A ROBUST PHENOMENOLOGICAL APPROACH TO INVESTIGATE COVID-19 DATA FOR FRANCE

QUENTIN GRIETTE, JACQUES DEMONGEOT, AND PIERRE MAGAL

Dedicated to the memory of Professor Mayan Mimura.

ABSTRACT. We provide a new method to analyze the COVID-19 cumulative reported case data based on a two-step process: first, we regularize the data by using a phenomenological model which takes into account the endemic or epidemic nature of the time period, then we use a mathematical model which reproduces the epidemic exactly. This allows us to derive new information on the epidemic parameters and to compute the effective basic reproductive ratio on a daily basis. Our method has the advantage of identifying robust trends in the number of new infectious cases and produces an extremely smooth reconstruction of the epidemic. The number of parameters required by the method is parsimonious: for the French epidemic between February 2020 and January 2021 we use only 11 parameters in total.

1. INTRODUCTION

Modeling endemic and epidemic phases of the infectious diseases such as smallpox which by the 16th century had become a predominant cause of mortality in Europe until the vaccination by E. Jenner in 1796, and present Covid-19 pandemic outbreak has always been a means of describing and predicting disease. D. Bernoulli proposed in 1760 a differential model [2] taking into account the virulence of the infectious agent and the mortality of the host, which showed a logistic formula [2, p.13] of the same type as the logistic equation by Verhulst [9]. The succession of an epidemic phase followed by an endemic phase had been introduced by Bernoulli and for example appears clearly in the Figures 9 and 10 in [5].

The aim of this article is to propose a new approach to compare epidemic models with data from reported cumulative cases. Here we propose a phenomenological model to fit the observed data of cumulative infectious cases of COVID-19 that describe the successive epidemic phases and endemic intermediate phases. This type of problem dates back to the 1970s with the work of London and York [8]. More recently, Chowell et al. [3] have proposed a specific function to model the temporal transmission speeds $\tau(t)$. In the context of COVID-19, a two-phase model has been proposed by Liu et al. [7] to describe the South Korean data with an epidemic phase followed by an endemic phase.

In this article, we use a phenomenological model to fit the data (see Figure 1). The phenomenological model is used in the modeling process between the data and the epidemic models. The difficulty here is to propose a simple phenomenological model (with a limited number of parameters) that would give a meaningful result for the time-dependent transmission rates $\tau(t)$. Many models could potentially

be used as phenomenological models to represent the data (ex. cubic spline and others). The major difficulty here is to provide a model that gives a good description of the tendency for the data. It has been observed in our previous work that it is difficult to choose between the possible phenomenological models (see Figures 12-14 in [4]). The phenomenological model can also be viewed as a regularization of data that should not fluctuate too much to keep the essential information. An advantage in our phenomenological model is the limited number of parameters (5 parameters during each epidemic phase and 2 parameters during each endemic phase). The last advantage of our approach is that once the phenomenological model has been chosen, we can compute some explicit formula for the transmission rate and derive some estimations for the other parameters.
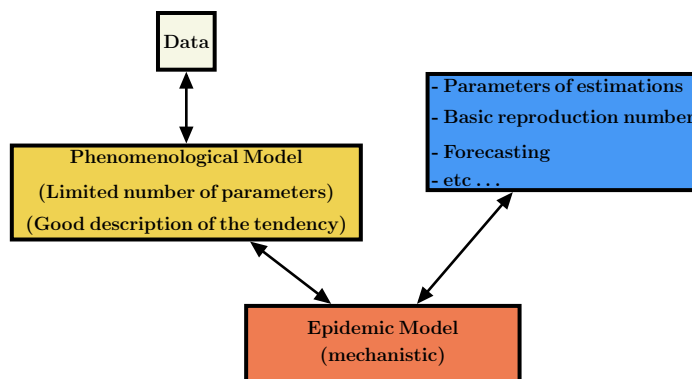

FIGURE 1. *We can apply statistical methods to estimate the parameters of the proposed phenomenological model and derive their average values with some confidence intervals. The phenomenological model is used at the first step of the modelling process, providing regularized data to the epidemic model and allowing the identification of its parameters.*

2. MATERIAL AND METHODS

2.1. **Phenomenological model.** In this article, the phenomenological model is compared with the cumulative reported case data taken from WHO [1]. The phenomenological model deals with data series of new infectious cases decomposed into two types of successive phases, 1) endemic phases, followed by 2) epidemic ones.

