Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany

P. Magal$^{(a)}$ and G. Webb$^{(b)}$

$^{(a)}$ *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.*
CNRS, IMB, UMR 5251, F-33400 Talence, France.
$^{(b)}$ *Mathematics Department, Vanderbilt University, Nashville, TN, USA*

March 20, 2020

Abstract

We model the COVID-19 coronavirus epidemic in South Korea, Italy, France, and Germany. We use early reported case data to predict the cumulative number of reported cases to a final size. The key features of our model are the timing of implementation of major public policies restricting social movement, the identification and isolation of unreported cases, and the impact of asymptomatic infectious cases.

Keywords: corona virus, reported and unreported cases, isolation, quarantine, public closings; epidemic mathematical model

1 Introduction

In a previous work [2], we developed a model of the COVID-19 epidemic in China. The purpose of the model was to predict forward in time the future number of cases in a time-line of the epidemic from early reported case data. The model in [2] focused on the Chinese government imposed public policies designed to contain the epidemic. Our goal here is to apply this analysis to the COVID-19 epidemics in South Korea, Italy, France, and Germany. In an early phase of the epidemic, the reported case data grows exponentially, which corresponds to a constant transmission rate. We assume that government measures and public awareness cause this early constant transmission rate to change to a time dependent exponentially decreasing rate. We identify the early constant transmission rate using a method developed in [1]. We then identify the time dependent exponentially decreasing transmission rate from reported case data, and project forward the time-line of the epidemic course. We refer to [3, 4] for more results about this topic.

Our model incorporates the following essential ingredients of COVID-19 epidemics: (1) the number of asymptomatic infectious individuals (with very mild or no symptoms), (2) the number of symptomatic reported infectious individuals (with severe symptoms) and (3) the number of symptomatic unreported infectious individuals (with mild symptoms).

COVID-19 epidemics can be decomposed into three phases:

Phase I: *The first phase of the epidemic corresponds to a linear growth in the number of reported cases, where the number of daily reported case is almost constant day after day;*

Phase II: *The second phase of the epidemic corresponds to an exponential increasing phase, where the number of cases grows exponentially, corresponding to a constant transmission rate.*

Phase III: *The third phase of the epidemic corresponds to a time dependent exponentially decreasing transmission rate, due to major public interventions and social distancing measures.*

Our analysis identifies the epidemics in South Korea, Italy, France, and Germany as Phase III. Our model is applicable to COVID-19 epidemics in any region with reported case data, which can be updated to higher accuracy with on-going day by day reported case data.

2 Data

We use data from the Korean Center for Disease Control, the Italian Ministry of Health, the French Public Agency of Health, and the Robert Koch Institute of Germany. The cumulative reported cases are given in Figure 1 and the daily reported cases are given in Figure 2.
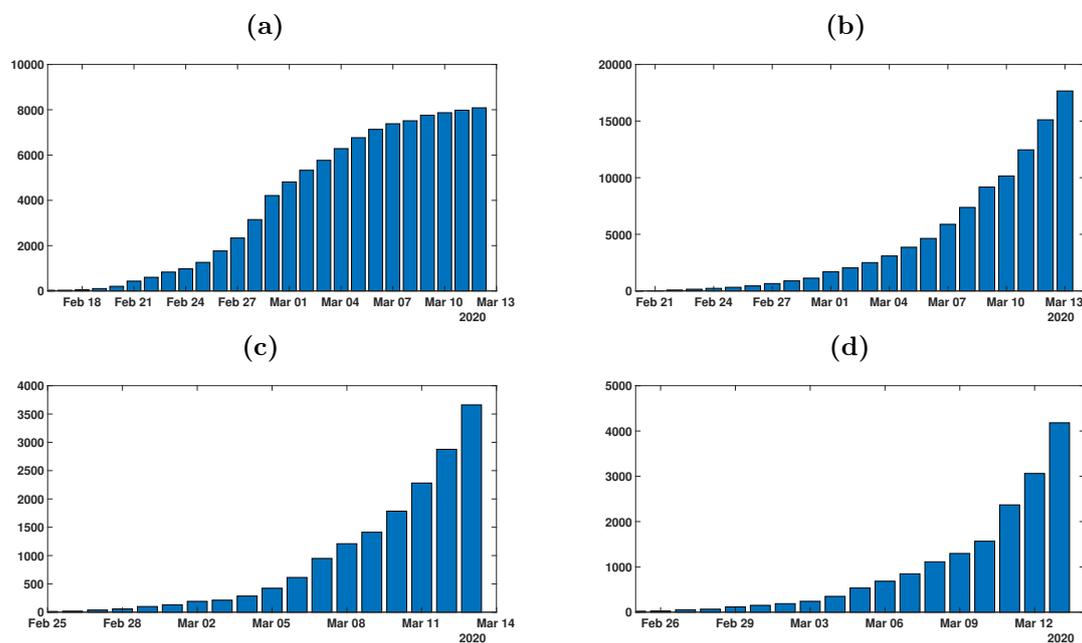

