SI epidemic model applied to COVID-19 data in mainland China

J. Demongeot[(a)], Q. Griette[(b)] and P. Magal[(b)*]

[(a)] *Université Grenoble Alpes, AGEIS EA7407, F-38700 La Tronche, France*

[(b)] *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.*

CNRS, IMB, UMR 5251, F-33400 Talence, France.

November 13, 2020

Abstract

The article is devoted to the parameters identification in the SI model. We consider several methods, starting with an exponential fit of the early cumulative data of Sars-CoV2 in mainland China. The present methodology provides a way to compute the parameters at the early stage of the epidemic. Next, we establish an identifiability result. Then we use the Bernoulli-Verhulst model as a phenomenological model to fit the data and derive some results on the parameters identification. The last part of the paper is devoted to some numerical algorithms to fit a daily piecewise constant rate of transmission.

Keywords: Corona virus, reported and unreported cases, parameters identification, epidemic mathematical model.

1 Introduction

Estimating the average transmission rate is one of the most crucial challenges in the epidemiology of communicable diseases. This rate conditions the entry into the epidemic phase of the disease and its return to the extinction phase, if it has diminished sufficiently. It is the combination of three factors, one, the coefficient of virulence, linked to the infectious agent (in the case of infectious transmissible diseases), the other, the coefficient of susceptibility, linked to the host (all summarized into the probability of transmission), and also, the number of contact per unit of time between individuals (see Magal and Ruan [1]). The coefficient of virulence may change over time due to mutation over the course of the disease history. The second and third also, if mitigation measures have been taken. This was the case in China from the start of the pandemic (see Qiu, Chen and Shi [2])). Monitoring the decrease in the average transmission rate is an excellent way to monitor the effectiveness of these mitigation measures. Estimating the rate is therefore a central problem in the fight against epidemics.

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

$$\begin{cases} S'(t) = -\tau(t)S(t)I(t), \\ I'(t) = \tau(t)S(t)I(t) - \nu I(t), \end{cases} \tag{1.1}$$

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

$$S(t_0) = S_0 \geq 0,\ I(t_0) = I_0 \geq 0. \tag{1.2}$$

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).

*Corresponding author. e-mail: pierre.magal@u-bordeaux.fr

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

$$\mathrm{CR}(t) = \mathrm{CR}_0 + \nu\, f\, \mathrm{CI}(t), \text{ for } t \geq t_0, \tag{1.3}$$

where

$$\mathrm{CI}(t) = \int_{t_0}^{t} I(\sigma)\mathrm{d}\sigma. \tag{1.4}$$

Assumption 1.1 *We assume that*

- $S_0 > 0$ *the number of susceptible individuals at time* t_0 *when we start to use the model;*
- $\dfrac{1}{\nu} > 0$ *the average duration of infectious period;*
- $f > 0$ *the fraction of reported individuals;*

are known parameters.

Throughout this paper, the parameter $S_0 = 1.4 \times 10^9$ will be the entire population of mainland China (since COVID-19 is a newly emerging disease). The actual number of susceptibles S_0 can be smaller since some individuals can be partially (or totally) immunized by previous infections or other factors. This is also true for Sars-CoV2, even if COVID-19 is a newly emerging disease. In fact, for COVID-19 the level of susceptibility may depend on blood group and genetic lineage. It is indeed suspected that the blood group O is associated with a lower susceptibility to SARS-CoV2 while a gene cluster inherited from Neanderthal has been identified as a risk factor for severe symptoms (see Zeberg et al. [3] and Guillon et al. [4]).

At the early beginning of the epidemic, the average duration of the infectious period $1/\nu$ is unknown, since the virus has never been investigated in the past. Therefore, at the early beginning of the COVID-19 epidemic, medical doctors and public health scientists used previously estimated average duration of the infectious period to make some public health recommendations. Here we show that the average infectious period is impossible to estimate by using only the time series of reported cases, and must therefore be identified by other means. Actually, with the data of Sars-CoV2 in mainland China, we will fit the cumulative number of the reported case almost perfectly for any non-negative value $1/\nu < 3.3$ days. In the literature, several estimations were obtained: 11 days in [6], 9.5 days in [7], 8 days in [8], and 3.5 days in [9]. The recent survey by Byrne et al. [5] focuses on this subject.

Result

In Section 3, our analysis shows that

- It is hopeless to estimate the exact value of the duration of infectiousness by using SI models. Several values of the average duration of the infectious period give the exact same fit to the data.
- We can estimate an upper bound for the duration of infectiousness by using SI models. In the case of Sars-CoV2 in mainland China, this upper bound is 3.3 days.

In [10], it is reported that transmission of COVID-19 infection may occur from an infectious individual who is not yet symptomatic. In [11] it is reported that COVID-19 infected individuals generally develop symptoms, including mild respiratory symptoms and fever, on average $5-6$ days after the infection date (with a confience of 95%, range $1-14$ days). In [12] it is reported that the median time prior to symptom onset is 3 days, the shortest 1 day, and the longest 24 days. It is evident that these time periods play an important role in understanding COVID-19 transmission dynamics. Here the fraction of reported individuals f is unknown as well.

Result

In Section 3, our analysis shows that:

- It is hopeless to estimate the fraction of reported by using the SI models. Several values for the fraction of reported give the exact same fit to the data.
- We can estimate a lower bound for the fraction of unreported. We obtain $3.83 \times 10^{-5} < f \leq 1$. This lower bound is not significant. Therefore we can say anything about the fraction of unreported from this class of models.

As a consequence, the parameters $1/\nu$ and f have to be estimated by another method, for instance by a direct survey methodology that should be employed on an appropriated sample in the population in order to evaluate the two parameters.

The goal of this article is to focus on the estimation of the two remaining parameters. Namely, knowing the above-mentioned parameters, we plan to identify

- I_0 the initial number of infectious at time t_0;
- $\tau(t)$ the rate of transmission at time t.

This problem has already been considered in several articles. In the early 70s, London and Yorke [13, 14] already discussed the time dependent rate of transmission in the context of measles, chickenpox and mumps. More recently, Wang and Ruan [15] the question of reconstructing the rate of transmission was considered for the 2002-2004 SARS outbreak in China. In Chowell et al. [16] a specific form was chosen for the rate of transmission and applied to the Ebola outbreak in Congo. Another approach was also proposed in Smirnova et al. [17].

In Section 2, we will explain how to apply the method introduce in Liu et al. [19] to fit the early cumulative data of Sars-CoV2 in China. This method provides a way to compute I_0 and $\tau_0 = \tau(t_0)$ at the early stage of the epidemic. In Section 3, we establish an identifiability result in the spirit of Hadeler [21].

In Section 4, we use the Bernoulli-Verhulst model as a phenomenological model to describe the data. As it was observed in several articles, the data from mainland China (and other countries as well) can be fitted very well by using this model. As a consequence, we will obtain an explicit formula for $\tau(t)$ and I_0 expressed as a function of the parameters of the Bernoulli-Verhulst model and the remaining parameters of the SI model. This approach gives a very good description of this set of data. The disadvantage of this approach is that it requires an evaluation of the final size CR_∞ from the early beginning (or at least it requires an estimation of this quantity).

Therefore, in order to be predictive, we will explore in the remaining sections of the paper the possibility of constructing a day by day rate of transmission. Here we should refer to Bakhta et al. [20] where another novel forecasting method was proposed.

In Section 5, we will prove that the daily cumulative data can be approached perfectly by at most one sequence of day by day piecewise constant transmission rates. In Section 6, we propose a numerical methods to compute such a (piecewise constant) rate of transmission. Section 7 is devoted to the discussion, and we will present some figures showing the daily basic reproduction number for the COVID-19 outbreak in mainland China.

2 Estimating $\tau(t_0)$ and I_0 at the early stage of the epidemic

In this section, we apply the method presented in [18] to the SI model. At the early stage of the epidemic, we can assume that $S(t)$ is almost constant and equal to S_0. We can also assume that $\tau(t)$ remains constant equal to $\tau_0 = \tau(t_0)$. Therefore, by replacing these parameters into the I-equation of system (1.1) we obtain

$$I'(t) = (\tau_0 S_0 - \nu) I(t).$$

Therefore

$$I(t) = I_0 \exp\left(\chi_2 \left(t - t_0\right)\right),$$

where

$$\chi_2 = \tau_0 S_0 - \nu. \tag{2.1}$$

By using (1.3), we obtain

$$\mathrm{CR}(t) = \mathrm{CR}_0 + \nu\, f\, I_0\, \frac{e^{\chi_2 (t-t_0)} - 1}{\chi_2}. \tag{2.2}$$

We obtain a first phenomenological model for the cumulative number of reported cases (valid only at the early stage of the epidemic)

$$\mathrm{CR}(t) = \chi_1 e^{\chi_2\, t} - \chi_3. \tag{2.3}$$

In Figure 1, we compare the model to the COVID-19 data for mainland China. The data used in the article are taken from [29, 30, 31] and reported in Section A. In order to estimate the parameter χ_3, we minimize the distance between $\mathrm{CR}_{\mathrm{Data}}(t) + \chi_3$ and the best exponential fit $t \to \chi_1 e^{\chi_2\, t}$ (i.e. we use the MATLAB function fit(t, data, 'exp1')).