Endemic phase. During the endemic phase, the dynamics of new cases appear to fluctuate around an average value independently of the number of cases. Therefore the average cumulative number of cases is given by

$$\mathrm{CR}(t) = N_0 + (t - t_0) \times a, \text{ for } t \in [t_0, t_1], \tag{2.1}$$

where t_0 denotes the beginning of the endemic phase. a is the average value of $\mathrm{CR}(t_0)$ and N_0 the average value of the daily number of new cases.

In other words, we assume that the average daily number of new cases is constant. Therefore the daily number of new cases is given by

$$\mathrm{CR}'(t) = a. \tag{2.2}$$

Epidemic phase. In the epidemic phase, the new cases contribute to produce secondary cases. Therefore the daily number of new cases is no longer constant but varies with time as follows

$$\mathrm{CR}(t) = N_{\mathrm{base}} + N(t), \text{ for } t \in [t_0, t_1], \tag{2.3}$$

where

$$N(t) = \frac{e^{\chi(t-t_0)}N_0}{\left[1 + \frac{N_0^\theta}{N_\infty^\theta}\left(e^{\chi\theta(t-t_0)} - 1\right)\right]^{1/\theta}}. \tag{2.4}$$

In other words, the daily number of new cases follows the Bernoulli-Verhulst [2, 9] equation. Namely, by setting $N(t) = \mathrm{CR}(t) - N_{\mathrm{base}}$ we obtain

$$N'(t) = \chi N(t)\left[1 - \left(\frac{N(t)}{N_\infty}\right)^\theta\right] \tag{2.5}$$

with the initial value

$$N(t_0) = N_0. \tag{2.6}$$

In the model $N_{\mathrm{base}} + N_0$ corresponds to the value $\mathrm{CR}(t_0)$ of the cumulative number of cases at time $t = t_0$. The parameter $N_\infty + N_{\mathrm{base}}$ is the maximal increase of the cumulative reported case after the time $t = t_0$. $\chi > 0$ is a Malthusian growth parameter, and θ regulates the speed at which the $\mathrm{CR}(t)$ increases to $N_\infty + N_{\mathrm{base}}$.

Regularized model. Because the formula for $\tau(t)$ involves derivatives of the phenomenological model regularizing $\mathrm{CR}(t)$ (see equation (2.13)), we need to connect the phenomenological models of the different phases as smoothly as possible. We let $\widetilde{\mathrm{CR}}(t)$ be the model obtained by placing phenomenological models side by side for different phases. Outside of the time window where phenomenological models are used, we consider that the function $\widetilde{\mathrm{CR}}(t)$ is constant. We define the regularized model by using the convolution formula:

$$\mathrm{CR}(t) = \int_{-\infty}^{+\infty} \widetilde{\mathrm{CR}}(t-s) \times \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{s^2}{2\sigma^2}}\,\mathrm{d}s = (\widetilde{\mathrm{CR}} * \mathcal{G})(t), \tag{2.7}$$

where $\mathcal{G}(t) := \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{t^2}{2\sigma^2}}$ is the Gaussian function with variance σ^2. The parameter σ controls the trade-off between smoothness and precision: increasing σ reduces the variations in $\mathrm{CR}(t)$ and decreasing σ reduces the distance between $\mathrm{CR}(t)$ and $\widetilde{\mathrm{CR}}(t)$. In any case the resulting function $\mathrm{CR}(t)$ is very smooth (as well as its derivatives) and close to the original model $\widetilde{\mathrm{CR}}(t)$ when σ is not too large. In numerical applications, we take $\sigma = 2$ days.

Procedure to fit the phenomenological model to the data. To fit the model to the data, we used the regularized model (2.7) where the periods of the different phases are fixed as in Table 1. We use a standard curve-fitting algorithm to find the parameters of the regularized model. In numerical applications we used the Levenberg–Marquardt nonlinear least-squares algorithm provided by the MATLAB© function `fit`. Our 95% confidence intervals are the ones provided as an output of this algorithm. The best-fit parameters and the corresponding confidence intervals are provided in Table 1.

