the expression Eq. (A.1) is also valid for the case $\lambda_r(t) = \rho \frac{\frac{\gamma}{\rho} + r}{1 + \rho\, t}$, we can define $\mu_r(t)$ and $\mu_\Delta(t)$ so that the same properties are achieved. Therefore, following the same steps we get the expression of the pmfs for the Polya-Lundberg process. A similar procedure provides the pmfs for the Yule process.

Appendix B. Mean value function for Pure Birth processes

Multipliying Eq. (1) by r and summing we obtain

$$
\sum_{r=1}^{\infty} r\, P'_r(t) = \lim_{R\to\infty} \sum_{r=1}^{R} r\, P'_r(t)
$$

$$
= \lim_{R\to\infty} -\sum_{r=1}^{R} r\, \lambda_r(t)\, P_r(t) + \sum_{r=0}^{R-1} (r+1)\, \lambda_r(t)\, P_r(t)
$$

$$
= \lim_{R\to\infty} -R\, \lambda_R(t)\, P_R(t) + \sum_{r=0}^{R-1} \lambda_r(t)\, P_r(t).
$$

Since the condition

$$
\lim_{r\to\infty} r\, \lambda_r(t)\, P_r(t) = 0. \tag{B.1}
$$

is attained for our model, we get

$$
\sum_{r=1}^{\infty} r\, P'_r(t) = \sum_{r=0}^{\infty} \lambda_r(t)\, P_r(t). \tag{B.2}
$$

By using the expression of $P_r(t)$ for our model (Eq. (A.8)), it is straightforward to see that $M(t)$ is finite using D'Alembert's criterion to show that the series $\sum_{r=1}^{\infty} rP_r(t)$ converges. Therefore, the left side of Eq. (B.2) is $M'(t)$. Then, we can recover the mean value function by solving the following differential equation:

$$
M'(t) = \sum_{r=0}^{\infty} \lambda_r(t)\, P_r(t). \tag{B.3}
$$

Replacing $\lambda_r(t)$ in Eq. (B.3) by Eq. (A.2)

$$
M'(t) = \sum_{r=0}^{\infty} \rho \frac{1 + \frac{\gamma}{\rho}\, r}{1 + \rho\, t} P_r(t)
$$

$$
= \frac{\rho}{1 + \rho\, t} \sum_{r=0}^{\infty} P_r(t) + \frac{\gamma}{1 + \rho\, t} \sum_{r=0}^{\infty} r\, P_r(t)
$$

$$
= \frac{\rho}{1 + \rho\, t} + \frac{\gamma}{1 + \rho\, t}\, M(t)
$$

$$
= \frac{\rho}{1 + \rho\, t} (1 + \frac{\gamma}{\rho}\, M(t)).
$$

we obtain Eq. (B.4), which can be solved to retrieve our model's mean value function, Eq. (11).

$$
M'(t) = \frac{\rho}{1 + \rho\, t} (1 + \frac{\gamma}{\rho}\, M(t)). \tag{B.4}
$$

The same procedure that leads to Eq. (B.3) is also valid for the Polya-Lundberg process. Replacing $\lambda_r(t)$ by its expression and solving the differential equation results in a linear mean value function. In the same way, with $\lambda_r(t) = \lambda\, r$, we get an exponential mean value function, which corresponds to the Yule process.

Appendix C. Mean time between infections calculation for the contagion model

We begin from the exponential waiting time (Eq. (3)), which we can rewrite as

$$
P(T_r \geq t | t_{r-1}) = \exp\left(-\int_{t_{r-1}}^{t} \lambda_{r-1}(u)\, du\right). \tag{C.1}
$$

This shows that the exponential waiting time is, in fact, the probability complement of $F(t)$, which is the distribution function of the (finite) time until the r-th infections. We shall write this as

$$
F(t, t < \infty | t_{r-1}) = 1 - \exp\left(-\int_{t_{r-1}}^{t} \lambda_{r-1}(u)\, du\right). \tag{C.2}
$$

Moreover, the probability of having an *infinite* time until the r-th infections can also be deduced from Eq. (C.1). We can then write its complement as

$$
P(t_r < \infty | t_{r-1}) = 1 - P(t_r = \infty | t_{r-1}) = 1 - \exp\left(-\int_{t_{r-1}}^{\infty} \lambda_{r-1}(u)\, du\right). \tag{C.3}
$$