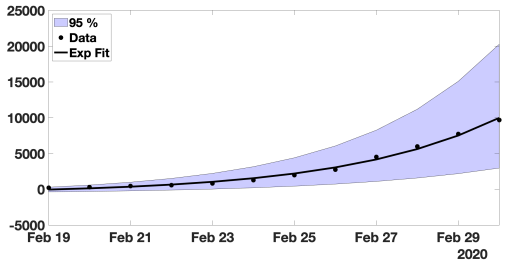

Figure 1: *In this figure, we plot the best fit of the exponential model to the cumulative number of reported cases of COVID-19 in mainland China between February 19 and March 1. We obtain* $\chi_1 = 3.7366$, $\chi_2 = 0.2650$ *and* $\chi_3 = 615.41$ *with* $t_0 = 19$ *Feb. The parameter* χ_3 *is obtained by minimizing the error between the best exponential fit and the data.*

The estimated initial number of infected and transmission rate

By using (1.3) and (2.3) we obtain

$$I_0 = \frac{\mathrm{CR}'(t_0)}{\nu\, f} = \frac{\chi_1\, \chi_2 e^{\chi_2\, t_0}}{\nu\, f}, \tag{2.4}$$

and by using (2.1)

$$\tau_0 = \frac{\chi_2 + \nu}{S_0}. \tag{2.5}$$

Remark 2.1 *Fixing* $f = 0.5$ *and* $\nu = 0.2$, *we obtain*

$$I_0 = 3.7366 \times 0.2650 \times \exp(0.2650 \times 19)/(0.2 \times 0.5) = 1521,$$

and

$$\tau_0 = \frac{0.2650 + 0.2}{1.4 \times 10^9} = 3.3214 \times 10^{-10}.$$

The influence of the errors made in the estimations (at the early stage of the epidemic) has been considered in the recent article by Roda et al. [22]. To understand this problem, let us first consider the case of the rate of transmission $\tau(t) = \tau_0$ in the model (1.1). In that case (1.1) becomes

$$\begin{cases} S'(t) = -\tau_0 S(t) I(t), \\ I'(t) = \tau_0 S(t) I(t) - \nu I(t). \end{cases} \tag{2.6}$$

By using the S-equation of model (2.6) we obtain

$$S(t) = S_0 \exp\left(-\tau_0 \int_{t_0}^{t} I(\sigma) \mathrm{d}\sigma\right) = S_0 \exp\left(-\tau_0 \mathrm{CI}(t)\right)$$

where $\mathrm{CI}(t)$ is the cumulated number of infectious individuals. Substituting $S(t)$ by this formula in the I-equation of (2.6) we obtain

$$I'(t) = S_0 \exp\left(-\tau_0 \mathrm{CI}(t)\right) \tau_0 \mathrm{CI}'(t) - \nu I(t).$$

Therefore, by integrating the above equation between t and t_0 we obtain

$$\mathrm{CI}'(t) = I_0 + S_0\left[1 - \exp\left(-\tau_0 \mathrm{CI}(t)\right)\right] - \nu \mathrm{CI}(t). \tag{2.7}$$

Remarkably, the equation (2.7) is monotone. We refer to Hal Smith [23] for a comprehensive presentation on monotone systems. By applying a comparison principle to (2.7), we are in a position to confirm the intuition about epidemics SI models. Notice that the monotone properties are only true for the cumulative number of infectious (this is false for the number of infectious).

Theorem 2.2 *Let $t > t_0$ be fixed. The cumulative number of infectious $\mathrm{CI}(t)$ is strictly increasing with respect to the following quantities*

(i) $I_0 > 0$ *the initial number of infectious individuals;*

(ii) $S_0 > 0$ *the initial number of susceptible individuals;*

(iii) $\tau > 0$ *the transmission rate;*

(iv) $1/\nu > 0$ *the average duration of the infectiousness period.*

Error in the estimated initial number of infected and transmission rate

Assume that the parameters χ_1 and χ_2 are estimated with a 95% confidence interval

$$\chi^-_{1,95\%} \leq \chi_1 \leq \chi^+_{1,95\%},$$

and

$$\chi^-_{2,95\%} \leq \chi_2 \leq \chi^+_{2,95\%}.$$

We obtain

$$I^-_{0,95\%} := \frac{\chi^-_{1,95\%}\,\chi^-_{2,95\%} e^{\chi^-_{2,95\%}\,t_0}}{\nu\, f} \leq I_0 \leq I^+_{0,95\%} := \frac{\chi^+_{1,95\%}\,\chi^+_{2,95\%} e^{\chi^+_{2,95\%}\,t_0}}{\nu\, f}, \tag{2.8}$$

and

$$\tau^-_{0,95\%} := \frac{\chi^-_{2,95\%} + \nu}{S_0} \leq \tau_0 \leq \tau^+_{0,95\%} := \frac{\chi^+_{2,95\%} + \nu}{S_0}. \tag{2.9}$$

Remark 2.3 *By using the data for mainland China we obtain*

$$\chi^-_{1,95\%} = 1.57,\ \chi^+_{1,95\%} = 5.89,\ \chi^-_{2,95\%} = 0.24,\ \chi^+_{2,95\%} = 0.28. \tag{2.10}$$

In Figure 2, we plot the upper and lower solutions $\mathrm{CR}^+(t)$ (obtained by using $I_0 = I^+_{0,95\%}$ and $\tau_0 = \tau^+_{0,95\%}$) and $\mathrm{CR}^-(t)$ (obtained by using $I_0 = I^-_{0,95\%}$ and $\tau_0 = \tau^-_{0,95\%}$) corresponding to the blue region and the black curve corresponds to the best estimated value $I_0 = 1521$ and $\tau_0 = 3.3214 \times 10^{-10}$.

Recall that the final size of the epidemic corresponds to the positive equilibrium of (2.7)

$$0 = I_0 + S_0\left[1 - \exp\left(-\tau_0 \mathrm{CI}_\infty\right)\right] - \nu \mathrm{CI}_\infty. \tag{2.11}$$

In Figure 2 the changes in the parameters I_0 and τ_0 (in (2.8)-(2.9)) do not affect significantly the final size.

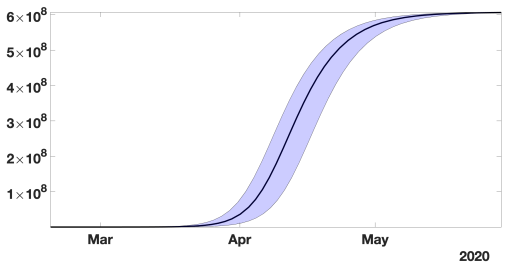

Figure 2: *In this figure, the black curve corresponds to the cumulative number of reported cases* $\mathrm{CR}(t)$ *obtained from the model (2.6) with* $\mathrm{CR}'(t) = \nu f I(t)$ *by using the values* $I_0 = 1521$ *and* $\tau_0 = 3.32 \times 10^{-10}$ *obtained from our method and the early data from February 19 to March 1. The blue region corresponds the* 95% *confidence interval when the rate of transmission* $\tau(t)$ *is constant and equal to the estimated value* $\tau_0 = 3.32 \times 10^{-10}$.

3 Theoretical formula for $\tau(t)$

By using the S-equation of model (1.1) we obtain

$$S(t) = S_0 \exp\left(-\int_{t_0}^{t} \tau(\sigma)\, I(\sigma) \mathrm{d}\sigma\right),$$

next by using the I-equation of model (1.1) we obtain

$$I'(t) = S_0 \exp\left(-\int_{t_0}^{t} \tau(\sigma)\, I(\sigma) \mathrm{d}\sigma\right) \tau(t)\, I(t) - \nu I(t),$$

and by taking the integral between t and t_0 we obtain a Volterra integral equation for the cumulative number of infectious

$$\mathrm{CI}'(t) = I_0 + S_0 \left[1 - \exp\left(-\int_{t_0}^{t} \tau(\sigma)\, I(\sigma) \mathrm{d}\sigma\right)\right] - \nu \mathrm{CI}(t), \tag{3.1}$$

which is equivalent to (by using (1.3))

$$\mathrm{CR}'(t) = \nu\, f \left(I_0 + S_0 \left[1 - \exp\left(-\frac{1}{\nu\, f}\int_{t_0}^{t} \tau(\sigma)\, \mathrm{CR}'(\sigma) \mathrm{d}\sigma\right)\right]\right) + \nu\, \mathrm{CR}_0 - \nu \mathrm{CR}(t). \tag{3.2}$$

The following result permits to obtain a perfect match between the SI model and the time-dependent rate of transmission $\tau(t)$.

