Clarifying predictions for COVID-19 from testing data: The example of New York State

Quentin Griette [a, b, 1], Pierre Magal [a, b, *, 1]

[a] Univ. Bordeaux, IMB, UMR 5251, F-33400, Talence, France
[b] CNRS, IMB, UMR 5251, F-33400, Talence, France

ABSTRACT

With the spread of COVID-19 across the world, a large amount of data on reported cases has become available. We are studying here a potential bias induced by the daily number of tests which may be insufficient or vary over time. Indeed, tests are hard to produce at the early stage of the epidemic and can therefore be a limiting factor in the detection of cases. Such a limitation may have a strong impact on the reported cases data. Indeed, some cases may be missing from the official count because the number of tests was not sufficient on a given day. In this work, we propose a new differential equation epidemic model which uses the daily number of tests as an input. We obtain a good agreement between the model simulations and the reported cases data coming from the state of New York. We also explore the relationship between the dynamic of the number of tests and the dynamics of the cases. We obtain a good match between the data and the outcome of the model. Finally, by multiplying the number of tests by 2, 5, 10, and 100 we explore the consequences for the number of reported cases.

© 2021 The Authors. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The epidemic of novel coronavirus (COVID-19) infections began in China in December 2019 and rapidly spread worldwide in 2020. Since the early beginning of the epidemic, mathematicians and epidemiologists have developed models to analyze the data and characterize the spread of the virus, and attempt to project the future evolution of the epidemic. Many of those models are based on the SIR or SEIR model which is classical in the context of epidemics. We refer to (Tang et al., 2020; Wu et al., 2020) for the earliest articles devoted to such a question and to (Anderson & May 1991; Bailey, 1957; Brauer & Castillo-Chavez, 2000; Brauer et al., 2008, 2019; Busenberg & Cooke, 1993; Diekmann et al., 2013; Hethcote, 2000; Keeling & Rohani, 2007; Murray, 1993; Thieme, 2003) for more models. In the course of the COVID-19 outbreak, it became clear for the scientific community that covert cases (asymptomatic or unreported infectious case) play an important role. An early description of an asymptomatic transmission in Germany was reported by Rothe et al. (Rothe et al., 2020). It was also observed on the Diamond Princess cruise ship in Yokohama in Japan by Mizumoto et al. (Mizumoto et al., 2020) that many of the passengers were tested

---

positive to the virus, but never presented any symptoms. We also refer to Qiu (Qiu, 2020) for more information about this problem. At the early stage of the COVID-19 outbreak, a new class of epidemic models was proposed in Liu et al. (Liu et al., 2020a) to take into account the contamination of susceptible individuals by contact with unreported infectious. Actually, this class of models was presented earlier in Arino et al. (Arino et al., 2006). In (Liu et al., 2020a) a new method to use the number of reported cases in SIR models was also proposed. This method and model was extended in several directions by the same group in (Liu et al., 2020b, 2020c, 2020d) to include non-constant transmission rates and a period of exposure. More recently the method was extended and successfully applied to a Japanese age-structured dataset in (Griette et al., 2020). The method was also extended to investigate the predictability of the outbreak in several countries including China, South Korea, Italy, France, Germany and the United Kingdom in (Liu et al., 2021). The application of the Bayesian method was also considered in (Cotta et al., 2020).

In parallel with these modeling ideas, Bayesian methods have been widely used to identify the parameters in the models used for the COVID-19 pandemic (see e.g. Roques et al. (Roques, Klein, Papaix, Sar, & Soubeyrand, 2020a, 2020b; Roques et al., 2020a, 2020b) where an estimate of the fatality ratio has been developed). A remarkable feature of those methods is to provide mechanisms to correct some of the known biases in the observation of cases, such as the daily number of tests. Here we embed the data for the daily number of tests into an epidemic model and compare the number of reported cases produced by the model and the data. Our goal is to understand the relationship between the data for the daily number of tests (which is an input of our model) and the data for the daily number of reported cases (which is an output of our model).

The plan of the paper is the following. In Section 2, we present a model involving the daily number of tests. In Section 3, we apply the method presented in (Liu et al., 2020a) to our new model. In Section 4, we present some numerical simulations and compare the model with the data. The last section is devoted to the discussion.

2. Epidemic with testing data

Let $n(t)$ be the number of tests per unit of time. Throughout this paper, we use one day as the unit of time. Therefore $n(t)$ can be regarded as the daily number of tests at time t. The function $n(t)$ is actually coming from a database for the New York State (https://covidtracking.com). Let $N(t)$ be the cumulative number of tests from the beginning of the epidemic. Then

$$N'(t) = n(t), \text{ for } t \geq t_1 \text{ and } N(t_1) = N_1. \tag{2.1}$$

Remark 2.1. Section 4 is devoted to numerical simulations. We use n(t) as a piecewise constant function that varies day by day. Each day, n(t) is equal to the number of tests that were performed that day. So n(t) should be understood as the black curve in Fig. 4.

The model consists of the following ordinary differential equation

$$\begin{cases} S'(t) = -\tau S(t)[I(t) + U(t) + D(t)], \\ E'(t) = \tau S(t)[I(t) + U(t)] + D(t)] - \alpha E(t), \\ I'(t) = \alpha E(t) - \nu I(t), \\ U'(t) = \nu (1-f) I(t) + n(t) (1-\sigma) g D(t) - \eta U(t), \\ D'(t) = \nu f I(t) - n(t) g D(t) - \eta D(t), \\ R'(t) = n(t) \sigma g D(t) - \eta R(t). \end{cases} \tag{2.2}$$

This system is supplemented by initial data (which are all non negative)

$$S(t_1) = S_1, E(t_1) = E_1, I(t_1) = I_1, U(t_1) = U_1, D(t_1) = D_1 \text{ and } R(t_1) = R_1. \tag{2.3}$$

The time t_1 corresponds to the time where the tests started to be used constantly. Therefore the epidemic started before t_1.