2.2. **SI Epidemic model.** The SI epidemic model used in this work is the same as in [4]. It is summarized by the flux diagram in Figure 2.
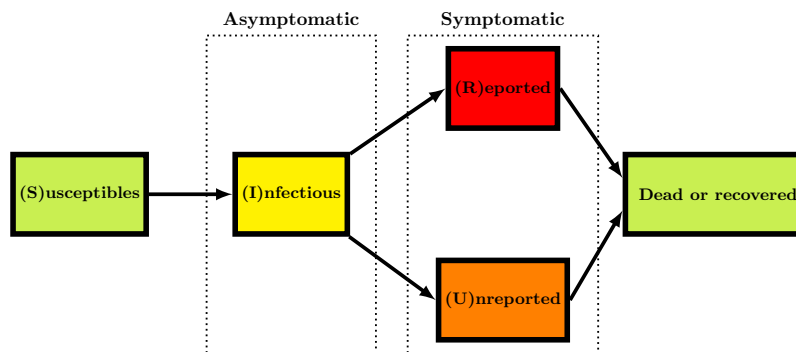
FIGURE 2. *Schematic view showing the different compartments and transition arrows in the epidemic model.*

The goal of this article is to understand how to compare the SI model to the reported epidemic data and therefore the model can be used to predict the future evolution of epidemic spread and to test various possible scenarios of social mitigation measures. For $t \geq t_0$, the SI model is the following

$$\begin{cases} S'(t) = -\tau(t)S(t)I(t), \\ I'(t) = \tau(t)S(t)I(t) - \nu I(t), \end{cases} \qquad (2.8)$$

where $S(t)$ is the number of susceptible and $I(t)$ the number of infectious at time t. This system is supplemented by initial data

$$S(t_0) = S_0 \geq 0, \; I(t_0) = I_0 \geq 0. \qquad (2.9)$$

In this model, the rate of transmission $\tau(t)$ combines the number of contacts per unit of time and the probability of transmission. The transmission of the pathogen from the infectious to the susceptible individuals is described by a mass action law $\tau(t)\,S(t)\,I(t)$ (which is also the flux of new infectious).

The quantity $1/\nu$ is the average duration of the infectious period and $\nu I(t)$ is the flux of recovering or dying individuals. At the end of the infectious period, we assume that a fraction $f \in (0,1]$ of the infectious individuals is reported. Let $\mathrm{CR}(t)$ be the cumulative number of reported cases. We assume that

$$\mathrm{CR}(t) = \mathrm{CR}_0 + \nu\, f\, \mathrm{CI}(t), \text{ for } t \geq t_0, \qquad (2.10)$$

where

$$\mathrm{CI}(t) = \int_{t_0}^{t} I(\sigma)\mathrm{d}\sigma. \qquad (2.11)$$

Assumption 2.1 (Given parameters)*.* We assume that

- the number of susceptible individuals when we start to use the model $S_0 = 67$ millions;
- the average duration of infectious period $\dfrac{1}{\nu} = 3$ days;
- the fraction of reported individuals $f = 0.9$;

are known parameters.

Parameters estimated in the simulations. As described in [4] the number of infectious at time t_0 is

$$I_0 = \frac{\mathrm{CR}'(t_0)}{\nu\, f} \qquad (2.12)$$

The rate of transmission $\tau(t)$ at time t is given by

$$\tau(t) = \frac{\nu f \left(\dfrac{\mathrm{CR}''(t)}{\mathrm{CR}'(t)} + \nu \right)}{\nu f \left(I_0 + S_0 \right) - \mathrm{CR}'(t) - \nu \left(\mathrm{CR}(t) - \mathrm{CR}_0 \right)}. \tag{2.13}$$

Parameters estimated in the endemic phase. The initial number of infectious is given by

$$I_0 = \frac{a}{\nu f},$$

and the transmission rate is given by the explicit formula

$$\tau(t) = \frac{\nu^2 f}{\nu f \left(I_0 + S_0 \right) - a - \nu(t - t_0) \times a}, \forall t \in [t_0, t_1].$$