Theorem 3.1 *Let* S_0, ν, f, $I_0 > 0$ *and* $\mathrm{CR}_0 \geq 0$ *be given. Let* $t \to I(t)$ *be the second component of system (1.1). Let* $\widehat{\mathrm{CR}} : [t_0, \infty) \to \mathbb{R}$ *be a two times continuously differentiable function satisfying*

$$\widehat{\mathrm{CR}}(t_0) = \mathrm{CR}_0, \tag{3.3}$$

$$\widehat{\mathrm{CR}}'(t_0) = \nu\, f\, I_0, \tag{3.4}$$

$$\widehat{\mathrm{CR}}'(t) > 0, \forall t \geq t_0, \tag{3.5}$$

and

$$\nu f\, (I_0 + S_0) - \widehat{\mathrm{CR}}'(t) - \nu \left(\widehat{\mathrm{CR}}(t) - \mathrm{CR}_0\right) > 0, \forall t \geq t_0. \tag{3.6}$$

Then

$$\widehat{\mathrm{CR}}(t) = \mathrm{CR}_0 + \nu f \int_{t_0}^{t} I\,(s)\, ds, \forall t \geq t_0, \tag{3.7}$$

if and only if

$$\tau(t) = \frac{\nu f \left(\dfrac{\widehat{\mathrm{CR}}''(t)}{\widehat{\mathrm{CR}}'(t)} + \nu\right)}{\nu f\, (I_0 + S_0) - \widehat{\mathrm{CR}}'(t) - \nu \left(\widehat{\mathrm{CR}}(t) - \mathrm{CR}_0\right)}. \tag{3.8}$$

Proof. Assume first (3.7) is satisfied. Then by using equation (3.1) we deduce that

$$S_0 \exp\left(-\int_{t_0}^{t} \tau(\sigma) I(\sigma) d\sigma\right) = I_0 + S_0 - I(t) - \nu \mathrm{CI}(t).$$

Therefore

$$\int_{t_0}^{t} \tau(\sigma) I(\sigma) d\sigma = \ln\left[\frac{S_0}{I_0 + S_0 - I(t) - \nu \mathrm{CI}(t)}\right] = \ln(S_0) - \ln\left[I_0 + S_0 - I(t) - \nu \mathrm{CI}(t)\right]$$

therefore by taking the derivative on both side

$$\tau(t) I(t) = \frac{I'(t) + \nu I(t)}{I_0 + S_0 - I(t) - \nu \mathrm{CI}(t)} \Leftrightarrow \tau(t) = \frac{\dfrac{I'(t)}{I(t)} + \nu}{I_0 + S_0 - I(t) - \nu \mathrm{CI}(t)} \tag{3.9}$$

and by using the fact that $\mathrm{CR}(t) - \mathrm{CR}_0 = \nu f \mathrm{CI}(t)$ we obtain (3.8).

Conversely, assume that $\tau(t)$ is given by (3.8). Then if we define $\widetilde{I}(t) = \widehat{\mathrm{CR}}'(t)/\nu f$ and $\widetilde{\mathrm{CI}}(t) = \left(\widehat{\mathrm{CR}}(t) - \mathrm{CR}_0\right)/\nu f$, by using (3.3) we deduce that

$$\widetilde{\mathrm{CI}}(t) = \int_{t_0}^{t} \widetilde{I}(\sigma) d\sigma,$$

and by using (3.4)

$$\widetilde{I}(t_0) = I_0. \tag{3.10}$$

Moreover from (3.8) we deduce that $\widetilde{I}(t)$ satisfies (3.9). By using (3.10) we deduce that $t \to \widetilde{\mathrm{CI}}(t)$ is a solution of (3.1). By uniqueness of the solution of (3.1), we deduce that $\widetilde{\mathrm{CI}}(t) = \mathrm{CI}(t), \forall t \geq t_0$ or equivalently $\mathrm{CR}(t) = \mathrm{CR}_0 + \nu f \int_{t_0}^{t} I(s)\, ds, \forall t \geq t_0$. The proof is completed. ∎

The formula (3.8) was already obtained by Hadeler [21, see Corollary 2].

4 Explicit formula for $\tau(t)$ and I_0

Many phenomenological models have been compared to the data during the first phase of the COVID-19 outbreak. We refer to the paper of Tsoularis and Wallace [24] for a nice survey on the generalized logistic equations. Let us consider here for example, the Bernoulli-Verhulst equation

$$\mathrm{CR}'(t) = \chi_2 \mathrm{CR}(t) \left(1 - \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^{\theta}\right), \forall t \geq t_0, \tag{4.1}$$

supplemented with the initial data

$$\mathrm{CR}(t_0) = \mathrm{CR}_0 \geq 0.$$

Let us recall the explicit formula for the solution of (4.1)

$$\mathrm{CR}(t) = \frac{e^{\chi_2 (t-t_0)} \mathrm{CR}_0}{\left[1 + \dfrac{\chi_2 \theta}{\mathrm{CR}_\infty^{\theta}} \int_{t_0}^{t} \left(e^{\chi_2 (\sigma - t_0)} \mathrm{CR}_0\right)^{\theta} d\sigma\right]^{1/\theta}} = \frac{e^{\chi_2 (t-t_0)} \mathrm{CR}_0}{\left[1 + \dfrac{\mathrm{CR}_0^{\theta}}{\mathrm{CR}_\infty^{\theta}} \left(e^{\chi_2 \theta (t-t_0)} - 1\right)\right]^{1/\theta}}. \tag{4.2}$$

Assumption 4.1 *We assume that the cumulative numbers of reported cases* $\mathrm{CR}_{\mathrm{Data}}(t_i)$ *are known for a sequence of times* $t_0 < t_1 < \cdots < t_{n+1}$.

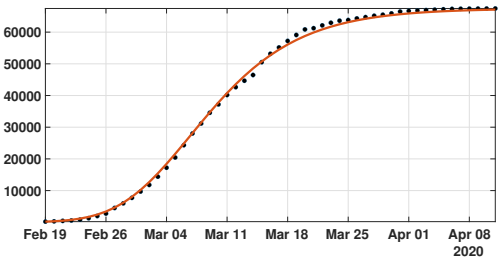

Figure 3: *In this figure, we plot the best fit of the Bernoulli-Verhulst model to the cumulative number of reported cases of COVID-19 in China. We obtain* $\chi_2 = 0.66$ *and* $\theta = 0.22$. *The black dots correspond to data for the cumulative number of reported cases and the blue curve corresponds to the model.*

Estimated initial number of infected

By combining (1.3) and the Bernoulli-Verhulst equation (4.1) for $t \to \mathrm{CR}(t)$, we deduce the initial number of infected

$$I_0 = \frac{\mathrm{CR}'(t_0)}{\nu\, f} = \frac{\chi_2\, \mathrm{CR}_0 \left(1 - \left(\frac{\mathrm{CR}_0}{\mathrm{CR}_\infty}\right)^\theta\right)}{\nu\, f}. \tag{4.3}$$

Remark 4.2 *We fix* $f = 0.5$, *from the COVID-19 data in mainland China and formula (4.3) (with* $\mathrm{CR}_0 = 198$*), we obtain*

$$I_0 = 1909 \text{ for } \nu = 0.1,$$

and

$$I_0 = 954 \text{ for } \nu = 0.2.$$

By using (4.1) we deduce that

$$\begin{aligned}\mathrm{CR}''(t) &= \chi_2\, \mathrm{CR}'(t) \left(1 - \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^\theta\right) - \frac{\chi_2 \theta}{\mathrm{CR}_\infty^\theta}\, \mathrm{CR}(t)\, (\mathrm{CR}(t))^{\theta-1}\, \mathrm{CR}'(t) \\ &= \chi_2\, \mathrm{CR}'(t) \left(1 - \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^\theta\right) - \frac{\chi_2 \theta}{\mathrm{CR}_\infty^\theta}\, (\mathrm{CR}(t))^\theta\, \mathrm{CR}'(t),\end{aligned}$$

therefore

$$\mathrm{CR}''(t) = \chi_2\, \mathrm{CR}'(t) \left(1 - (1+\theta) \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^\theta\right). \tag{4.4}$$

Estimated rate of transmission

By using the Bernoulli-Verhulst equation (4.1) and substituting (4.4) in (3.8), we obtain

$$\tau(t) = \frac{\nu\, f \left(\chi_2 \left(1 - (1+\theta) \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^\theta\right) + \nu\right)}{\nu\, f\, (I_0 + S_0) + \nu \mathrm{CR}_0 - \mathrm{CR}(t) \left(\chi_2 \left(1 - \left(\frac{\mathrm{CR}(t)}{\mathrm{CR}_\infty}\right)^\theta\right) + \nu\right)}. \tag{4.5}$$

This formula (4.5) combined with (4.2) gives an explicit formula for the rate of transmission.

Since $\mathrm{CR}(t) < \mathrm{CR}_\infty$, by considering the sign of the numerator and the denominator of (4.5), we obtain the following proposition.