Figure 1: *Cumulative number of reported cases for COVID-19 for (a) South Korea between January 20 and March 9; (b) Italy between January 31 and March 8; (c) France between February 25 and March 9; (d) Germany between February 24 and March 9.*
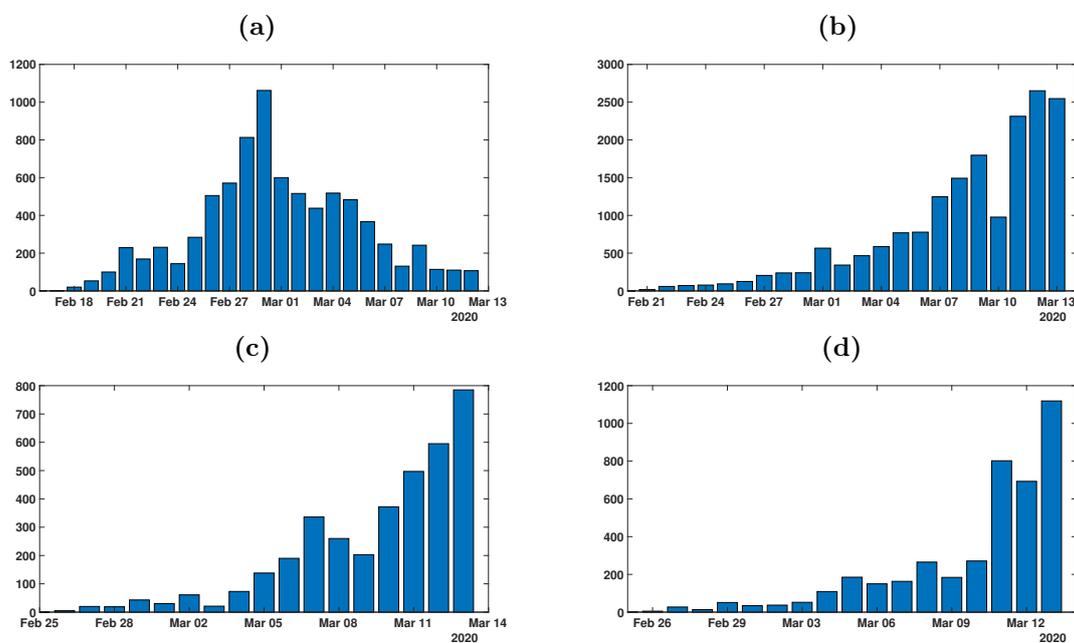

Figure 2: *Daily number of reported cases for COVID-19 for (a) South Korea between January 20 and March 9; (b) Italy between January 31 and March 8; (c) France between February 25 and March 9; (d) Germany between February 24 and March 9.*

2

3 Model

The model consists of the following system of ordinary differential equations:

$$
\begin{cases}
S'(t) = -\tau(t)S(t)[I(t) + U(t)], \\
I'(t) = \tau(t)S(t)[I(t) + U(t)] - \nu I(t), \\
R'(t) = \nu_1 I(t) - \eta R(t), \\
U'(t) = \nu_2 I(t) - \eta U(t).
\end{cases}
\tag{3.1}
$$

This system is supplemented by initial data

$$
S(t_0) = S_0 > 0,\ I(t_0) = I_0 > 0,\ R(t_0) = 0 \text{ and } U(t_0) = U_0 \geq 0.
\tag{3.2}
$$

Here $t \geq t_0$ is time in days, t_0 is the beginning date of the model of the epidemic, $S(t)$ is the number of individuals susceptible to infection at time t, $I(t)$ is the number of asymptomatic infectious individuals at time t, $R(t)$ is the number of reported symptomatic infectious individuals at time t, and $U(t)$ is the number of unreported symptomatic infectious individuals at time t.