Here $t \geq t_1$ is the time in days. $S(t)$ is the number of individuals susceptible to infection. $E(t)$ is the number of exposed individuals (i.e. who are incubating the disease but not infectious). $I(t)$ is the number of individuals incubating the disease, but already infectious. $U(t)$ is the number of undetected infectious individuals (i.e. who are expressing mild or no symptoms), and the infectious that have been tested with a false negative result, are therefore not candidates for testing. $D(t)$ is the number of individuals who express severe symptoms and are candidates for testing. $R(t)$ is the number of individuals who have been tested positive to the disease. The flux diagram of our model is presented in Fig. 1.

Susceptible individuals $S(t)$ become infected by contact with an infectious individual $I(t), U(t)D(t)$. When they get infected, susceptibles are first classified as exposed individuals $E(t)$, that is to say that they are incubating the disease but not yet infectious. The average length of this exposed period (or noninfectious incubation period) is $1/\alpha$ days.

After the exposure period, individuals are becoming asymptomatic infectious $I(t)$. The average length of the asymptomatic infectious period is $1/\nu$ days. After this period, individuals are becoming either mildly symptomatic individuals U(t) or individuals with severe symptoms $D(t)$. The average length of this infectious period is $1/\eta$ days. Some of the U-individuals may show no symptoms at all.
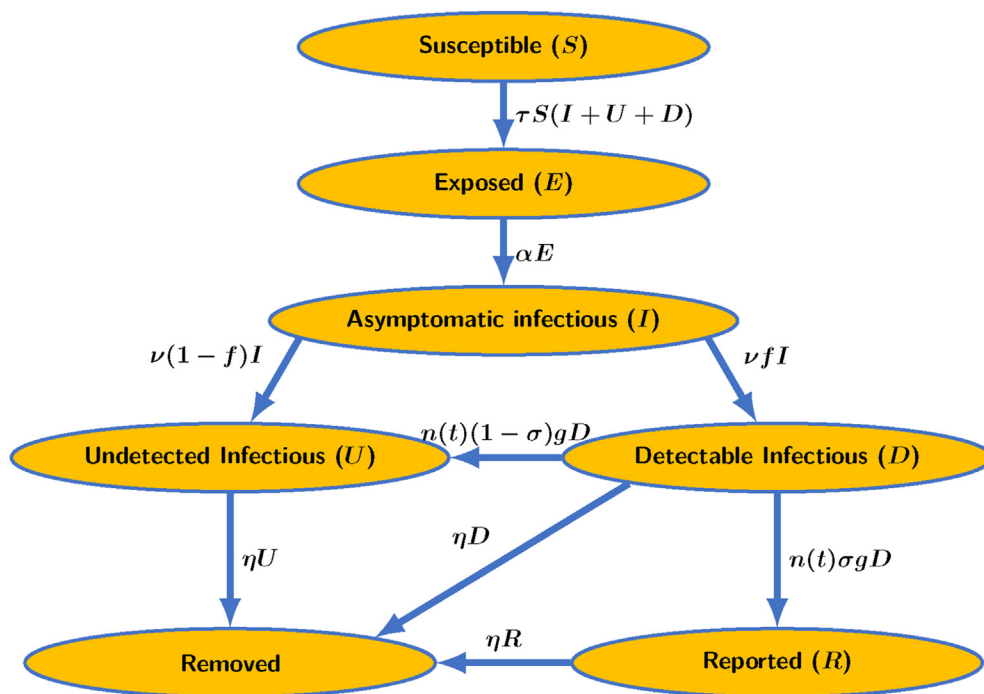
Fig. 1. Flow chart of the epidemic model with tests (2.2). In this diagram $n(t)$ is the daily number of tests at time t. We consider a fraction $(1 - \sigma)$ of false negative tests and a fraction σ of true positive tests. The parameter g reflects the fact that the tests are devoted not only to the symptomatic patients but also to a large fraction of the population of New York state.

In our model, the transmission can occur between a S-individual and an I-,U- or R-individual. Transmissions of SARS-CoV-2 are described in the model by the term $\tau S(t)[I(t) + U(t) + D(t)]$ where τ is the transmission rate. Here, even though a transmission from R-individuals to a S-individuals is possible in theory (e.g. if a tested patient infects its medical doctor), we consider that such a case is rare and we neglect it.

The last part of the model is devoted to the testing. The parameter σ is the fraction of true positive tests and $(1 - \sigma)$ is the fraction of false negative tests. The quantity σ has been estimated at $\sigma = 0.7$ in the case of nasal or pharyngeal swabs for SARS-CoV-2 (Wang et al., 2020).

Among the detectable infectious, we assume that only a fraction g are tested per unit of time. This fraction corresponds to individuals with symptoms suggesting a potential infection to SARS-CoV-2. The fraction g is the frequency of testable individuals in the population of New York state. We can rewrite g as

$$g = \frac{1}{\kappa P} \tag{2.4}$$

where P is the total number of individuals in the population of the state of New York and $0 \leq \kappa \leq 1$ is the fraction total population with mild or sever symptoms that may induce a test.

Individuals who were tested positive $R(t)$ are infectious on average during a period of $1/\eta$ days. But we assume that they become immediately isolated and do not contribute to the epidemic anymore. In this model we focus on the testing of the D-individuals. The quantity $n(t)\,\sigma\,g\,D$ is a flux of successfully tested D-individuals which become R-individuals. The flux of tested D-individuals which are false negatives is $n(t)\,(1 - \sigma)\,g\,D$ which go from the class of D-individuals to the U-individuals. The parameters of the model and the initial conditions of the model are listed in Table 1.