Proposition 4.3 *The rate of transmission* $\tau(t)$ *given by (4.5) is non negative for all* $t \geq t_0$ *if*

$$\nu \geq \chi_2\, \theta, \tag{4.6}$$

and

$$f\, (I_0 + S_0) + \nu \mathrm{CR}_0 > \mathrm{CR}_\infty\, (\chi_2 + \nu)\,. \tag{4.7}$$

Compatibility of the model SI with the COVID-19 data for mainland China

The model SI is compatible with the data only when $\tau(t)$ stays positive for all $t \geq t_0$. From our estimation of the Chinese's COVID-19 data we obtain $\chi_2\,\theta = 0.14$. Therefore from (4.6) we deduce that model is compatible with the data only when

$$1/\nu \leq 1/0.14 = 3.3 \text{ days}. \tag{4.8}$$

This means that the average duration of infectious period $1/\nu$ must be shorter than 3.3 days.

Similarly the condition (4.7) implies

$$f \geq \frac{\mathrm{CR}_\infty \chi_2 + (\mathrm{CR}_\infty - \mathrm{CR}_0)\,\nu}{S_0 + I_0} \geq \frac{\mathrm{CR}_\infty \chi_2 + (\mathrm{CR}_\infty - \mathrm{CR}_0)\,\chi_2\,\theta}{S_0 + I_0}$$

and since we have $CR_0 = 198$ and $\mathrm{CR}_\infty = 67102$, we obtain

$$f \geq \frac{67102 \times 0.66 + (67102 - 198) \times 0.14}{1.4 \times 10^9} \geq 3.83 \times 10^{-5}. \tag{4.9}$$

So according to this estimation the fraction of unreported $0 < f \leq 1$ can be almost as small as we want.

Figure 4 illustrates the Proposition 4.3. We observe that the formula for the rate of transmission (4.5) becomes negative whenever $\nu < \chi_2\theta$.

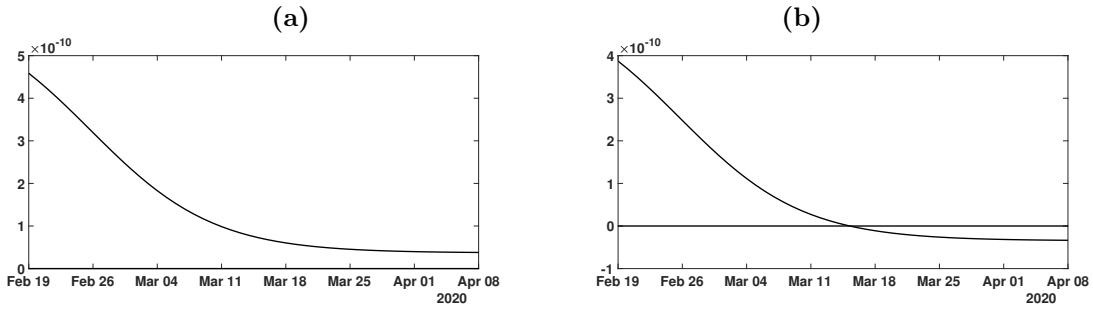

Figure 4: *In this figure, we plot the rate of transmission obtained from formula (4.5) with* $f = 0.5$, $\chi_2\,\theta = 0.14 < \nu = 0.2$ *(in Figure (a)) and* $\nu = 0.1 < \chi_2\,\theta = 0.14$ *(in Figure (b)),* $\chi_2 = 0.66$ *and* $\theta = 0.22$ *and* $\mathrm{CR}_\infty = 67102$ *which is the latest value obtained from the cumulative number of reported cases for China.*

In Figure 5 we plot the numerical simulation obtained from (1.1)-(1.3) when $t \to \tau(t)$ is replaced by the explicit formula (4.5). It is surprising that we can reproduce perfectly to the original Bernoulli-Verhulst even when $\tau(t)$ becomes negative. This was not guaranteed at first, since the I-class of individuals is losing some individuals which are recovering.

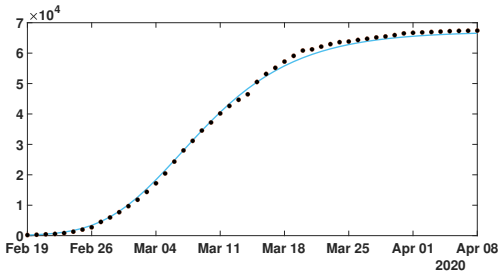

Figure 5: *In this figure, we plot the number of reported cases by using model (1.1) and (1.3), and the rate of transmission is obtained in (4.5). The parameters values are* $f = 0.5$, $\nu = 0.1$ *or* $\nu = 0.2$, $\chi_2 = 0.66$ *and* $\theta = 0.22$ *and* $\mathrm{CR}_\infty = 67102$ *is the latest value obtained from the cumulative number of reported cases for China. Furthermore, we use* $S_0 = 1.4 \times 10^9$ *for the total population of China and* $I_0 = 954$ *which is obtained from formula (4.3). The black dots correspond to data for the cumulative number of reported cases observed and the blue curve corresponds to the model.*

5 Computing numerically a day by day piecewise constant rate of transmission

Assumption 5.1 *We assume that the rate of transmission* $\tau(t)$ *is piecewise constant and for each* $i = 0, \dots, n$,

$$\tau(t) = \tau_i, \ \text{whenever } t_i \le t < t_{i+1}. \tag{5.1}$$

For $t \in [t_{i-1}, t_i]$, we deduce by using Assumption 5.1 that

$$\int_{t_0}^{t} \tau(\sigma)\, \mathrm{CR}'(\sigma)\mathrm{d}\sigma = \sum_{j=0}^{i-2} \int_{t_j}^{t_{j+1}} \tau_j\, \mathrm{CR}'(\sigma)\mathrm{d}\sigma + \int_{t_{i-1}}^{t} \tau_{i-1}\, \mathrm{CR}'(\sigma)\mathrm{d}\sigma.$$

Therefore by using (3.2), for $t \in [t_{i-1}, t_i]$, we obtain

$$\mathrm{CR}'(t) = \nu f \left(I_0 + S_0 \left[1 - \Pi_{i-1} \exp\left(-\frac{\tau_{i-1}}{\nu f} \left[\mathrm{CR}(t) - \mathrm{CR}(t_{i-1}) \right] \right) \right] \right) + \nu\, \mathrm{CR}_0 - \nu \mathrm{CR}(t), \tag{5.2}$$

where

$$\Pi_{i-1} = \exp\left(-\sum_{j=0}^{i-2} \frac{\tau_j}{\nu f} \left[\mathrm{CR}(t_{j+1}) - \mathrm{CR}(t_j) \right] \right). \tag{5.3}$$

By fixing $\tau_{i-1} = 0$ on the right hand side of (5.2) we get

$$\mathrm{CR}'(t) \ge \nu f \left(I_0 + S_0 \left[1 - \Pi_{i-1} \right] \right) + \nu\, \mathrm{CR}_0 - \nu \mathrm{CR}(t),$$

and when $\tau_{i-1} \to \infty$ we obtain

$$\mathrm{CR}'(t) \le \nu f \left(I_0 + S_0 \right) + \nu\, \mathrm{CR}_0 - \nu \mathrm{CR}(t).$$

By using the theory of monotone ordinary differential equations (see Smith [23]) we deduce that the map $\tau_i \to \mathrm{CR}(t_i)$ is monotone increasing, and we get the following result.

Theorem 5.2 *Let assumptions 1.1, 4.1 and 5.1 be satisfied. Let* I_0 *be fixed. Then we can find a unique sequence* $\tau_0, \tau_1, \dots, \tau_n$ *of non negative numbers such that* $t \to \mathrm{CR}(t)$ *the solution of (3.2) fits exactly the data at any time* t_i, *that is to say that*

$$\mathrm{CR}(t_i) = \mathrm{CR}_{\mathrm{Data}}(t_i), \forall i = 1, \dots, n+1,$$

if and only if the two following two conditions are satisfied for each $i = 0, 1, \dots, n+1$,

$$\mathrm{CR}_{\mathrm{Data}}(t_i) \ge e^{-\nu(t_i - t_{i_1})} \mathrm{CR}_{\mathrm{Data}}(t_{i-1}) + \int_{t_{i-1}}^{t_i} \nu e^{-\nu(t_i - \sigma)} \mathrm{d}\sigma \left(f \left(I_0 + S_0 \left[1 - \Pi_{i-1}^{\mathrm{Data}} \right] \right) + \mathrm{CR}_0 \right), \tag{5.4}$$

where

$$\Pi_{i-1}^{\text{Data}} = \exp\left(-\sum_{j=0}^{i-2} \frac{\tau_j}{\nu\, f}\left[\text{CR}_{\text{Data}}(t_{j+1}) - \text{CR}_{\text{Data}}(t_j)\right]\right), \tag{5.5}$$

and

$$\text{CR}_{\text{Data}}(t_i) \le e^{-\nu(t_i - t_{i_1})}\text{CR}_{\text{Data}}(t_{i-1}) + \int_{t_{i-1}}^{t_i} \nu e^{-\nu(t_i-\sigma)}\mathrm{d}\sigma\left(f\left(I_0 + S_0\right) + \text{CR}_0\right). \tag{5.6}$$

Remark 5.3 *The above theorem means that the data are identifiable for this model SI if and only if the conditions (5.4) and (5.6) are satisfied. Moreover, in that case, we can find a unique sequence of transmission rates $\tau_i \ge 0$ which gives a perfect fit to the data.*

6 Numerical simulations

In this section, we propose a numerical method to fit the day by day rate of transmission. The goal is to take advantage of the monotone property of $\text{CR}(t)$ with respect to τ_i on the time interval $[t_i, t_{i+1}]$. Recently more sophisticated methods were proposed by Bakha et al. [20] by using several types of approximation methods for the rate of transmission.