The transmission rate at time t is $\tau(t)$. Asymptomatic infectious individuals $I(t)$ are infectious for an average period of $1/\nu$ days. Reported symptomatic individuals $R(t)$ are infectious for an average period of $1/\eta$ days, as are unreported symptomatic individuals $U(t)$. We assume that reported symptomatic infectious individuals $R(t)$ are reported and isolated immediately, and cause no further infections. The asymptomatic individuals $I(t)$ can also be viewed as having a low-level symptomatic state. All infections are acquired from either $I(t)$ or $U(t)$ individuals. The fraction f of asymptomatic infectious become reported symptomatic infectious, and the fraction $1 - f$ become unreported symptomatic infectious. The rate asymptomatic infectious become reported symptomatic is $\nu_1 = f\nu$, the rate asymptomatic infectious become unreported symptomatic is $\nu_2 = (1 - f)\nu$, where $\nu_1 + \nu_2 = \nu$. The cumulative number of reported cases at time t is given by the formula

$$
CR(t) = \nu_1 \int_{t_0}^{t} I(\sigma)d\sigma, \text{ for } t \geq t_0,
\tag{3.3}
$$

and the cumulative number of unreported at time t is given by the formula

$$
CU(t) = \nu_2 \int_{t_0}^{t} I(\sigma)d\sigma, \text{ for } t \geq t_0.
\tag{3.4}
$$

The parameters and initial conditions of the model are given in Table 1 and a flow diagram of the model is given in Figure 3.

Symbol	Interpretation	Method
t_0	Time at which the epidemic started	fitted
S_0	Number of susceptible at time t_0	fixed
I_0	Number of asymptomatic infectious at time t_0	fitted
U_0	Number of unreported symptomatic infectious at time t_0	fitted
$\tau(t)$	Transmission rate at time t	fitted
$1/\nu$	Average time during which asymptomatic infectious are asymptomatic	fixed
f	Fraction of asymptomatic infectious that become reported symptomatic infectious	fixedt
$\nu_1 = f\nu$	Rate at which asymptomatic infectious become reported symptomatic	fitted
$\nu_2 = (1-f)\nu$	Rate at which asymptomatic infectious become unreported symptomatic	fitted
$1/\eta$	Average time symptomatic infectious have symptoms	fixed

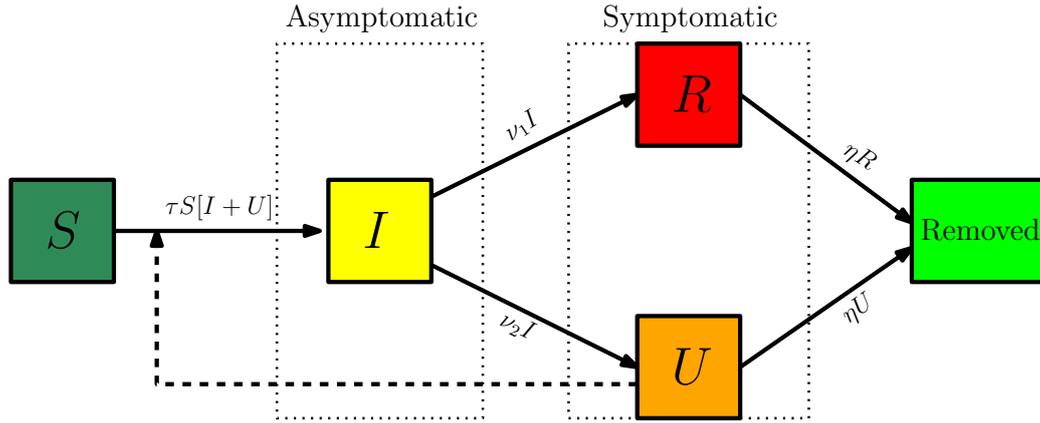Table 1: *Parameters and initial conditions of the model.*

Figure 3: *Compartments and flow chart of the model.*

4 Method to estimate the parameters

We assume $f = 0.6$ or $f = 0.1$, which means that 40% or 90% of symptomatic infectious cases go unreported. The actual value of f is unknown. We assume $\eta = 1/7$, which means that the average period of infectiousness of both unreported symptomatic infectious individuals and reported symptomatic infectious individuals is 7 days. We assume $\nu = 1/7$, which means that the average period of infectiousness of asymptomatic infectious individuals is 7 days. These values can be modified as further epidemiological information becomes known.