Before describing our method we list a few variables and parameters in Table 2. The cumulative number of reported cases is obtained by using the following equation

$$CR'(t) = n(t)\,\sigma\,g\,D(t). \tag{2.5}$$

The daily number of reported cases $DR'(t)$ is given by

$$DR(t)' = n(t)\,\sigma\,g\,D(t) - DR(t). \tag{2.6}$$

The cumulative number of detectable cases is given by

Table 1
Parameters and initial conditions of the model.

Symbol	Interpretation	Method
t_1	Date when the tests start to be used extensively	fixed
S_1	Number of susceptible at time t_1	fixed
E_1	Number of exposed at time t_1	fitted
I_1	Number of asymptomatic infectious at time t_1	fitted
U_1	Number of undetectable infectious at time t_1	fitted
D_1	Number of detectable infectious at time t_1	fitted
R_1	Number of reported (tested positive) cases at time t_1	fitted
T	Transmission rate	fitted
$n(t)$	Number of tests per unit of time	fixed
$1/\alpha$	Average length of exposure	fixed
$1/\nu$	Average length of asymptomatic infectiousness	fixed
$1/\eta$	Average length of symptomatic infectiousness	fixed
F	Frequency of infectious with sever symptoms	fixed
Σ	Fraction of true positive tests	fixed
G	Frequency of testable individuals	fixed

Table 2
Variables used in the model.

Symbol	Interpretation	Equation
T	Time (in days)	
$S(t)$	Number of susceptible at time t	(2.2)
$E(t)$	Number of exposed at time t	(2.2)
$I(t)$	Number of asymptomatic infectious at time t	(2.2)
$U(t)$	Number of undetectable infectious at time t	(2.2)
$D(t)$	Number of detectable infectious at time t	(2.2)
$R(t)$	Number of reported (tested infectious) cases at time t	(2.2)
$CR(t)$	Cumulative number of reported (tested infectious) cases at time t	(2.5)
$DR(t)$	Daily number of reported (tested infectious) cases at time t	(2.6)
$CD(t)$	Cumulative number of detectable infectious at time t	(2.7)
$CU(t)$	Cumulative number of undetectable infectious at time t	(2.8)

$$CD'(t) = \nu f I(t), \tag{2.7}$$

and the cumulative number of undetectable cases is given by

$$CU'(t) = \nu(1-f)I(t) + n(t)(1-\sigma)gD(t). \tag{2.8}$$

3. Method to fit the cumulative number of reported cases

In order to deal with data, we need to understand how to set the parameters as well as some components of the initial conditions (see Fig. 2). In order to do so, we extend the method presented first in (Liu et al., 2020a). The main novelty here concerns the cumulative number of tests which is assumed to grow linearly at the beginning. This property is satisfied for the New York State data as we can see in Fig. 3. The black curve in this figure is close to a line from March 15 to April 15. Fig. 4
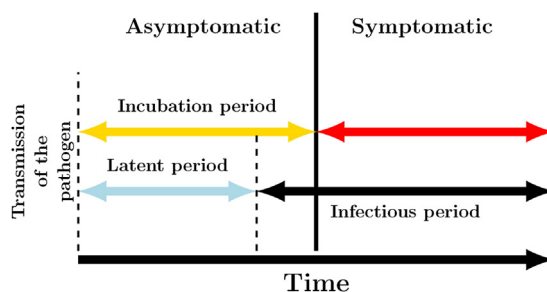

Fig. 2. Key time periods of COVID-19 infection: the latent or exposed period before the onset of symptoms and transmissibility, the incubation period before symptoms appear, the symptomatic period, and the transmissibility period, which may overlap the asymptomatic period.
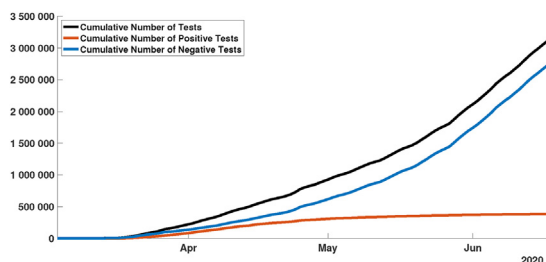
Fig. 3. In this figure, we plot the cumulative number of tests for the New York State. The black curve, orange curve, and blue curve correspond respectively to the number of tests, the number of positive tests, and the number of negative tests. We can see that at the early beginning of the epidemic, the cumulative number of tests (black curve) grows linearly from mid-March to mid-April.

shows day-by-day fluctuations of the number of tests while in Fig. 3 the day-by-day fluctuations are not visible and the cumulative data allow to understand the growth tendency of the number of tests.

Phenomenological models for the tests: We fit a line to the cumulative number of tests in a suitable interval of days $[t_1, t_2]$. This means that we can find a pair of numbers a and b such that

$$N(t) = a \times (t - t_1) + N_1, \text{ for } t_1 \leq t \leq t_2. \tag{3.1}$$

where a the daily number of tests and N_1 is the cumulative number of tests on day t_1.

By using the fact that $N(t)' = n(t)$ we deduce that

$$n(t) = a, \text{ for } t_1 \leq t \leq t_2. \tag{3.2}$$

Remark 3.1. In the simulations we fit a line to the cumulative number of tests from mid-March to mid-April. Fig. 3 shows that the linear growth assumption is reasonable for the New York State cumulative testing data.