We start with the simplest Algorithm 1 in order to show the difficulties to identify the rate of transmission.

Algorithm 1

Step 1: *We fix $S_0 = 1.4 \times 10^9$, $\nu = 0.1$ or $\nu = 0.2$ and $f = 0.5$. We consider the system*

$$\begin{cases} S'(t) = -\tau S(t)I(t), \\ I'(t) = \tau S(t)I(t) - \nu I(t), \\ \text{CR}'(t) = \nu f I(t), \end{cases} \tag{6.1}$$

on the interval of time $t \in [t_0, t_1]$. This system is supplemented by initial values $S(t_0) = S_0$ and $I(t_0) = I_0$ is given by formula (2.4) (if we consider the data only at the early stage) or formula (4.3) (if we consider all the data) and $\text{CR}(t_0) = \text{CR}_{\text{Data}}(t_0)$ is obtained from the data.

The map $\tau \to \text{CR}(t_1)$ being monotone increasing, we can apply a bisection method to find the unique value τ_0 solving

$$\text{CR}(t_1) = \text{CR}_{\text{Data}}(t_1).$$

Then we proceed by induction.

Step i: *For each integer $i = 1, \ldots, n$ we consider the system*

$$\begin{cases} S'(t) = -\tau S(t)I(t), \\ I'(t) = \tau S(t)I(t) - \nu I(t), \\ \text{CR}'(t) = \nu f I(t), \end{cases} \tag{6.2}$$

on the interval of time $t \in [t_i, t_{i+1}]$. This system is supplemented by initial values $S(t_i)$ and $I(t_i)$ obtained from the previous iteration and with $\text{CR}(t_i) = \text{CR}_{\text{Data}}(t_i)$ obtained from the data.

The map $\tau \to \text{CR}(t_i)$ being monotone increasing, we can apply a bisection method to find the unique value τ_i solving

$$\text{CR}(t_i) = \text{CR}_{\text{Data}}(t_i).$$

In Figure 6, we plot an example of such a perfect fit, which is the same for $\nu = 0.1$ and $\nu = 0.2$. In Figure 7 we plot the rate of transmission obtained numerically for $\nu = 0.2$ in (a) and $\nu = 0.1$ in (b). This is an example of a negative rate of transmission. Figure 7 should be compared to Figure 4 which gives similar result.

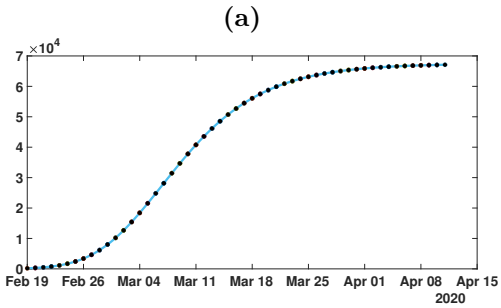

Figure 6: *In this figure, we plot the perfect fit of the cumulative number of reported cases of COVID-19 in China. We fix the parameters* $f = 0.5$ *and* $\nu = 0.2$ *or* $\nu = 0.1$ *and we apply our algorithm 1 to obtain the perfect fit. The black dots correspond to data for the cumulative number of reported cases and the blue curve corresponds to the model.*

In Figures 8-10 we use Algorithm 1 and we plot the rate of transmission obtained by using the reported cases of COVID-19 in China where the parameters are fixed as $f = 0.5$ and $\nu = 0.2$. In Figures 8-10, we observe an oscillating rate of transmission which is alternatively positive and negative back and forth. These oscillations are due to the amplification of the error in the numerical method itself. In Figure 8, we run the same simulation than in Figure 9 but during a shorter period. In Figure 8, we can see that the slope of $\mathrm{CR}(t)$ at the $t = t_i$ between two days (the black dots) is amplified one day to the next.

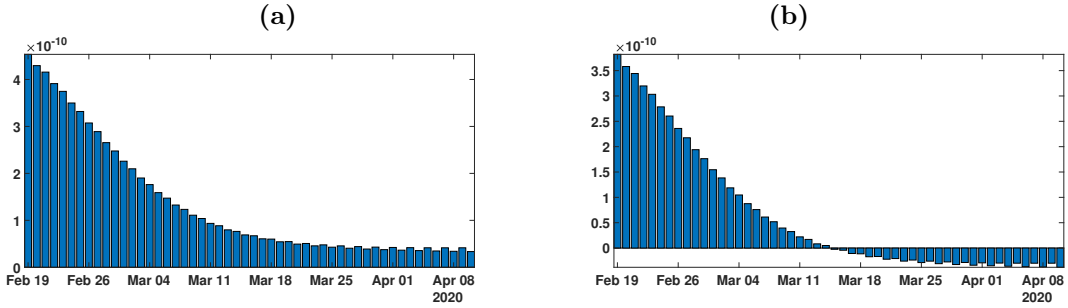

Figure 7: *In this figure, we plot the rate of transmission obtained for the reported cases of COVID-19 in China with the parameters* $f = 0.5$ *and* $\nu = 0.2$ *in figure (a) and* $\nu = 0.1$ *in figure (b). This rate of transmission corresponds to the perfect fit obtained in Figure 6.*

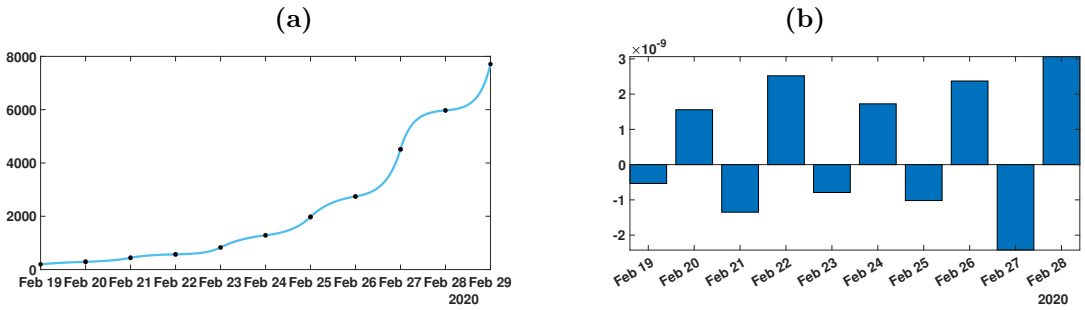

Figure 8: *In figure (a), we plot the cumulative number of reported cases obtained from the data (black dots) and the model (blue curve). In figure (b), we plot the daily rate of transmission obtained by using Algorithm 1. We see that we can fit the data perfectly. But the method is very unstable. We obtain a rate of transmission that oscillates from positive to negative values back and forth.*

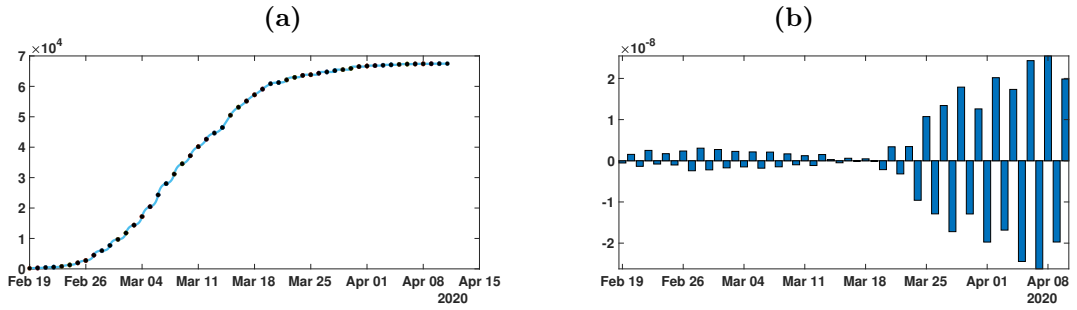

Figure 9: *In figure (a), we plot the cumulative number of reported cases obtained from the data (black dots) and the model (blue curve). In figure (b), we plot the daily rate of transmission obtained by using Algorithm 1. We see that we can fit the data perfectly. But the method is very unstable. We obtain a rate of transmission that oscillates from positive to negative values back and forth.*