We assume that in the second phase of a COVID-19 epidemic, the cumulative number of reported cases $CR(t)$ grows approximately exponentially, according to the formula:

$$CR(t) = \chi_1 \exp(\chi_2 t) - \chi_3, \ t \geq t_0. \tag{4.1}$$

We fix the value $\chi_3 = 1$. The values of χ_1 and χ_2 are fitted to cumulative reported case data in the early phase of the epidemic, when it is recognized that $CR(t)$ is growing exponentially (i.e. we use an exponential fit $\chi_1 \exp(\chi_2 t)$ to fit the data $CR(t) + 1$). We assume the initial value S_0, corresponds to the population of the region of the reported case data. The value of the susceptible population $S(t)$ is assumed to be only slightly changed by removal of the number of people infected in the beginning of the second phase. The other initial conditions are

$$I_0 = \frac{\chi_2}{f(\nu_1 + \nu_2)}, \ U_0 = \left(\frac{(1 - f)(\nu_1 + \nu_2)}{\eta + \chi_2} \right) I_0, \ R_0 = 1.0. \tag{4.2}$$

The value of the transmission rate $\tau(t)$, during the second phase of the epidemic, when the cumulative number of reported cases grows approximately exponential, is the constant value

$$\tau_0 = \left(\frac{\chi_2 + \nu_1 + \nu_2}{S_0} \right) \left(\frac{\eta + \chi_2}{\nu_2 + \eta + \chi_2} \right). \tag{4.3}$$

The initial time for the beginning of the early exponential growth phase is

$$t_0 = \frac{1}{\chi_2} - \log(\chi_1). \tag{4.4}$$

The value of the basic reproductive number is

$$\mathcal{R}_0 = \left(\frac{\tau_0 S_0}{\nu_1 + \nu_2} \right) \left(1 + \frac{\nu_2}{\eta} \right). \tag{4.5}$$

These formulas for I_0, U_0, t_0, τ_0, and $\mathcal{R}_0$ were derived in [1]. Their numerical values, which are technically theoretic, are essential for identification of the exponential growth rate of $CR(t)$ in the second phase.

During the second phase $\tau(t) \equiv \tau_0$ is constant. When strong government measures such as isolation, quarantine, and public closings are implemented, the third phase begins. The actual effects of these

measures are complex, and we use an exponential decrease for a time-dependent decreasing transmission rate $\tau(t)$ in the third phase to incorporate these effects. The formula for $\tau(t)$ during the third phase is

$$\begin{cases} \tau(t) = \tau_0, \, 0 \leq t \leq N, \\ \tau(t) = \tau_0 \exp\left(-\mu\left(t - N\right)\right), \, N < t. \end{cases} \quad (4.6)$$

The date N and the value of μ are chosen so that the cumulative reported cases in the numerical simulation of the epidemic aligns with the cumulative reported case data after day N, when the public measures take effect. In this way we are able to project forward the time-path of the epidemic after the government imposed public restrictions take effect. We illustrate $\tau(t)$ in Figure 3 for a typical case.
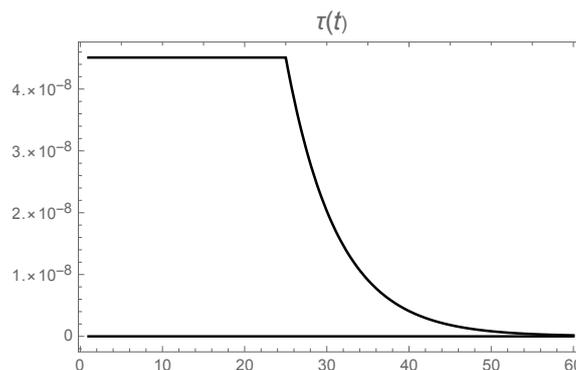

Figure 4: *Graph of $\tau(t)$ with $\tau_0 = 4.5 \times 10^{-8}$, $N = 25$, and $\mu = 0.16$. The transmission rate is effectively 0 after day 53.*

The daily number of reported cases from the model can be obtained by computing the solution of the following equation:

$$DR'(t) = \nu \, f \, I(t) - DR(t), \text{ for } t \geq t_0 \text{ and } DR(t_0) = DR_0. \quad (4.7)$$

5   Predicting the cumulative number of cases

In this section we apply the method described in section 4 to the data coming from South Korea, Italy, France and Germany [5, 6, 7, 8]. In Table 2 we choose the data (1) February 19 to Feb 30 for South Korea; (2) February 21 to March 3 for Italy; (3) February 25 to March 7 for France; (4) February 24 to March 6 for Germany.