Parameters estimated in the epidemic phase: The initial number of infectious is given by

$$I_0 = \frac{\chi N_0 \left[1 - \left(\dfrac{N_0}{N_\infty} \right)^\theta \right]}{\nu f},$$

and the transmission rate is given by the explicit formula

$$\tau(t) = \frac{\nu f \left(\dfrac{N''(t)}{N'(t)} + \nu \right)}{\nu f \left(I_0 + S_0 \right) - N'(t) - \nu \left(N(t) - N_0 \right)},$$

and since (recall (2.5))

$$N'(t) = \chi N(t) \left[1 - \left(\frac{N(t)}{N_\infty} \right)^\theta \right]$$

and

$$N''(t) = \chi N'(t) \left[1 - (1 + \theta) \left(\frac{N(t)}{N_\infty} \right)^\theta \right], \tag{2.14}$$

we obtain an explicit formula

$$\tau(t) = \frac{\nu f \left(\chi \left[1 - (1 + \theta) \left(\dfrac{N(t)}{N_\infty} \right)^\theta \right] + \nu \right)}{\nu f \left(I_0 + S_0 \right) - \chi N(t) \left[1 - \left(\dfrac{N(t)}{N_\infty} \right)^\theta \right] - \nu \left(N(t) - N_0 \right)}, \tag{2.15}$$

where $N(t)$ is given by (2.4):

$$N(t) = \frac{e^{\chi(t - t_0)} N_0}{\left[1 + \dfrac{N_0^\theta}{N_\infty^\theta} \left(e^{\chi \theta (t - t_0)} - 1 \right) \right]^{1/\theta}}.$$

By using the Bernoulli-Verhulst model to represent the data, the daily number of new cases is nothing but the derivative $N'(t)$ (whenever the unit of time is one day). The daily number of new cases reaches its maximum at the turning point $t = t_p$, and by using (2.14), we obtain

$$N''(t_p) = 0 \Leftrightarrow N(t_p) = \left(\frac{1}{1 + \theta} \right)^{1/\theta} N_\infty.$$

Therefore by using (2.5), the maximum of the daily number of cases equals

$$N'(t_p) = \chi\, N(t_p) \left[1 - \left(\frac{N(t_p)}{N_\infty} \right)^\theta \right].$$

By using the above formula, we obtain a new indicator for the amplitude of the epidemic.

Theorem 2.2. *The maximal daily number of cases in the course of the epidemic phase is given by*

$$\chi \times N_\infty \times \theta \times \left(\frac{1}{1+\theta} \right)^{\frac{1}{\theta}+1}. \tag{2.16}$$

2.3. **Parameter bounds.** The epidemic model (2.8) with time-dependent transmission rate is consistent only insofar as the transmission rate remains positive. This gives us a criterion to judge if a set of epidemic parameters has a chance of being consistent with the observed data: since we know the parameters N_0, N_∞, χ and θ from the phenomenological model, the formula (2.15) allows us to compute a criterion on ν and f which decides whether a given parameter values are compatible with the observed data or not. That is to say that, a set of parameter values is compatible if the transmission rate $\tau(t)$ in (2.15) remains positive for all $t \geq t_0$, and it is not compatible if the sign of $\tau(t)$ in (2.15) changes for some $t \geq t_0$. We refer to [4, Proposition 4.3] for more results.

The value of the parameter ν is compatible with the model (2.15) if and only if

$$0 \leq \frac{1}{\nu} \leq \frac{1}{\chi\theta}, \tag{2.17}$$

and the value of the parameter f is compatible with the model (2.15) if and only if

$$f \geq \frac{N_\infty - N_0}{I_0 + S_0}. \tag{2.18}$$

Therefore, we obtain an information on the parameters ν and f, even though they are not directly identifiable (two different values of ν or f can produce exactly the same cumulative reported cases).

2.4. **Computation of the basic reproduction number.** In order to compute the reproduction number in Figure 5 we use the Algorithm 2 in [4] and the day-by-day values of the phenomenological model.

3. Results

3.1. **Phenomenological model compared to the French data.** In Figure 3 we present the best fit of our phenomenological model for the cumulative reported case data of COVID-19 epidemic in France. The yellow regions correspond to the endemic phases and the blue regions correspond to the epidemic phases. Here we consider the two epidemic waves for France, and the chosen period, as well as the parameters values for each period, are listed in Table 1. In Table 1 we also give 95% confidence intervals for the fitted parameters values.
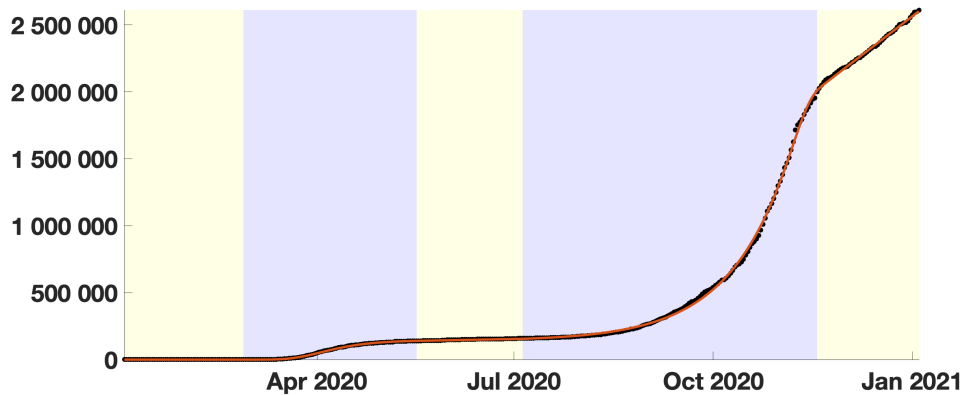
FIGURE 3. *The red curve corresponds to the phenomenological model and the black dots correspond to the cumulative number of reported cases in France.*