Dividing Eq. (C.2) by Eq. (C.3) we recover the distribution function of the time variable t conditioned on it being finite, which we can differentiate to obtain the pdf

$$
f(t | t_{r-1}, t_r < \infty) = \frac{\lambda_{r-1}(t) \cdot \exp\left(-\int_{t_{r-1}}^{t_r} \lambda_{r-1}(u)\, du\right)}{1 - \exp\left(-\int_{t_{r-1}}^{\infty} \lambda_{r-1}(u)\, du\right)}. \tag{C.4}
$$

Note that all these probabilities are conditioned on the previous detected infection cases. Taking the conditional expectation of t results in Eq. (6), and replacing Eq. (8) therein yields Eq. (12), where r and t denote $r-1$ and t_{r-1} respectively. We shall write this more precisely as

$$
MTBI_r(r-1, t_{r-1}) = \frac{1 + \rho t_{r-1}}{\gamma\, (r-1)}. \tag{C.5}
$$

Under the assumption that exactly the *expected* number of infections have been detected, we can replace $r-1$ in Eq. (C.5) to get Eq. (13), which we rewrite as

$$
MTBI_r(t_r | r-1 = \mu(t_{r-1})) = \frac{1}{\rho} \frac{(1 + \rho\, t_{r-1})}{(1 + \rho\, t_{r-1})^{\frac{\gamma}{\rho}} - 1}. \tag{C.6}
$$

This is clearly a conditional *MTBI*. Eq. (C.6) is useful to predict future values of the MTBI, when the real $r-1$ is unknown.

References

[1] Barraza NR. Software reliability modeled on contagion. In: 2016 IEEE international symposium on software reliability engineering workshops (ISSREW); 2016. p. 49–50. doi:10.1109/ISSREW.2016.22.

[2] Alajaji F, Fuja T. A communication channel modeled on contagion. IEEE Trans Inf Theory 1994;40(6):2035–41. doi:10.1109/18.340476.

[3] Banerjee A, Burlina P, Alajaji F. Image segmentation and labeling using the polya urn model. IEEE Trans Image Process 1999;8(9):1243–53. doi:10.1109/83.784436.

[4] Feller W. An Introduction to Probability Theory and Its Applications, Vol. 1. 3rd. Wiley; 1968. ISBN 0471257087.

[5] Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. Physics of Life Reviews 2016;18:66–97. doi:10.1016/j.plrev.2016.07.005. http://www.sciencedirect.com/science/article/pii/S1571064516300641.

[6] Li MY. An Introduction to Mathematical Modeling of Infectious Diseases. 1st. Springer International Publishing; 2018. doi:10.1007/978-3-319-72122-4. ISBN 978-3-319-72121-7.

[7] Anderson R, Anderson B, May R. Infectious Diseases of Humans: Dynamics and Control. Dynamics and Control. OUP Oxford; 1992. ISBN 9780198540403, https://books.google.com.ar/books?id=HT0--xXBguQC.

[8] Chen D. Modeling the spread of infectious diseases: a review. John Wiley & Sons, Ltd; 2015. p. 19–42. ISBN 9781118630013. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118630013.ch2.

[9] Getz WM, Salter R, Muellerklein O, Yoon HS, Tallam K. Modeling epidemics: A primer and numerus model builder implementation. Epidemics 2018;25:9–19. doi:10.1016/j.epidem.2018.06.001. http://www.sciencedirect.com/science/article/pii/S1755436517301342.

[10] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A, Contain Paper Math Phys Charact 1927;115(772):700–21.

[11] Ali I, Ullah Khan S. Analysis of stochastic delayed sirs model with exponential birth and saturated incidence rate. Chaos Solitons Fractal 2020;138:110008. doi:10.1016/j.chaos.2020.110008. http://www.sciencedirect.com/science/article/pii/S0960077920304069.

[12] Contreras S, Biron-Lattes JP, Villavicencio HA, Medina-Ortiz D, Llanovarced-Kawles N, Álvaro Olivera-Nappa. Statistically-based methodology for revealing real contagion trends and correcting delay-induced errors in the assessment of covid-19 pandemic. Chaos Soliton Fractal 2020;139:110087. doi:10.1016/j.chaos.2020.110087. http://www.sciencedirect.com/science/article/pii/S0960077920304847.

[13] Ganyani T, Faes C, Hens N. Inference of the generalized-growth model via maximum likelihood estimation: A reflection on the impact of overdispersion. Journal of Theoretical Biology 2020;484. doi:10.1016/j.jtbi.2019.110029. http://www.sciencedirect.com/science/article/pii/S0022519319303996.