Phenomenological models for the reported cases: At the early stage of the epidemic, we assume that all the infected components of the system grow exponentially while the number of susceptible remains unchanged during a relatively short period of time $t \not\ni [t_1, t_2]$. Therefore, we assume that

$$E(t) = E_1 e^{\chi_2(t-t_1)}, I(t) = I_1 e^{\chi_2(t-t_1)}, D(t) = D_1 e^{\chi_2(t-t_1)} \text{ and } U(t) = U_1 e^{\chi_2(t-t_1)}. \tag{3.3}$$

We deduce that the cumulative number of reported cases satisfies

$$CR(t) = CR(t_1) + \int_{t_1}^{t} a\sigma g D(\theta) d\theta \tag{3.4}$$

hence by replacing $D(t)$ by the exponential formula (3.3)

$$CR(t) = CR(t_1) + \frac{a\sigma g}{\chi_2} D_1 \left(e^{\chi_2(t-t_1)} - 1 \right) \tag{3.5}$$

and it makes sense to assume that $CR(t) - CR(t_1)$ has the following form

$$CR(t) - CR(t_1) = \chi_1 e^{\chi_2(t-t_1)} - \chi_3. \tag{3.6}$$

By identifying (3.5) and (3.6) we deduce that

$$\chi_1 = \chi_3 = \frac{a \sigma g}{\chi_2} D_1. \tag{3.7}$$

Moreover by using (3.2) and the fact that the number of susceptible $S(t)$ remains constant equalling S_1 on the time interval $t \not\ni [t_1, t_2]$, the E-equation, I-equation, U-equation and D-equation of the model (2.2) become
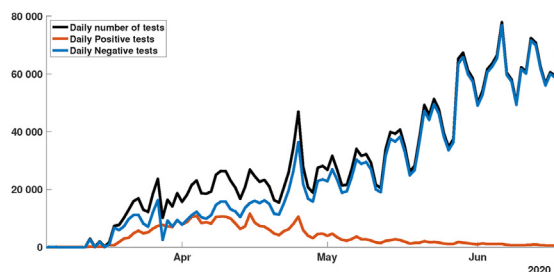
Fig. 4. In this figure, we plot the daily number of tests for the New York State. The black curve, orange curve, and blue curve correspond respectively to the number of tests, the number of positive tests, and the number of negative tests.

$$
\begin{cases}
E'(t) = \tau S_1 [I(t) + U(t) + D(t)] - \alpha E(t), \\
I'(t) = \alpha E(t) - \nu I(t), \\
U'(t) = \nu (1 - f) I(t) + a (1 - \sigma) g D(t) - \eta U(t), \\
D'(t) = \nu f I(t) - a g D(t) - \eta D(t).
\end{cases}
$$

By using (3.3) we obtain

$$
\begin{cases}
\chi_2 E_1 = \tau S_1 [I_1 + U_1 + D_1] - \alpha E_1, \\
\chi_2 I_1 = \alpha E_1 - \nu I_1, \\
\chi_2 U_1 = \nu (1 - f) I_1 + a (1 - \sigma) g D_1 - \eta U_1, \\
\chi_2 D_1 = \nu f I_1 - a g D_1 - \eta D_1.
\end{cases}
$$

Computing further, we get

$$
\begin{cases}
E_1 = \dfrac{\tau_1 S_1 (I_1 + U_1 + D_1)}{\chi_2 + \alpha} \\[2mm]
I_1 = \dfrac{\alpha E_1}{\chi_2 + \nu} \\[2mm]
U_1 = \dfrac{\nu I_1 + a (1 - \sigma) g D_1}{\chi_2 + \eta} \\[2mm]
D_1 = \dfrac{\nu f I_1}{\chi_2 + a g + \eta}.
\end{cases}
\tag{3.8}
$$

Finally by using (3.7)

$$
D_1 = \frac{\chi_2 \chi_3}{\sigma a g}.
\tag{3.9}
$$

and by using (3.8) we obtain

$$
\begin{cases}
I_1 \quad = \dfrac{\chi_2 + a g + \eta}{\nu f} D_1 = \dfrac{\chi_2 + a g + \eta}{\nu} \times \dfrac{\chi_2 \chi_3}{f \sigma a g} \\[3mm]
U_1 \quad = \dfrac{\nu I_1 + (1 - \sigma) a g D_1}{\chi_2 + \eta} = \dfrac{(\chi_2 + \eta + [1 + f(1 - \sigma)] a g)}{\chi_2 + \eta} \times \dfrac{\chi_2 \chi_3}{f \sigma a g} \\[3mm]
E_1 \quad = \dfrac{(\chi_2 + \nu)}{\alpha} I_1 = \dfrac{(\chi_2 + \nu)}{\alpha} \times \dfrac{(\chi_2 + a g + \eta)}{\nu} \times \dfrac{\chi_2 \chi_3}{f \sigma a g} \\[3mm]
\tau_1 \quad = \dfrac{(\chi_2 + \alpha)}{S_1 (I_1 + U_1 + D_1)} E_1 \\[3mm]
\quad = \dfrac{(\chi_2 + \alpha)(\chi_2 + \nu)(\chi_2 + \eta)(\chi_2 + a g + \eta)}{\alpha S_1 ([\chi_2 + a g + \eta + \nu(f + 1)](\chi_2 + \eta) + \nu[1 + f(1 - \sigma)] a g)},
\end{cases}
\tag{3.10}
$$

where I_1 is the number of incubating infectious individuals at time t_1, U_1 is the number of unreported infectious individuals at time t_1, E_1 is the number of incubating non-infectious individuals at time t_1 (see (3.3)), and finally τ_1 is the transmission rate at time t_1.

4. Numerical simulations

We assume that the transmission coefficient takes the form

$$\tau(t) = \tau_0((1 - \gamma)\exp(-\mu(t - T_m)_+) + \gamma), \tag{4.1}$$

where $\tau_0 > 0$ is the initial transmission coefficient, $T_m > 0$ is the time at which the social distancing starts in the population, and $\mu > 0$ controls the speed at which this social distancing is taking place.