Figure 4 shows the corresponding daily number of new reported case data (black dots) and the first derivative of our phenomenological model (red curve).
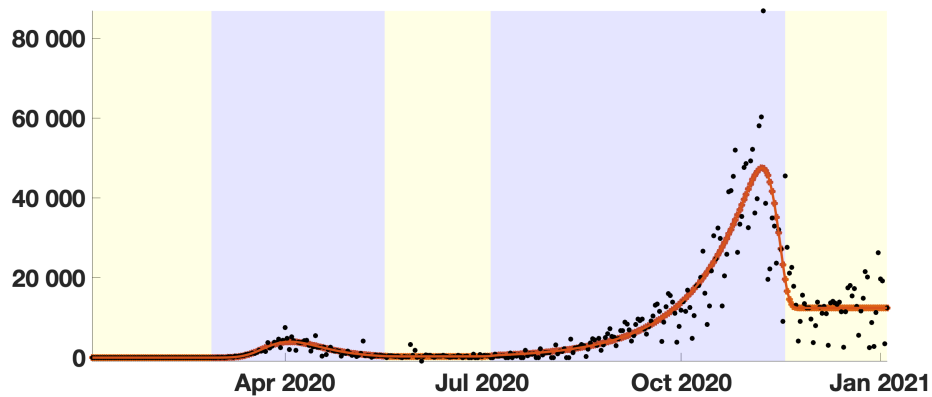

FIGURE 4. *The red curve corresponds to the first derivative of the phenomenological model and the black dots correspond to daily number of new reported cases in France.*

3.2. **SI epidemic model compared to the French data.** Some parameters of the model are known as $S_0 = 67$ millions for France (this is questionable). Some parameters of the epidemic model can not be precisely evaluated [4].

Period	Parameters value	Method	95% Confidence interval
Period 1: Endemic phase Jan 03 - Feb 27	$N_0 = -4.368$	computed	
	$a = 1.099 \times 10^{-1}$	fitted	$a \in [-8.582 \times 10^1, 8.604 \times 10^1]$
Period 2: Epidemic phase Feb 27 - May 17	$N_{base} = 0$	fixed	
	$N_0 = 1.675$	fitted	$N_0 \in [-3.807 \times 10^1, 4.142 \times 10^1]$
	$N_\infty = 1.445 \times 10^5$	fitted	$N_\infty \in [1.367 \times 10^5, 1.523 \times 10^5]$
	$\chi = 1.263$	fitted	$\chi \in [-1.171 \times 10^1, 1.424 \times 10^1]$
	$\theta = 6.315 \times 10^{-2}$	fitted	$\theta \in [-6.086 \times 10^{-1}, 7.349 \times 10^{-1}]$
Period 3: Endemic phase May 17 - Jul 05	$N_0 = 1.405 \times 10^5$	computed	
	$a = 3.11 \times 10^2$	computed	
Period 4: Epidemic phase Jul 05 - Nov 18	$N_{base} = 1.403 \times 10^5$	fitted	$N_{base} \in [1.367 \times 10^5, 1.439 \times 10^5]$
	$N_0 = 1.517 \times 10^4$	fitted	$N_0 \in [1.427 \times 10^4, 1.607 \times 10^4]$
	$N_\infty = 1.953 \times 10^6$	fitted	$N_\infty \in [1.92 \times 10^6, 1.986 \times 10^6]$
	$\chi = 3.671 \times 10^{-2}$	fitted	$\chi \in [3.62 \times 10^{-2}, 3.722 \times 10^{-2}]$
	$\theta = 7.679$	fitted	$\theta \in [6.256, 9.102]$
Period 5: Endemic phase Nov 18 - Jan 04	$N_0 = 4.45 \times 10^{-84}$	computed	
	$a = 1.099 \times 10^{-1}$	fitted	$a \in [1.222 \times 10^4, 1.265 \times 10^4]$