[14] Silva PC, Batista PV, Lima HS, Alves MA, Guimarães FG, Silva RC. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos Soliton Fractal 2020;139:110088. doi:10.1016/j.chaos.2020.110088. http://www.sciencedirect.com/science/article/pii/S0960077920304859.

[15] Van Dyke Parunak H, Savit R, Riolo RL. Agent-based modeling vs equation-based modeling: a case study and users' guide. In: Sichman JS, Conte R, Gilbert N, editors. Multi-agent systems and agent-based simulation. Berlin, Heidelberg: Springer Berlin Heidelberg; 1998. p. 10–25. ISBN 978-3-540-49246-7

[16] Ribeiro MHDM, da Silva RG, Mariani VC, dos Santos Coelho L. Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil. Chaos Soliton Fractal 2020;135:109853. doi:10.1016/j.chaos.2020.109853. http://www.sciencedirect.com/science/article/pii/S0960077920302538.

[17] da Silva RG, Ribeiro MHDM, Mariani VC, dos Santos Coelho L. Forecasting brazilian and american covid-19 cases based on artificial intelligence coupled with climatic exogenous variables. Chaos Soliton Fractal 2020;139:110027. doi:10.1016/j.chaos.2020.110027. http://www.sciencedirect.com/science/article/pii/S0960077920304252.

[18] Simon M. Sir epidemics with stochastic infectious periods. Stochastic Processes and their Applications 2020;130(7):4252–74. doi:10.1016/j.spa.2019.12.003. http://www.sciencedirect.com/science/article/pii/S030441491830512X.

[19] Cao Z, Shi Y, Wen X, Su H, Li X. Dynamic behaviors of a two-group stochastic sirs epidemic model with standard incidence rates. Physica A: Statistical Mechanics and its Applications 2020;554:124628. doi:10.1016/j.physa.2020.124628. http://www.sciencedirect.com/science/article/pii/S037843712030306X.

[20] Liu Y, Zhang Y, Wang Q. A stochastic sir epidemic model with lévy jump and media coverage. Adv Diff Eqs 2020;2020. doi:10.1186/s13662-020-2521-6.

[21] Clancy D, O'Neill P. Approximation of epidemics by inhomogeneous birth-and-death processes. Stochastic Processes and their Applications 1998;73(2):233–45. doi:10.1016/S0304-4149(97)00092-6. http://www.sciencedirect.com/science/article/pii/S0304414997000926.

[22] Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. Epidemics 2016;15:27–37. doi:10.1016/j.epidem.2016.01.002. http://www.sciencedirect.com/science/article/pii/S1755436516000037.

[23] Pell B, Kuang Y, Viboud C, Chowell G. Using phenomenological models for forecasting the 2015 ebola challenge. Epidemics 2018;22:62–70. doi:10.1016/j.epidem.2016.11.002. The RAPIDD Ebola Forecasting Challenge; http://www.sciencedirect.com/science/article/pii/S1755436516300433.

[24] Lawler G. Introduction to Stochastic Processes, Second Edition. Taylor & Francis; 2006. ISBN 9781584886518, https://books.google.com.ar/books?id=LATJdzBlajUC.

[25] Pena G, Barraza NR. Increasing failure rate software reliability models for agile projects: A comparative study. System performance modeling. Anand A, editor. De Gryuter; 2020. To be published

[26] Mishra S, Fisman DN, Boily M-C. The abc of terms used in mathematical models of infectious diseases. J Epidemiol Commun Health 2011;65(1):87–94. doi:10.1136/jech.2009.097113. https://jech.bmj.com/content/65/1/87.

[27] Ross S. Stochastic Processes. 2nd. Wiley India Pvt. Limited; 2008. ISBN 9788126517572, https://books.google.com.ar/books?id=HVHqPgAACAAJ.

[28] Pena G. covid19_contagion_model. 2020. https://github.com/GabrielPenaU3F/covid19_contagion_model

[29] Sendova K, Minkova L. Introducing the non-homogeneous compound-birth process. Stochastics 2019:1–19. doi:10.1080/17442508.2019.1666132.

[30] Stroock DW. An Introduction to Markov Processes. Graduate Texts in Mathematics, 230. 1st. Springer-Verlag Berlin Heidelberg; 2005. ISBN 978-3-540-23451-7