To take into account the effect of social distancing and public measures, we assume that the transmission coefficient $\tau(t)$ can be modulated by γ. Indeed by closing schools and non-essential shops and by imposing social distancing in New York State, the number of contacts per day is reduced. This effect was visible on the news during the first wave of the COVID-19 epidemic in New York city since the streets were almost empty at some point. The parameter $\gamma > 0$ is the percentage of the number of transmissions that remain after a transition period (depending on μ), compared to a normal situation. A similar non-constant transmission rate was considered by Chowell et al. (Chowell et al., 2004).

In Fig. 5 we consider a constant transmission rate $\tau(t) \equiv \tau_0$ which corresponds to $\gamma = 1$ in (4.1). In order to evaluate the distance between the model and the data, we compare the distance between the cumulative number of cases CR produced by the model and the data (see the orange dots and orange curve in Fig. 5-(a)). In Fig. 5-(c) we observe that the cumulative number of cases increases up more than 14 millions of people, which indeed is not realistic. Nevertheless by choosing the parameter $g = 3.08 \times 10^{-7} = 1/\left(\frac{S_0}{6}\right)$ in Fig. 5-(d) we can see that the orange dots and the blue curve match very well. In the rest of this section, we focus on the model with confinement (or social distancing) measures. We assume that such social distancing measures have a strong impact on the transmission rate by assuming that $\gamma = 0.2 < 1$. It means that only 20% of the transmissions remain after a transition period.
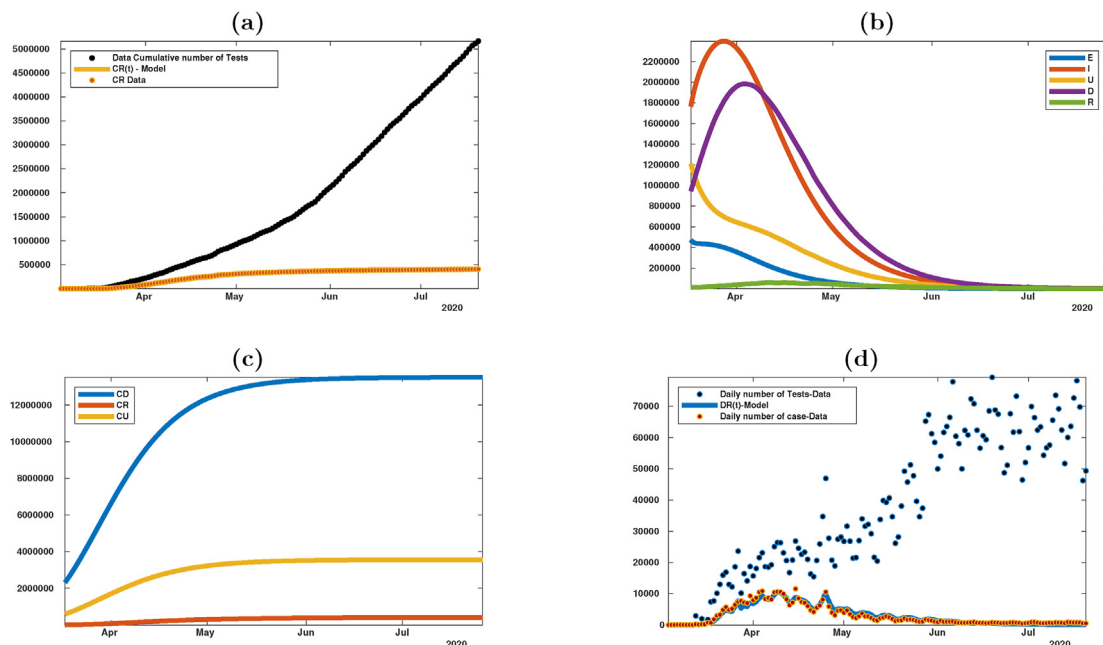