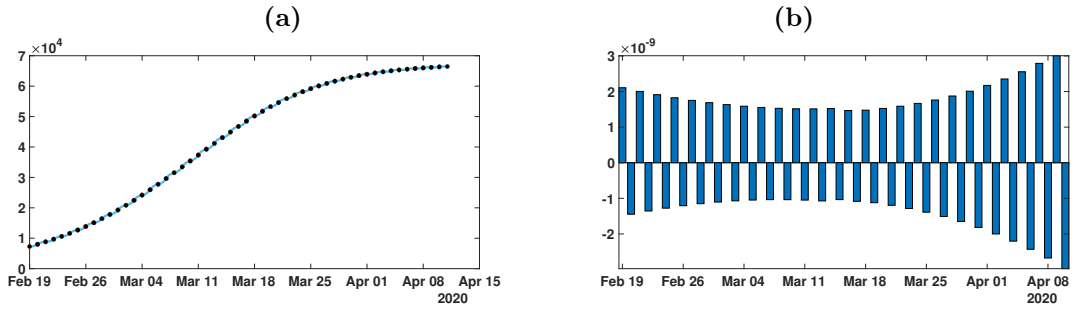

Figure 10: *We apply Algorithm 1 to the regularized data. In figure (a), we plot the regularized cumulative number of reported cases obtained from the data (black dots) and the model (blue curve). In figure (b), we plot the daily rate of transmission obtained by using Algorithm 1. We see that we can fit the data perfectly. But the method is very unstable. We obtain a rate of transmission that oscillates from positive to negative values back and forth.*

In Figure 10, we first smooth the original cumulative data by using the MATLAB function $\mathrm{CR}_{\mathrm{Data}} = \mathrm{smoothdata}(\mathrm{CR}_{\mathrm{Data}}, 'gaussian', 50)$ to regularize the data and we apply Algorithm 1. Unfortunately, smoothing the data does not help to solve the instability problem in Figure 10.

We need to introduce a correction when choosing the next initial value $I(t_i)$. In Algorithm 1 the errors are due to the following relationship which is not respected

$$\mathrm{CR}'(t) = \nu f I(t)$$

at the points $t = t_i$ which should be reflected by the algorithm.

In Figure 11, we smooth the data first by using the MATLAB function $\mathrm{CR}_{\mathrm{Data}} = \mathrm{smoothdata}(\mathrm{CR}_{\mathrm{Data}}, 'gaussian', 50)$, and we apply Algorithm 2 by approximating equation (6.6) by

$$I_i = [\mathrm{CR}_{\mathrm{Data}}(t_i) - \mathrm{CR}_{\mathrm{Data}}(t_{i-1})]/(\nu \times f). \tag{6.3}$$

In Figure 11 we no longer observe the oscillations of the rate of transmission.

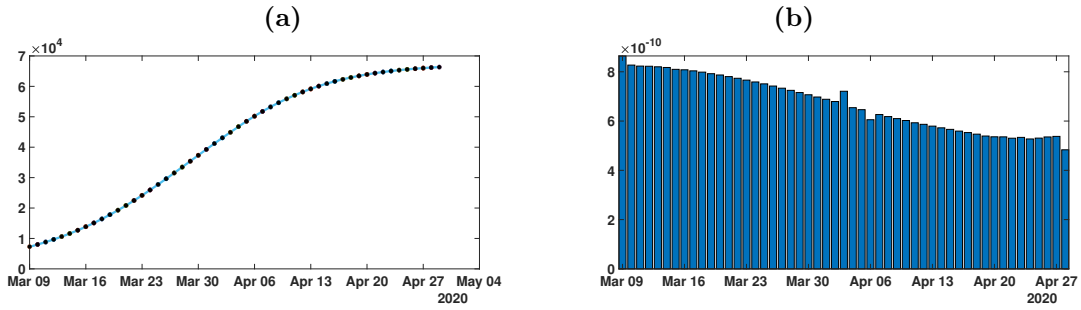

Figure 11: *In this figure, we plot the rate of transmission obtained by using the reported cases of COVID-19 in China with the parameters* $f = 0.5$ *and* $\nu = 0.2$. *We first regularize the data by applying the MATLAB function* $\mathrm{CR}_{\mathrm{Data}} = \mathrm{smoothdata}(\mathrm{CR}_{\mathrm{Data}}, 'gaussian', 50)$. *Then we apply Algorithm 2 to the regularized data. In figure (a), we plot the regularized cumulative number of reported cases obtained after smoothing (black dots) and the model (blue curve). In figure (b), we plot the daily rate of transmission obtained by using the Algorithm 2. We see that we can fit the data perfectly and this time the rate of transmission is becoming reasonable.*

Algorithm 2

We fix $S_0 = 1.4 \times 10^9$, $\nu = 0.1$ *or* $\nu = 0.2$ *and* $f = 0.5$. *Then we fit the data by using the method described in Section 2 to estimate the parameters* χ_1, χ_2 *and* χ_3 *from day* 1 *to* 10. *Then we use*

$$\begin{aligned} &S_0 = 1.40005 \times 10^9, \\ &I_0 = \chi_2\, \chi_1\, [\exp(\chi_2\, (t_0 - 1))]/(f\, \nu), \\ &\mathrm{CR}_0 = \chi_1\, exp(\chi_2\, t_0) - \chi_3. \end{aligned} \tag{6.4}$$

For each integer $i = 0, \ldots, n$, *we consider the system*

$$\begin{cases} S'(t) = -\tau S(t) I(t), \\ I'(t) = \tau S(t) I(t) - \nu I(t), \\ \mathrm{CR}'(t) = \nu f I(t), \end{cases} \tag{6.5}$$

for $t \in [t_i, t_{i+1}]$. *Then the map* $\tau \to \mathrm{CR}(t_{i+1})$ *being monotone increasing, we can apply a bisection method to find the unique* τ_i *solving*

$$\mathrm{CR}(t_{i+1}) = \mathrm{CR}_{\mathrm{Data}}(t_{i+1}).$$

The key idea of this new algorithm is the following correction on the I-component of the system. We start a new step by using the value $S(t_i)$ *obtained from the previous iteration and*

$$I_i = \mathrm{CR}'_{\mathrm{Data}}(t_i)/(\nu\, f), \tag{6.6}$$

and

$$\mathrm{CR}_i = \mathrm{CR}_{\mathrm{Data}}(t_i). \tag{6.7}$$

In Figure 12 we plot several types of regularized cumulative data in figure (a) and several types of regularized daily data in figure (b). Among the different regularization methods, an important one is the Bernoulli-Verhulst best fit approximation.

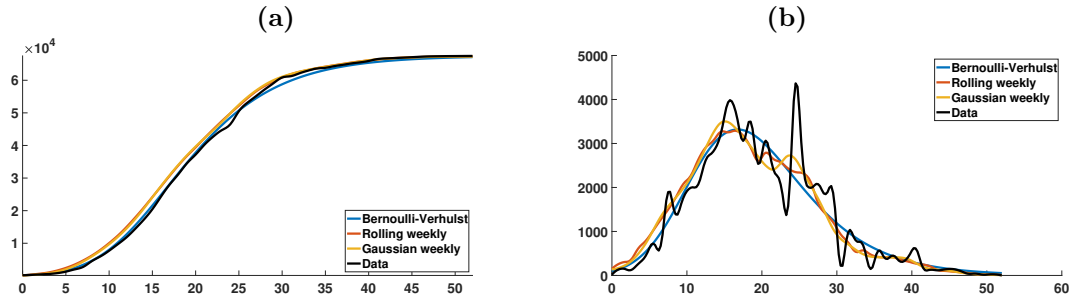

Figure 12: *In this figure, we plot the cumulative number of reported cases (left) and the daily number of reported cases (right). The black curves are obtained by applying the cubic spline matlab function "spline(Days,DATA)" to the cumulative data. The left-hand side is obtained by using the cubic spline function and right-hand side is obtained by using the derivative of the cubic spline interpolation. The blue curves are obtained by using cubic spline function to the day by day values of cumulative number of cases obtained from the best fit of the Bernoulli-Verhulst model. The orange curves are obtained by computing the rolling weekly daily number of cases (we use the matlab function "smoothdata(DAILY,'movmean',7)") and then by applying the cubic spline function the corresponding cumulative number of cases. The yellow curves are obtained by Gaussian the rolling weekly to the daily number of cases (we use the matlab function "smoothdata(DAILY,'gaussian',7)") and then by applying the cubic spline function to the corresponding cumulative number of cases.*

In Figure 13 we plot the rate of transmission $t \to \tau(t)$ obtained by using Algorithm 2. We can see that the original data gives a negative transmission rate while at the other extreme the Bernoulli-Verhulst seems to give the most regularized transmission rate. In Figure 13-(a) we observe that we now recover almost perfectly the theoretical transmission rate obtained in Section 4. In Figure 13-(b) the rolling weekly average regularization and in Figure 13-(c) the Gaussian weekly average regularization still vary a lot and in both cases the transmission rate becomes negative after some time. In Figure 13-(c) the original data gives a transmission rate that is negative from the beginning. We conclude that it is crucial to find a "good" regularization of the daily number of case. So far the best regularization method is obtained by using the best fit of the Bernoulli-Verhulst model.