Country	χ_1	χ_2	χ_3	t_0	μ	N	S_0	f	τ_0
South Korea	0.624	0.294	1	Feb. 2	1.3	Feb 28	51.47×10^6	0.6	7.51×10^{-9}
								0.1	6.56×10^{-9}
Italy	0.3459	0.2792	1	Feb. 4	0.032	Feb. 24	60.48×10^6	0.6	6.15×10^{-9}
								0.1	6.56×10^{-9}
France	0.0020	0.3627	1	Feb 17	0.032	Feb. 28	66.99×10^6	0.6	6.78×10^{-9}
								0.1	6.02×10^{-9}
Germany	0.0039	0.3371	1	Feb 16	0.016	Feb. 28	82.79×10^6	0.6	5.18×10^{-9}
								0.1	4.57×10^{-9}

Table 2: *The parameters χ_1, χ_2, χ_3 are estimated by using the data in Figures 1 and 2 to fit $\chi_1 \exp(\chi_2 t) - \chi_3$ to the data $CR(t)$ between the following periods for each country: (1) February 19 to Feb 30 for South Korea; (2) February 21 to March 3 for Italy; (3) February 25 to March 7 for France; (4) February 24 to March 6 for Germany. The parameters $\nu = 1/7$ and $\eta = 1/7$. The values of I_0 U_0, τ_0, and t_0 are obtained by using (4.2) to (4.4).*