Fig. 5. Best fit of the model without confinement (or social distancing) measures (i.e. $\gamma = 1$). **Fitted parameters:** The transmission rate $\tau(t) \equiv \tau_0$ is constant according to formula (4.1) with $\gamma = 1$ and τ_0 is fixed to the value τ_1 computed by using (3.10). **Parameter values:** $S_0 = 19453561$, $\alpha = 1$, $\nu = 1/6$, $\eta = 1/7$, $\sigma = 0.7$, $f = 0.8$ and $g = 6/S_0 = 3.08 \times 10^{-7}$. $t_1 = $ *march* 18, $t_2 = $ march 29, $a = 1.4874 \times 10^4$, $b = -2.1781 \times 10^5$, $\chi_1 = 2.8814 \times 10^4$, $\chi_2 = 0.1013$, $\chi_3 = 2.9969 \times 10^4$. In figure **(a)** we plot the cumulative number of tests (black dots), the cumulative number of positive cases (red dots) for the state of New York and the cumulative number of cases CD(t) (yellow curve) obtained by fitting the model to the data. In figures **(b)**–**(c)** we plot the number of cases obtained from the model. We observe that most of the cases are unreported. In figure **(d)** we plot the daily number of tests (black dots), the daily number of positive cases (red dots) for the state of New York and the daily number of cases DD(t) obtained from the data.
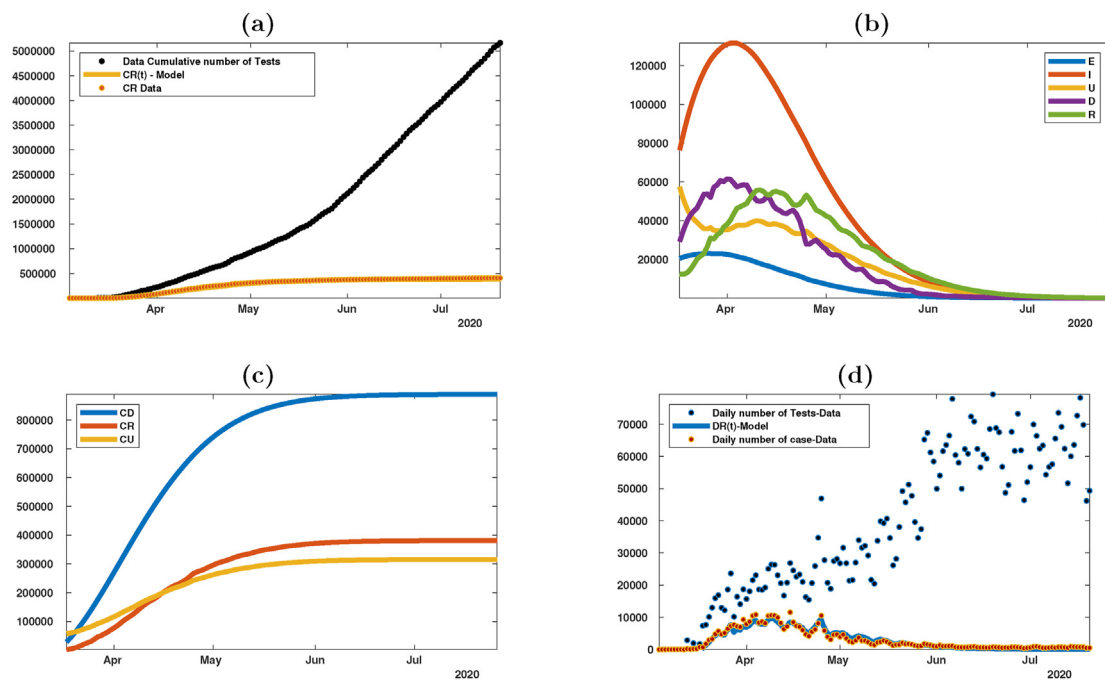
Fig. 6. Best fit of the model with confinement (or social distancing) measures. **Parameter values:** Same as in Fig. 5, except the transmission coefficient which is not constant in time with $\gamma = 0.2$, $T_m = 15$ Mar (starting day of public measures), $\mu = 0.0251$, $g = 10^{-5}$ and τ_0 is fixed at the value τ_1 computed by using (3.10). In figure **(a)** we plot the cumulative number of tests (black dots), the cumulative number of positive cases (red dots) for the state of New York and the cumulative number of cases CD(t) (yellow curve) obtained by fitting the model to the data. In figures **(b)–(c)** we plot the corresponding number of cases obtained from the model. With this set of parameters we observe that most of the cases are unreported. In figure **(d)** we plot the daily number of tests (black dots), the daily number of positive cases (red dots) for the state of New York and the daily number of cases DD(t) obtained from the data.
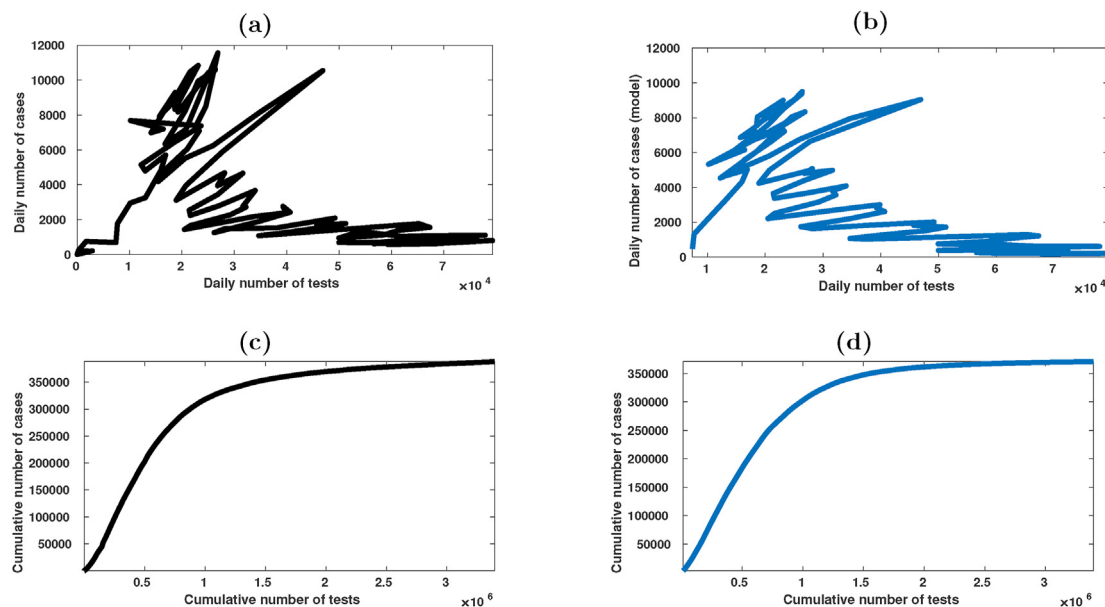

Fig. 7. In this figure we plot the curves of the number of reported cases as a function of the number of tests parametrized by the time. The top figures (a) and (b) correspond to the daily number of cases and the bottom figures (c) and (d) correspond to the cumulative number of cases. On the left-hand side we plot the data (a) and (c) while on the right-hand side we plot the model (b) and (d). **Parameter values:** Same as in Fig. 6. In figure **(a)** we plot the daily number of cases coming from the data as a function of the daily number of tests. In figure **(b)** we plot the daily number of cases given by the model as a function of the daily number of tests coming from the data. In figure **(c)** we plot the cumulative number of cases coming from the data as a function of the cumulative number of tests. In figure **(d)** we plot the cumulative number of cases coming from the model as a function of the cumulative number of tests from the data.