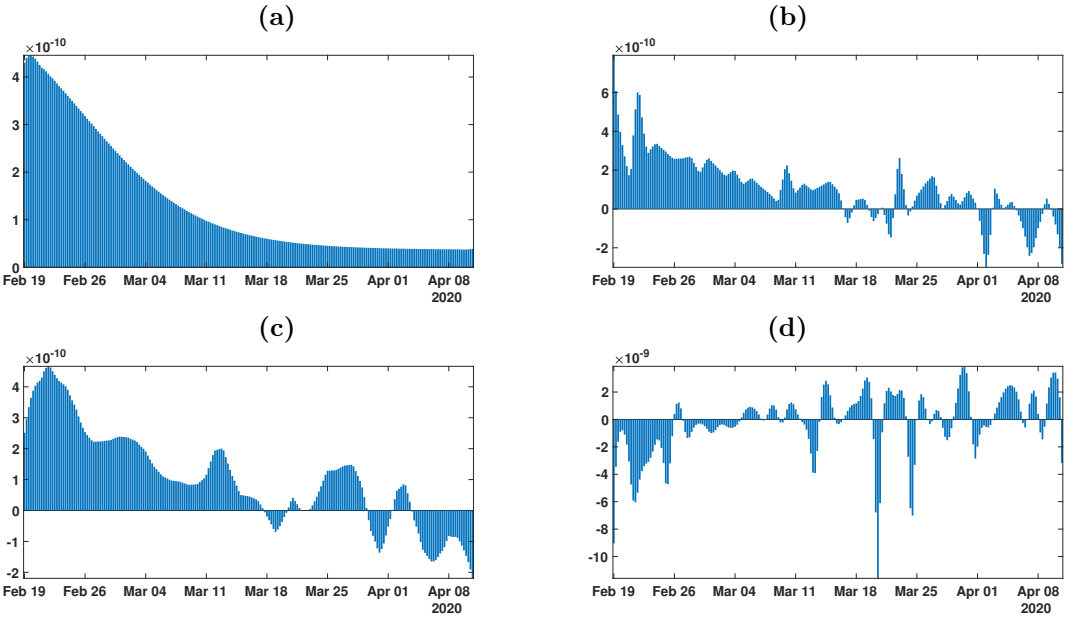

Figure 13: *In this figure we plot the transmission rates $t \to \tau(t)$ obtained by using Algorithm 2 with the parameters $f = 0.5$ and $\nu = 0.2$. In figure (a) we use the cumulative data obtained by using the Bernoulli-Verhulst regularization. In figure (b) we use the cumulative data obtained by using the rolling weekly average regularization. In figure (c) we use the cumulative data obtained by using the Gaussian weekly average regularization. In figure (d) we use the original cumulative data.*

Remark 6.1 *For each simulation Figure 13-(b) and Figure 13-(c), it is possible to obtain a transmission* $t \to \tau(t)$ *that is non negative for all time* t *by increasing sufficiently the parameter* ν*. Nevertheless, we do not present these simulations here because the corresponding values of* ν *to obtain a non negative* $\tau(t)$ *are unrealistic.*

In Figure 14 (a) (b) (c) and (d) (respectively) we plot the daily basic reproduction number corresponding to the Figure 13 (a) (b) (c) and (d) (respectively). The red line corresponds to $R_0 = 1$. We see some complex behavior for the Figure 14 (b) (c) and the figure (d) is again unrealistic.

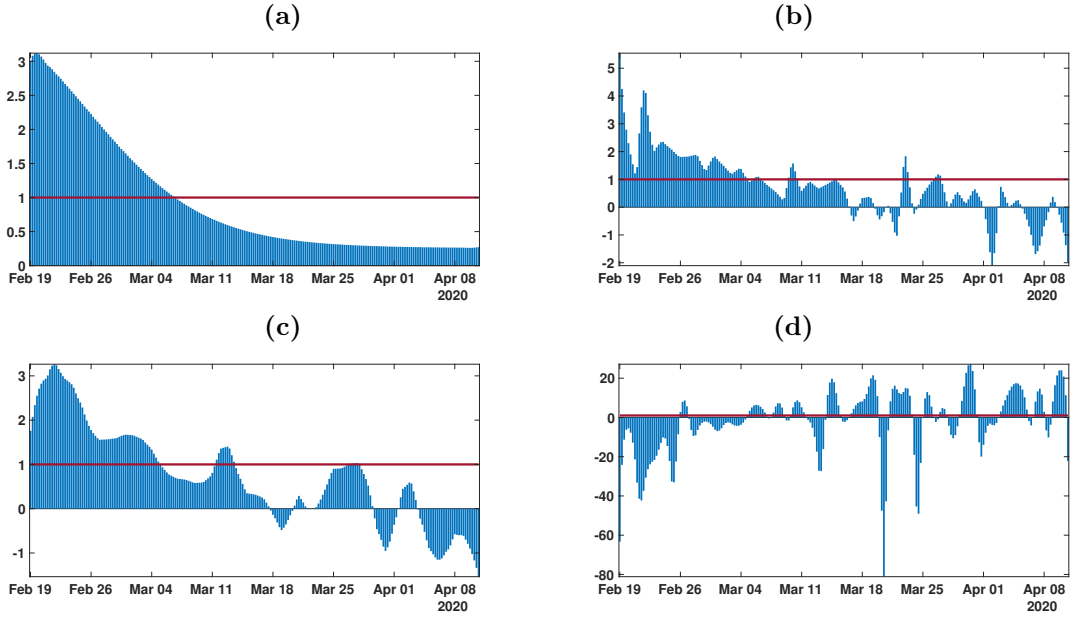

Figure 14: *In this figure we plot the daily basic reproduction number* $t \to R_0(t) = \tau(t)S(t)/\nu$ *obtained by using Algorithm 2 with the parameters* $f = 0.5$ *and* $\nu = 0.2$*. In figure (a) we use the cumulative data obtained by using the Bernoulli-Verhulst regularization. In figure (b) we use the cumulative data obtained by using the rolling weekly average regularization. In figure (c) we use the cumulative data obtained by using the Gaussian weekly average regularization. In figure (d) we use the original cumulative data.*

7 Discussion

Estimating the parameters of an epidemiological model is always difficult and generally requires strong assumptions about their value and their consistency and constancy over time. Despite this, it is often shown that many sets of parameter values are compatible with a good fit of the observed data. The new approach developed in this article consists first of all in postulating a phenomenological model of growth of infectious, based on the very classic model of Verhulst, proposed in demography in 1838 [28]. Then, obtaining explicit formulas for important parameter values such as the transmission rate or the initial number of infected (or for lower and/or upper limits of these values), gives an estimate allowing an almost perfect reconstruction of the observed dynamics.

The uses of phenomenological models can also be regarded as a way a of smoothing the data. Indeed, the errors concerning the observations of new infected cases are numerous:

- the census is rarely regular and many countries report late cases that occurred during the weekend and at varying times over-add data from specific counts, such as those from homes for the elderly;
- the number of cases observed is still underestimated and the calculation of cases not reported new cases of infected is always a difficult problem [18];
- the raw data are sometimes reduced for medical reasons of poor diagnosis or lack of detection tools, or for reasons of domestic policy of states.

For all these causes of error, it is important to choose the appropriate smoothing method (moving average, spline, Gaussian kernel, auto-regression, generalized linear model, etc.). In this article, several methods were used and the one which allowed the model to perfectly match the smoothed data was retained.

In this article, we developed several methods to understand how to reconstruct the rate of transmission from the data. In Section 2, we reconsidered the method presented in [18] based on an exponential fit of the early data. The approach gives a first estimation of I_0 and τ_0. In Section 3, we prove a result to connect the time dependent cumulative reported data and the transmission rate. In Section 4, we compare the data to the Bernoulli-Verhulst model and we use this model as a phenomenological model. The Bernoulli-Verhulst model fits the data for mainland China very well. Next by replacing the data by the solution of the Bernoulli-Verhulst model, we obtain an explicit formula for the transmission rate. So we derive some conditions on the parameters for the applicability of the SI model to the data for mainland China. In Section 5, we discretized the rate of transmission and we observed that given some daily cumulative data, we can get at most one perfect fit the data. Therefore, in Section 6, we provide two algorithms to compute numerically the daily rates of transmission. Such numerical questions turn out to be a delicate problem. This problem was previously considered by another French group Bakhta, Boiveau, Maday and Mula [20]. Here we use some simple ideas to approach the derivative of the cumulative reported cases combined with some smoothing method applied to the data.

To conclude this article we plot the daily basic reproduction number

$$R_0(t) = \frac{\tau(t)S(t)}{\nu}$$

as a function of the time t and the parameters f or ν. The above simple formula for R_0 is not the real basic reproductive number in the sense of the number of newly infected produced by a single infectious. But this is a simple formula which gives a tendency about the growth or decay of the number of infectious. In Figure 14-(a), the daily basic reproduction number is almost independent of f, while in Figure 14-(b), $R_0(t)$ is depending on ν mostly for the small value of ν. The red curve on each surfaces in Figure 14 corresponds to the turning point (i.e. time $t \geq t_0$ for which $R_0(t) = 1$). We also see that turning point is not depending much on these parameters.