5.1 Predicting the number of cases for South Korea
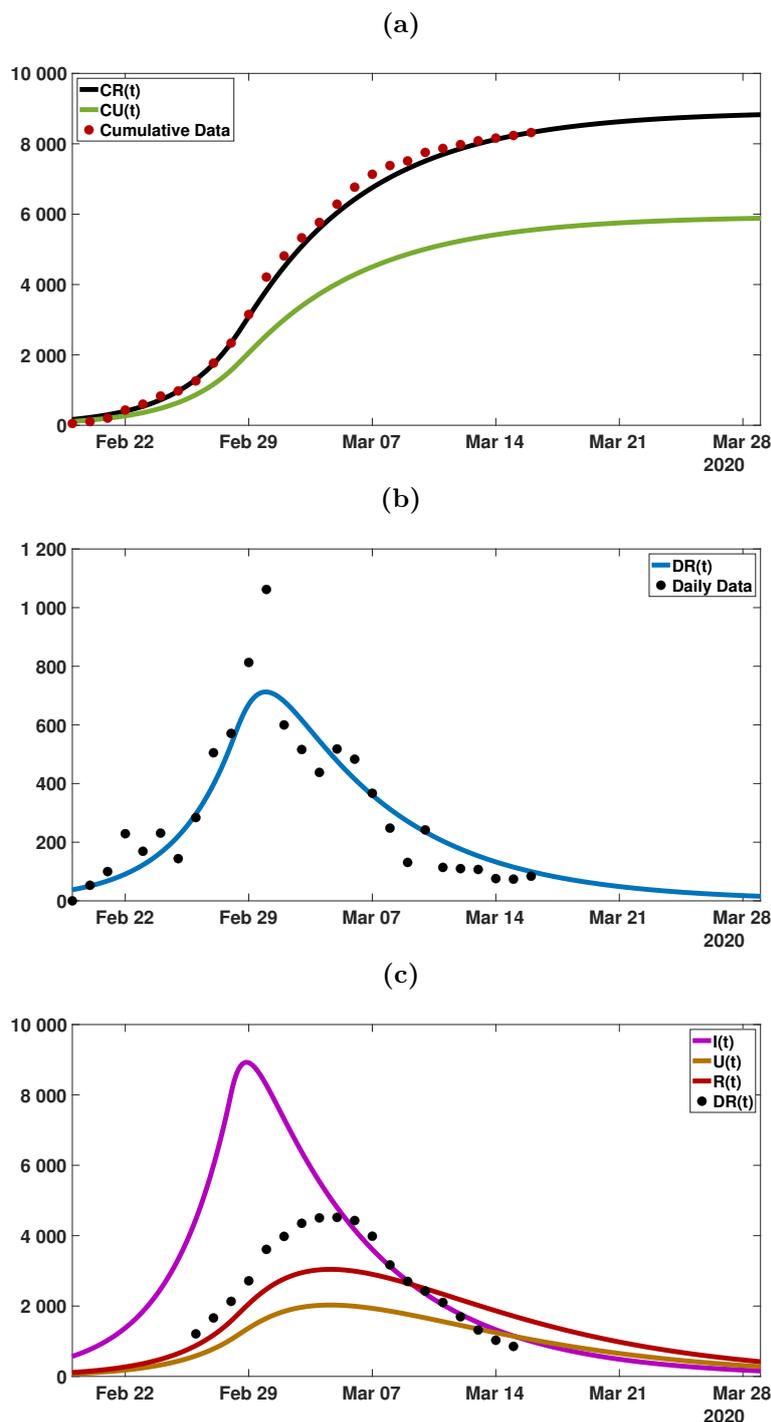
(a)

(b)

(c)