In Fig. 6-(c) we can observe that the cumulative number of cases increases up to 800 000 (blue curve) while the cumulative number reported cases goes up to 350 000. In Fig. 6-(d) we can see that the orange dots and the blue curve match very well again. In order to get this fit we fix the parameter $g = 10^{-5}$.

Fig. 7 (a) and (b), we aim at understanding the connection between the daily fluctuations of the number of reported cases (epidemic dynamic) and the daily number of tests (testing dynamics). The combination of the testing dynamics and the infection dynamics gives indeed a very complex curve parametrized by the time. It seems that the only reasonable comparison that we can make is between the cumulative number of reported cases and the cumulative number of tests. In Fig. 7 (c) and (d), the comparison of the model and the data gives a very decent fit. In Fig. 7, all the curves are time dependent parametrized curves. The abscissa is the number of tests (horizontal axis) and the ordinate is the number of reported cases (vertical axis). It corresponds (with our notations) to the parametric functions $t \rightarrow (n_{data}(t), DR(t))$ in figures (a) and (b) and their cumulative equivalent $t \rightarrow (N_{data}(t), CR(t))$ in figures (c) and (d). In figures (a) and (c) we use only the data, that is to say that we plot $t \rightarrow (n_{data}(t), DR_{data}(t))$ and $t \rightarrow (N_{data}(t), CR_{data}(t))$. In figures (b) and (d) we use only the model for the number of reported cases, that is to say that we plot $t \rightarrow (n_{data}(t), DR_{model}(t))$ and $t \rightarrow (N_{data}(t), CR_{model}(t))$.

In Fig. 8, our goal is to investigate the effect of a change in the testing policy in the New York State. We are particularly interested in estimating the effect of an increase of the number of tests on the epidemic. Indeed increasing the number of tests may be thought as beneficial to reduce the number of cases. Here we challenge this idea by comparing an increase in the number of tests to the quantitative output of our model. In Fig. 8, we replace the daily number of tests $n_{data}(t)$ (coming from the data for New York's state) in the model by either $2 \times n_{data}(t)$, $5 \times n_{data}(t)$, $10 \times n_{data}(t)$ or $100 \times n_{data}(t)$.

As expected, an increase of the number of tests is helping to reduce the number of cases at first. However, after increasing 10 times the number of tests, there is no significant difference (in the number of reported) between 10 times and 100 times more tests. Therefore there must be an optimum between increasing the number of tests (which costs money and other limited resources) and being efficient to slow down the epidemic.

5. Discussion

In this article, we proposed a new epidemic model involving the daily number of tests as an input of the model. The model itself extends our previous models presented in (Griette et al., 2020; Liu et al., 2020a, 2020b, 2020c, 2020d, 2021). We proposed a new method to use the data in such a context based on the fact that the cumulative number of tests grows linearly at the early stage of the epidemic. Fig. 3 shows that this is a reasonable assumption for the New York State data from mid-March to mid-April.
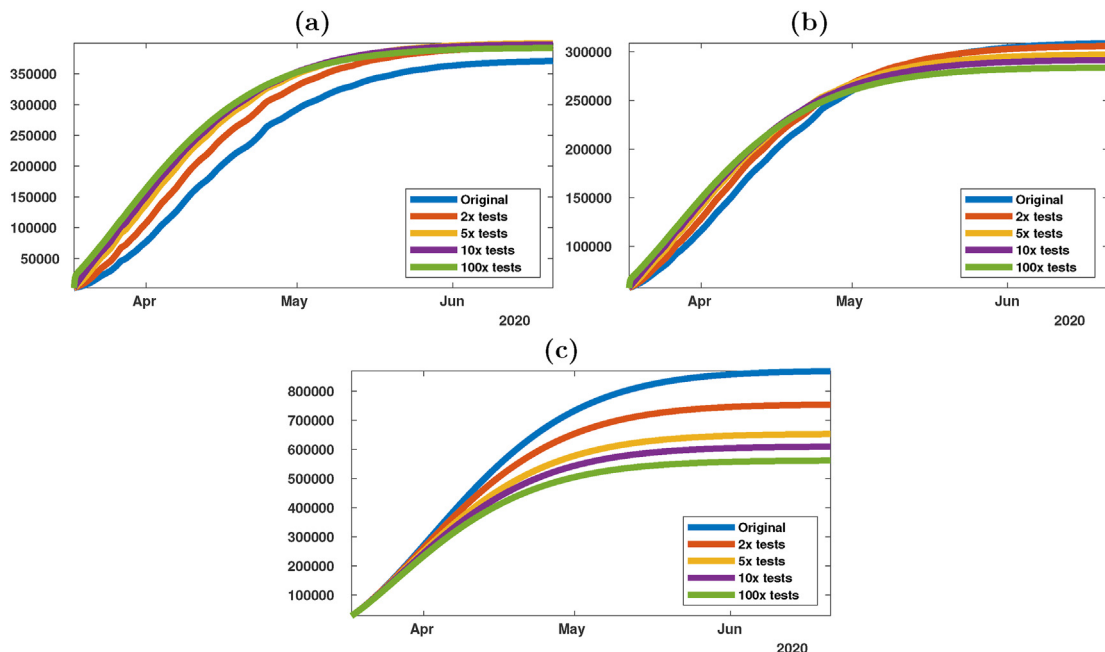

Fig. 8. Cumulative number of cases for different testing strategies: Original (blue curve), doubled (red curve), multiplied by 5 (yellow curve), multiplied by 10 (purple line) and multiplied by 100 (green curve). The transmission coefficient depends on the time, according to formula (4.1) with $\gamma = 0.2$, and τ_0 is fitted by using (3.10). **Parameter values:** they are the same as in Fig. 6. In figure **(a)** we plot the cumulated number of cases CR(t) as a function of time. In figure **(b)** we plot the cumulative number of undetectable cases CU(t) as a function of time. In figure **(c)** we plot the cumulative number of cases (including covert cases) CD(t) as a function of time. Note that the total number of cases (including covert cases) is reduced by 35% when the number of tests is multiplied by 100.