TABLE 1. *Fitted parameters and computed parameters for the whole epidemic going from January 03 2020 to January 04 2021.*

Result

By using (2.17) we obtain the following conditions for the average duration of infectious period

- $0 < \dfrac{1}{\nu} \leq 1/(\chi\theta) = 12.5$ days during the first epidemic wave;

- $0 < \dfrac{1}{\nu} \leq 1/(\chi\theta) = 3.5$ days during the second epidemic wave.

We obtain no constraint for the fraction $f \in (0, 1]$ of reported new cases (between 0 and 1 for France).

Moreover by using the formula (2.16) we deduce that the maximal daily number of cases is

- 4110 during the first epidemic wave;
- 47875 during the second epidemic wave.

Importantly, by combining the phenomenological model from Section 3.1 and the epidemiological model from Section 2.2, we can reconstruct the time-dependent transmission rate given by (2.13) and the corresponding time-dependent basic reproduction number $\mathcal{R}_0(t) = \tau(t)S(t)/\nu$ (sometimes called "effective basic reproductive ratio"). The obtained basic reproduction number is presented in Figure 5. We observe that $\mathcal{R}_0(t)$ is decreasing during each epidemic wave, except at the very end where it becomes increasing. This is not necessarily surprising since the lockdown becomes less strictly respected towards the end. During the endemic phases, the $\mathcal{R}_0(t)$ becomes effectively equal to one, except again near the end. The variations observed close to the transition between two phases may be partially due to the smoothing method, which has an impact on the size of the "bumps". However, they remain very limited in number and size.
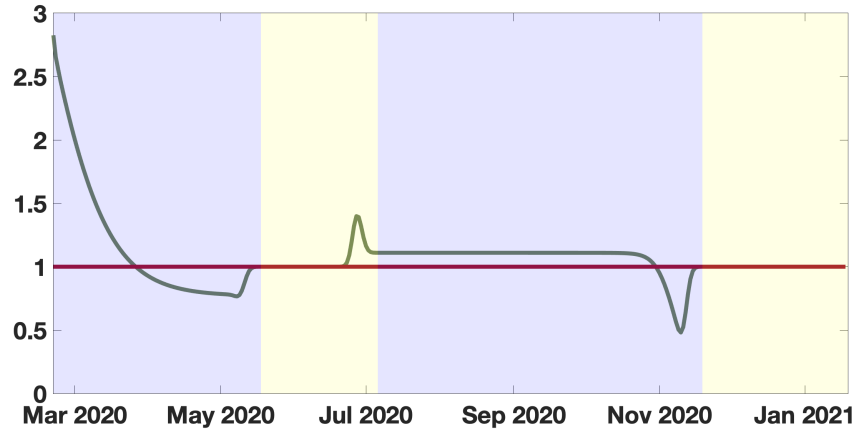
FIGURE 5. *In this figure we plot the time dependent basic reproduction number* $\mathcal{R}_0(t) :=$ $\tau(t)S(t)/\nu$. *We fix the average length of the asymptotic infectious period to 3 days. Notice that, contrary to Figure 3 and 4, we do not plot to first endemic phase because the basic reproduction number is meaningless before the first wave.*

4. Discussion

In our paper, we use a phenomenological model to reduce the number of parameters necessary for summarizing observed data without loss of pertinent information. The process of reduction consists of three stages: qualitative or quantitative detection of the boundaries between the different phases of the dynamics (here endemic and epidemic phases), choice of a reduction model (among different possible approaches: logistic, regression polynomials, splines, autoregressive time series, etc.) and smoothing of the derivatives at the boundary points corresponding to the breaks in the model.

In Figure 3, we have a very good agreement between the data and the phenomenological model, for both the original curve and its derivative. The relative error in Figure 3 is of order 10^{-2}, which means that the error is at most of the order of $100\,000$ individuals. In Figure 4 the red curve also gives a good tendency of the black dots corresponding to raw data.