Figure 5: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.6$ (60% of the cases are reported). $I_0 = 3.43$, $U_0 = 0.45$, $\mathcal{R}_0 = 3.79$. The final size of the epidemic is approximately $8,800$ reported cases, and $5,900$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately February 30. (c) Plot of $I(t)$ (purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately March 4.*
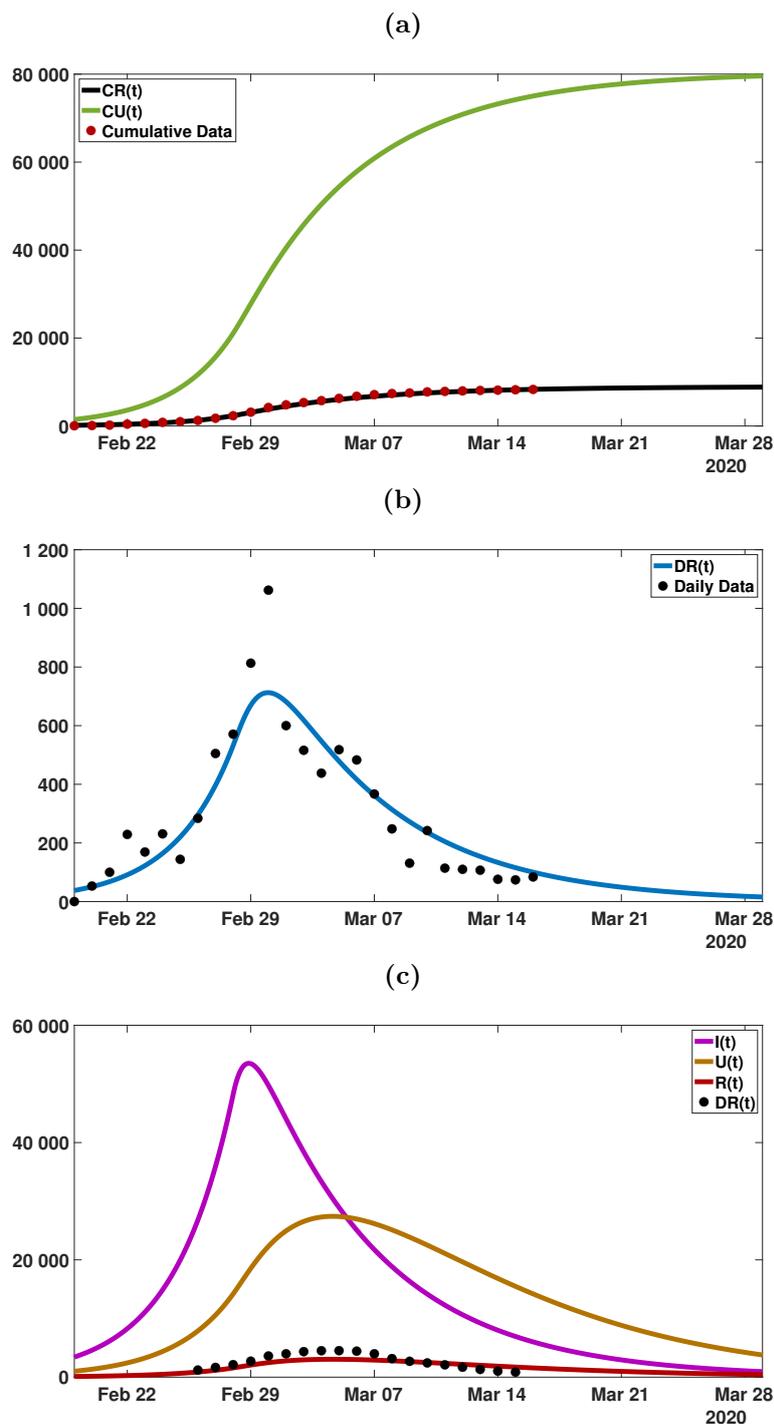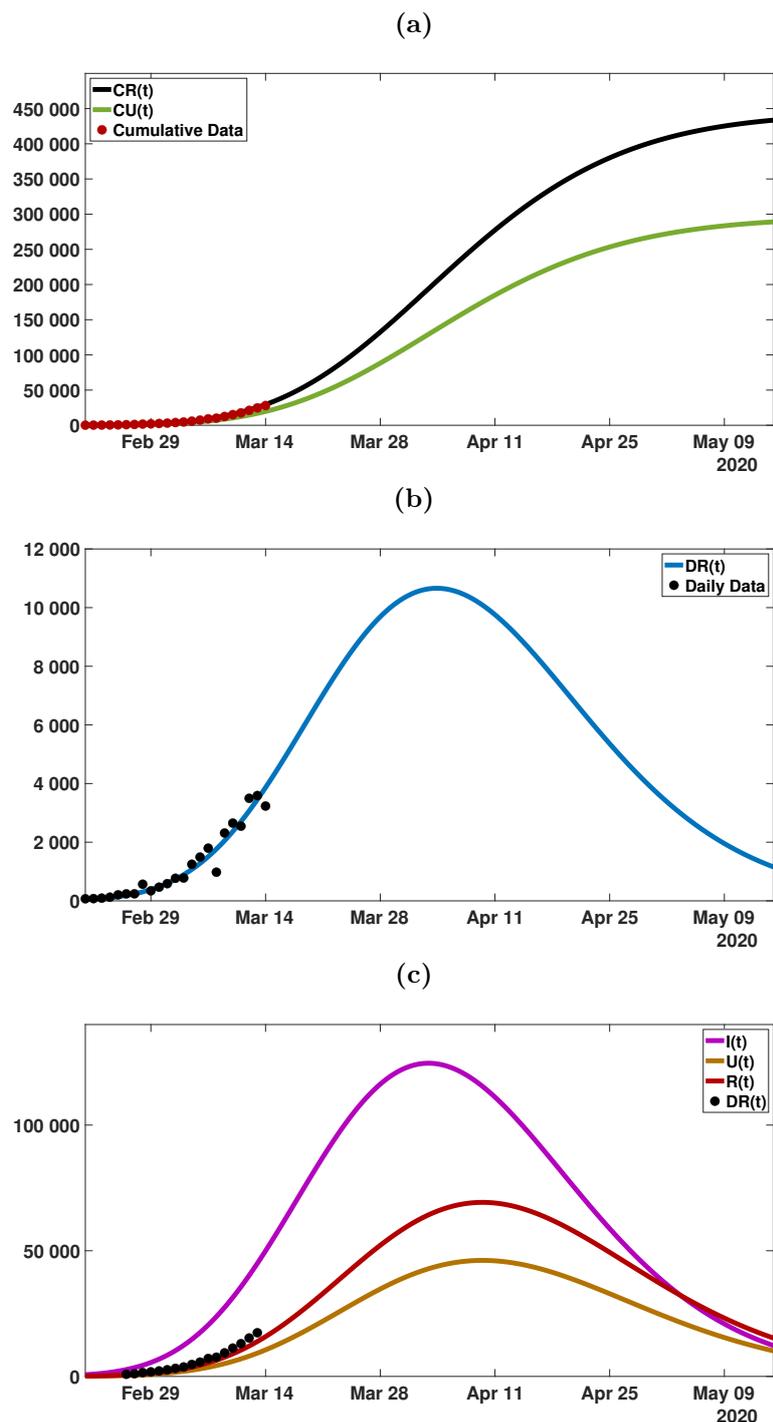
Figure 6: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.1$ (10% of the cases are reported). $I_0 = 20.6$, $U_0 = 6.1$, $\mathcal{R}_0 = 4.49$. The final size of the epidemic is approximately $10,000$ reported cases, and $80,000$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately February 30. (c) Plot of $I(t)$ (purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately March 4.*