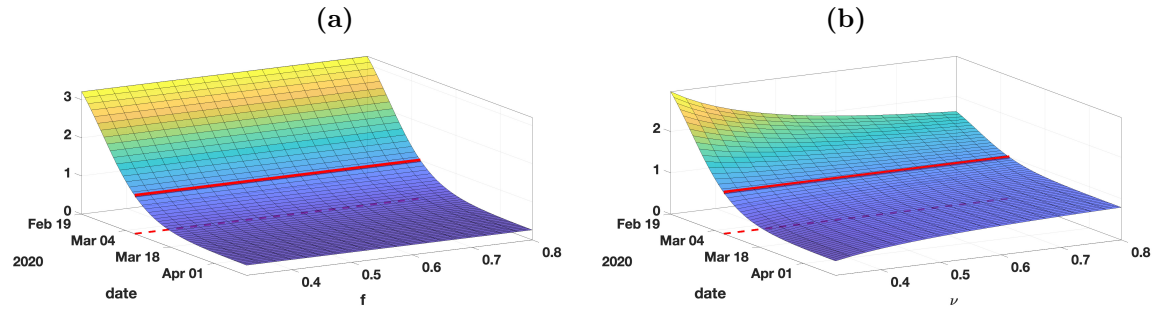

Figure 15: *In this figure plot* $R_0(t) = \frac{\tau(t)S(t)}{\nu}$ *the daily basic reproduction number and we vary the parameter* f *(left) and* ν *(right).*

Concerning contagious diseases, public health physicians are constantly facing four challenges. The first concerns the estimation of the average transmission rate. Until now, no explicit formula had been obtained in the case of the SIR model, according to the observed data of the epidemic, that is to say the number of reported cases of infected patients. Here, from realistic simplifying assumptions, a formula is provided (formula (4.5)), making it possible to accurately reconstruct theoretically the curve of the observed cumulative cases. The second challenge concerns the estimation of the mean duration of the infectious period for infected patients. As for the transmission rate, the same realistic assumptions make it possible to obtain an upper limit to this duration (inequality (4.8)), which makes it possible to better guide the individual quarantine measures decided by the authorities in charge of public health. This upper bound also makes it possible to obtain a lower bound for the percentage of unreported infected patients (inequality (4.8)), which gives an idea of the quality of the census of cases of infected patients, which is the third challenge faced by epidemiologists, specialists of contagious diseases. The fourth challenge is the estimation of the average transmission rate for each day of the infectious period (dependent on the distribution of the transmission over the "ages" of infectivity), which will be the subject of further work and which poses formidable problems, in particular those related to the age (biological age or civil age) class of the patients concerned. Another interesting prospect is the extension of methods developed in the present paper to the contagious non-infectious diseases (i.e., without causal infectious agent), such as social contagious diseases, the best example being that of the pandemic linked to obesity [25, 26, 27], for which many concepts and modeling methods remain available.

A Supplementary tables

We use cumulative reported data from the National Health Commission of the People's Republic of China and the Chinese CDC for mainland China. Before February 11, the data was based on confirmed testing. From February 11 to February 15, the data included cases that were not tested for the virus, but were clinically diagnosed based on medical imaging showing signs of pneumonia. There were 17,409such cases from February 10 to February 15. The data from February 10 to February 15 specified both types of reported cases. From February 16, the data did not separate the two types of reporting, but reported the sum of both types. We subtracted 17,409 cases from the cumulative reported cases after February 15 to obtain the cumulative reported cases based only on confirmed testing after February 15.The data is given in Table 2 with this adjustment.

January						
19	20	21	22	23	24	25
198	291	440	571	830	1287	1975
26	27	28	29	30	31	
2744	4515	5974	7711	9692	11791	
February						
1	2	3	4	5	6	7
14380	17205	20438	24324	28018	31161	34546
8	9	10	11	12	13	14
37198	40171	42638	44653	46472	48467	49970
15	16	17	18	19	20	21
51091	70548 − 17409	72436 − 17409	74185 − 17409	75002 − 17409	75891 − 17409	76288 − 17409
22	23	24	25	26	27	28
76936 − 17409	77150 − 17409	77658 − 17409	78064 − 17409	78497 − 17409	78824 − 17409	79251 − 17409
29						
79824 − 17409						
March						
1	2	3	4	5	6	7
79824 − 17409	79824 − 17409	79824 − 17409	80409 − 17409	80552 − 17409	80651 − 17409	80695 − 17409
8	9	10	11	12	13	14
80735 − 17409	80754 − 17409	80778 − 17409	80793 − 17409	80813 − 17409	80824 − 17409	80844 − 17409
15	16	17	18			
80860 − 17409	80881 − 17409	80894 − 17409	80928 − 17409			

Table 1: *Cumulative data describing confirmed cases in mainland China from January 20, 2020 to March 18, 2020. The data are taken from [29, 30, 31].*

Funding: This research was funded by the Agence Nationale de la Recherche in France (Project name : MPCUII (PM) and (QG)).

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest." The authors declare no conflict of interest.

References

[1] P. Magal and S. Ruan, Susceptible-Infectious-Recovered Models Revisited: From the Individual Level to the Population Level, *Mathematical Biosciences* **250** (2014), 26-40.

[2] Y. Qiu, X. Chen, and W . Shi, Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19), *China. J Popul Econ* **33** (2020), 1127-1172.

[3] H. Zeberg and S. Pääbo, The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* (2020).

[4] P. Guillon, et al., Inhibition of the interaction between the SARS-CoV spike protein and its cellular receptor by anti-histo-blood group antibodies, *Glycobiology* **18.12** (2008), 1085-1093.

[5] A. W. Byrne, et al., Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases, *BMJ Open.* 2020; **10(8)**: e039856.

[6] F. Zhou et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395** (2020), 1054-1062.

[7] Z. Hu et al. Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China Life Sciences* **63** (2020), 706-711.

[8] S. Ma et al. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *medRxiv* 2020.

[9] R. Li et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS- CoV-2). *Science*, **368(6490)** (2020), 489-493.

[10] C. Rothe, Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany, *N Engl J Med* **382** (2020), 970-971.

[11] Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) : https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[12] Z. Yang, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, *Journal of Thoracic Disease* **12.3** (2020), 165.

[13] W. P. London and J. A. Yorke, Recurrent outbreaks of measles, chickenpox and mumps: I. Seasonal variation in contact rates. *American Journal of Epidemiology*, **98(6)** (1973), 453-468.

[14] J. A. Yorke and W. P. London, Recurrent outbreaks of measles, chickenpox and mumps: II. Systematic differences in contact rates and stochastic effects. *American Journal of Epidemiology*, **98(6)** (1973), 469-482.

[15] W. Wang and S. Ruan, Simulating the SARS outbreak in Beijing with limited data, *J. Theoretical Biology* **227** (2004), 369-379.

[16] 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. *Journal of Theoretical Biology*, **229(1)** (2004), 119-126.

[17] A. Smirnova, L. deCamp and G. Chowell, Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model, *Bulletin of Mathematical Biology* **81** (2019), 4343-4365.

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

[19] 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, *Mathematical Biosciences and Engineering* **17(4)** (2020), 3040-3051.

[20] A. Bakhta, T. Boiveau, Y. Maday, O. Mula, Epidemiological short-term Forecasting with Model Reduction of Parametric Compartmental Models. Application to the first pandemic wave of COVID-19 in France. *arXiv preprint* arXiv:2009.09200, 2020.

[21] K.P. Hadeler, Parameter identification in epidemic models, *Math. Biosci.*, **229** (2011), 185-189.

[22] W. C. Roda, M. B. Varughese, D. Han and M. Y. Li, Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* **5** (2020) 271-281.

[23] H. L. Smith, *Monotone Dynamical Systems, an introduction to the theory of competitive and cooperative systems*, Math. Surveys and Monographs, **41**, American Mathematical Society, Providence, Rhode Island (1995).

[24] A. Tsoularis and J. Wallace, Analysis of logistic growth models. *Math. Biosci.*, **179(1)** (2002), 21-55.

[25] J. Demongeot and C. Taramasco, Evolution of social networks: the example of obesity. *Biogerontology*, **15** (2014), 611-626.

[26] J. Demongeot , O. Hansen and C. Taramasco, Complex systems and contagious social diseases: example of obesity. *Virulence*, **7** (2015), 129-140.

[27] J. Demongeot, M. Jelassi and C. Taramasco, From Susceptibility to Frailty in social networks: the case of obesity. *Math. Pop. Studies*, **24** (2017), 219-245.

[28] Pierre-François Verhulst, Notice sur la loi que la population pursuits dans son increase, *Correspondance mathématique et physique*, **10** (1838), 113-121.

[29] Data sourced Wikipedia who used from NHC daily reports : https://en.wikipedia.org/wiki/COVID-19_pandemic_in_mainland_China

[30] The National Health Commission of the People's Republic of China : http://www.nhc.gov.cn/yjb/pzhgli/new_list.shtml

[31] Chinese Center for Disease Control and Prevention : http://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/jszl_11809/