Our numerical simulations show a very good concordance between the number of reported cases produced by the model and the data in two very different situations. Indeed, Figs. 5 and 6 correspond respectively to an epidemic without and with public intervention to limit the number of transmissions. This is an important observation since this shows that testing data and reported cases are not sufficient to evaluate the real amplitude of the epidemic. To solve this problem, the only solution seems to include a different kind of data to the models. This could be done by studying statistically representative samples in the population. Otherwise, biases can always be suspected. Such a question is of particular interest in order to evaluate the fraction of the population that has been infected by the virus and their possible immunity.

In Fig. 7, we compared the testing dynamic (day to day variation in the number of tests) and the reported cases dynamic (day to day variation in the number of reported). Indeed, the dynamics of daily cases is extremely complex, but we also obtain a relatively robust curve for the cumulative numbers. Our model gives a good fit for this cumulative cases.

In Fig. 8, we compared multiple testing strategies. By increasing 2, 5, 10 and 100 times the number of tests, we can project the efficiency of an increase in the daily number of tests. We observe that it is efficient to increase this number up to 10 but the relative gain in absolute number of infected individuals rapidly drops after that. In particular, our projections do not show a big difference between a 10-times increase in the number of tests and a 100-times increase. Therefore there is a balance to find between the number of test and the efficiency in the evaluation of the number of cases, the optimal strategy being dependent on other factors like the monetary cost of the tests.

Declaration of competing interest

None declared.

References

Anderson, R. M., & May, R. M. (1991). *Infective diseases of humans: Dynamics and control.* Oxford: Oxford University Press.

Arino, J., Brauer, F., van den Driessche, P., Watmough, J., & Wu, J. (2006). Simple models for containment of a pandemic. *Journal of The Royal Society Interface, 3*(8), 453–457.

Bailey, N. T. J. (1957). *The mathematical theory of epidemics.* London: Charles Griffin.

Brauer, F., & Castillo-Chavez, C. (2000). *Mathematical models in population biology and epidemiology.* New York: Springer.

Brauer, F., Castillo-Chavez, C., & Feng, Z. (2019). *Mathematical models in epidemiology.* New York: Springer.

Brauer, F., van den Driessche, P., & Wu, J. (Eds.). (2008). *Mathematical epidemiology.* Berlin: Springer.

Busenberg, S., & Cooke, K. (1993). Vertically transmitted diseases: Models and dynamics. In *Lecture notes in biomathematics* (Vol. 23)Berlin: Springer-Verlag.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., & Hyman, J. M. (2004). 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), 119–126.

Cotta, R. M., Naveira-Cotta, C. P., & Magal, P. (2020). Modelling the COVID-19 epidemics in Brasil: Parametric identification and public health measures influence. *Biology, 9*(8), 220, 2020.

https://covidtracking.com/data/state/new-york#historical.

Diekmann, O., Heesterbeek, H., & Britton, T. (2013). *Mathematical tools for understanding infectious disease dynamics.* Princeton: Princeton University Press.

Griette, Q., Magal, P., & Seydi, O. (2020). Unreported cases for age dependent COVID-19 outbreak in Japan. *Biology, 9*, 132.

Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review, 42*(4), 599–653.

Keeling, M., & Rohani, P. (2007). *Modeling infectious diseases in humans and animals.* Princeton: Princeton University Press.

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

Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. *Mathematical Biosciences and Engineering, 17*(4), 3040–3051.

Liu, Z., Magal, P., Seydi, O., & Webb, G. (2020). A COVID-19 epidemic model with latency period. *Infectious Disease Modelling, 5*, 323–337.

Liu, Z., Magal, P., Seydi, O., & Webb, G. (May 01 2020). *A model to predict COVID-19 epidemics with applications to South Korea, Italy, and Spain.* SIAM News.

Liu, Z., Magal, P., & Webb, G. (2021). Predicting the number of reported and unreported cases for the COVID-19 epidemics in China, South Korea, Italy, France, Germany and United Kingdom. *Journal of Theoretical Biology, 509*(21), Article 110501. https://doi.org/10.1016/j.jtbi.2020.110501

Mizumoto, K., Kagaya, K., Zarebski, A., & Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveillance, 25*(10), 2000180.

Murray, J. D. (1993). *Mathematical biology.* Berlin: Springer.

Qiu, J. (2020). Covert coronavirus infections could be seeding new outbreaks. *Nature.* https://www.nature.com/articles/d41586-020-00822-x.

Roques, L., Klein, E. K., Papaix, J., Sar, A., & Soubeyrand, S. (2020). Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology, 9*, 97.

Roques, L., Klein, E. K., Papaix, J., Sar, A., & Soubeyrand, S. (2020). Effect of a one-month lockdown on the epidemic dynamics of COVID-19 in France. *Frontiers of Medicine, 7*, 274.

Rothe, C., et al. (2020). Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine, 382*(10), 970–971.

Tang, B., Wang, X., Li, Q., Bragazzi, N. L., Tang, S., Xiao, Y., & Wu, J. (2020). Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine, 9*, 462.

Thieme, H. R. (2003). *Mathematics in population biology.* Princeton: Princeton University Press.

Wang, C., et al. (2020). *Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China.* https://doi.org/10.1101/2020.03.03.20030593. Preprint at medRxiv.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and inter-national spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. *The Lancet, 395*, 689–697.