In Figure 5, the phenomenological models are necessary to derive a significant basic reproduction number. Otherwise the resulting $\mathcal{R}_0(t)$ is not interpretable and even not computable after sometime. Similar results were obtained in Figures 12-14 in [4]. The method to compute $\mathcal{R}_0$ can also be applied directly to the original data. We did not show the result here because the noise in the data is amplified by the method and the results are not usable. This shows that it is important to use the phenomenological model to provide a good regularization of the data.

In Figure 5, the major difficulty is to know how to make the transition from an epidemic phase to an endemic phase and vice versa. This is a non-trivial problem that is solved by our regularization approach (using a convolution with a Gaussian). As we can see in Figure 5, the number of oscillations is very limited between two phases. Without regularization, there is a sharp corner at the transition between two phases which leads to infinite values in $\tau(t)$. The choice of the convolution with a Gaussian kernel for the regularization method is the result of an experimental process. We tried several different regularization methods, including a smooth explicit interpolation function and Hermite polynomials. Eventually, the convolution with a Gaussian kernel gives the best results.

To minimize the variations of the curve of $\mathcal{R}_0(t)$, the choice of the transition dates between two phases is critical. In Figure 5 we choose the transition dates so that the derivatives of the phenomenological model do not oscillate too much. Other choices lead to higher variations or increase the number of oscillations. Finally, the qualitative shape of the curve presented in Figure 5 is very robust to changes in the epidemic parameters, even though the quantitative values of $\mathcal{R}_0(t)$ are different for other values of the parameters ν and f.

In Figure 5, we observe that the quantitative value of the $\mathcal{R}_0(t)$ during the first part of the second epidemic wave (second blue region) is almost constant and equals 1.11. This value is significantly lower than the one observed at the beginning of the first epidemic wave (first blue region). Yet the number of cases produced during the second wave is much higher than the number of cases produced during the first wave.

We observe that the values of the parameters of the phenomenological model are quantitatively different between the first wave and the second wave. Several phenomena can explain this difference. The population was better prepared for the second wave. The huge difference in the number of daily reported cases during the second phase can be partially attributed to the huge increase in the number of tests in France during this period. But this is only a partial explanation for the explosion of cases during the second wave. We also observe that the average duration of the infectious period varies between the first epidemic wave (12.5 days) and the second epidemic wave (3.5 days). This may indicate a possible adaptation of the virus SARS CoV-2 circulating in France during the two periods, or the effect of the mitigation measures, with better respect of the social distancing and compulsory mask-wearing.

The huge difference between the initial values of $\mathcal{R}_0(t)$ in the first and the second waves is an apparent paradox which shows that $\mathcal{R}_0(t)$ has a limited explanatory value regarding the severity of the epidemic: even if the quantitative value of $\mathcal{R}_0(t)$ is higher at the start of the first wave, the number of cases produced during an equivalent period in the second wave is much higher. This paradox can be partially resolved by remarking that the $\mathcal{R}_0(t)$ behaves like an exponential rate and the number of secondary cases produced in the whole population is therefore very sensitive to the number of active cases at time t. In other words, $\mathcal{R}_0(t)$ is blind to the epidemic state of the population and cannot be used as a reliable indicator of the severity of the epidemic. Other indicators have to be found for that purpose; we propose, for instance, the maximal value of the daily number of new cases, which can be forecasted by our method (see equation (2.16)), although other indicators can be imagined.

In Figure 6 we present an exploratory scenario assuming that during the endemic period preceding the second epidemic wave (May 17 - Jul 05) the daily number of cases is divided by 10. The resulting cumulative number of cases obtain is five lower the original one. We summarize this observation into the following statement.

> **Result**
>
> - The level of the daily number of cases during an endemic phase preceding an epidemic phase strongly influences the severity of this epidemic wave.
>
> - In other words, maintaining social distancing between epidemic waves is essential.
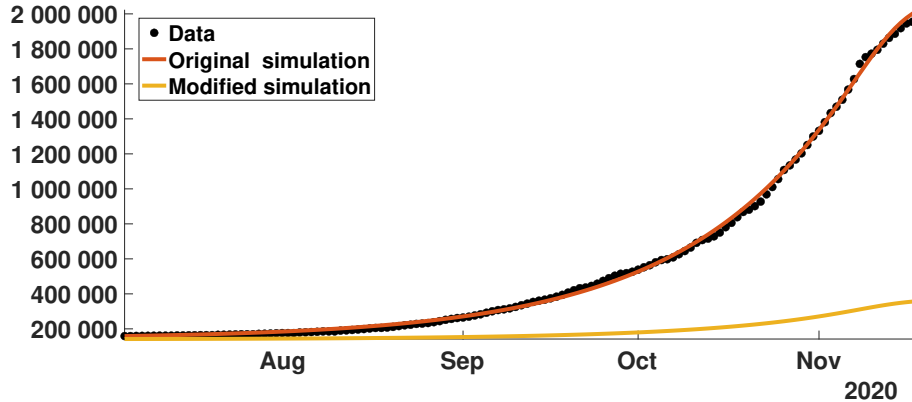
FIGURE 6. *Cumulative number of cases for the second epidemic wave obtained by using the SI model (2.8) with $\tau(t)$ given by (2.13), the parameters from Table 1. We start the simulation at time $t_0 = $ July 05 with the initial value $I_0 = \frac{CR'(t_2)}{\nu f}$ for the red curve and with $I_0 = \frac{1}{10} \frac{CR'(t_2)}{\nu f}$ for yellow curve. The remaining parameters used are $\nu = 1/3$, $f = 0.9$, $S_0 = 66841266$. We observe the number is five times lower than then the original number of cases.*

In Figure 4, there is two order of magnitude in the daily number of cases in between $CR'(t_1) \approx 1$ (with $t_1 = $ Feb 27) at the early beginning of the first epidemic wave and $CR'(t_2) = 422$ (with $t_2 = $ July 05) at the early beginning of the second epidemic wave. That confirms our result. After the second wave, the average daily number of cases $CR'(t)$ in France is stationary and approximately equal to 12440. Therefore, if the above observation remains true and if a third epidemic wave occurs, the third epidemic wave is expected to be more severe than the first and second epidemic waves.

Our study can be extended in several directions. A statistical study of the parameters obtained by using our phenomenological model with data at the regional scale could be interesting. We could in particular investigate statistically the correlations existing between the parameters changes and the variations with demographic parameters as the median age and the population density, as well as geo-climatic factors as the elevation and temperature, etc. We also plan to extend our method to more realistic epidemic models, like the SEIUR model from [6], which includes the possibility of transmission from asymptomatic unreported patients.

Author Contributions: QG, JD and PM conceived and designed the study. QG and PM analyzed the data, carried out the analysis and performed numerical simulations, JD and PM conducted the literature review. All authors participated in writing and reviewing of the manuscript.

References

1. Data from WHO: `https://covid19.who.int/WHO-COVID-19-global-data.csv`, accessed June 9, 2021.

2. D. Bernoulli, *Essai d'une nouvelle analyse de la petite Vérole, & des avantages de l'Inoculation pour la prévenir*, Histoire de l'Académie royale des sciences avec les mémoires de mathématique et de physique tirés de cette Académie (1766), 1–45.

3. G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. M. Hyman, *The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda*, J. Theor. Biol. **229** (2004), no. 1, 119–126.

4. J. Demongeot, Q. Griette, and P. Magal, *SI epidemic model applied to COVID-19 data in mainland China*, Roy. Soc. Open Sci. **7** (2020), no. 12, 201878.

5. K. Dietz and J. A. P. Heesterbeek, *Daniel Bernoulli's epidemiological model revisited*, Math. Biosci. **180** (2002), 1–21.

6. Z. Liu, P. Magal, O. Seydi, and G. Webb, *Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data*, Math. Biosci. Eng. **17** (2020), no. 4, 3040–3051.

7. _____ , *Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions*, Biology **9** (2020), no. 3, 50.

8. W. P. London and J. A. Yorke, *Recurrent outbreaks of measles, chickenpox and mumps: I. Seasonal variation in contact rates*, Am. J. Epidemiol. **98** (1973), no. 6, 453–468.

9. P. F. Verhulst, *Notice sur la loi que la population poursuit dans son accroissement*, Correspondance mathématique et physique **10** (1838), 113–121.

Univ. Bordeaux, IMB, CNRS UMR 5251, F-33400 Talence, France.
E-mail address: `quentin.griette@u-bordeaux.fr`

Univ. Grenoble Alpes, AGEIS EA7407, F-38700 La Tronche, France
E-mail address: `jacques.demongeot@univ-grenoble-alpes.fr`

Corresponding author, Univ. Bordeaux, IMB, CNRS UMR 5251, F-33400 Talence, France
E-mail address: `pierre.magal@u-bordeaux.fr`