5.2   Predicting the number of cases for Italy

(a)

(b)

(c)

Figure 7: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.6$ (60% of the cases are reported). $I_0 = 3.4$, $U_0 = 0.45$, $\mathcal{R}_0 = 3.79$. The final size of the epidemic is approximately $435,000$ reported cases, and $280,000$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately April 2. (c) Plot of $I(t)$ (purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately March 31.*
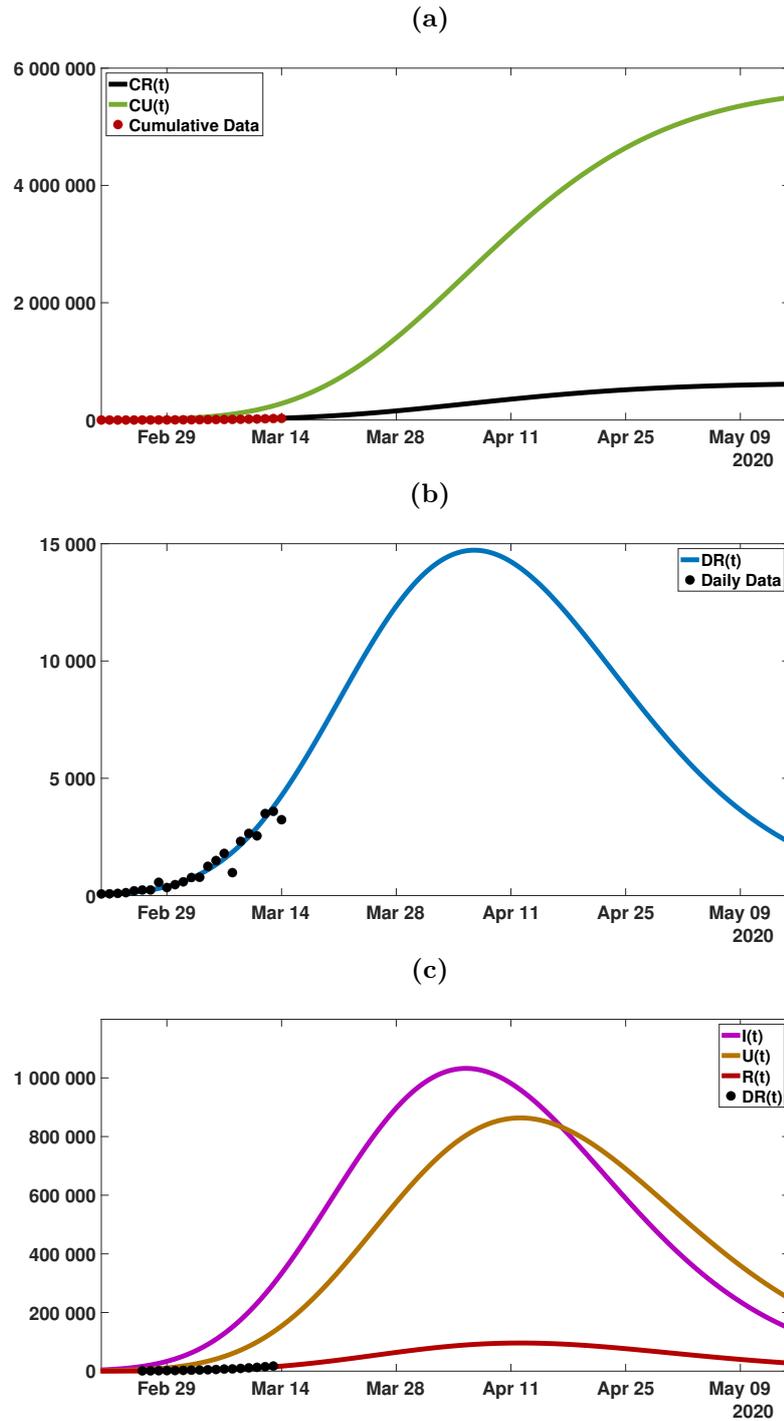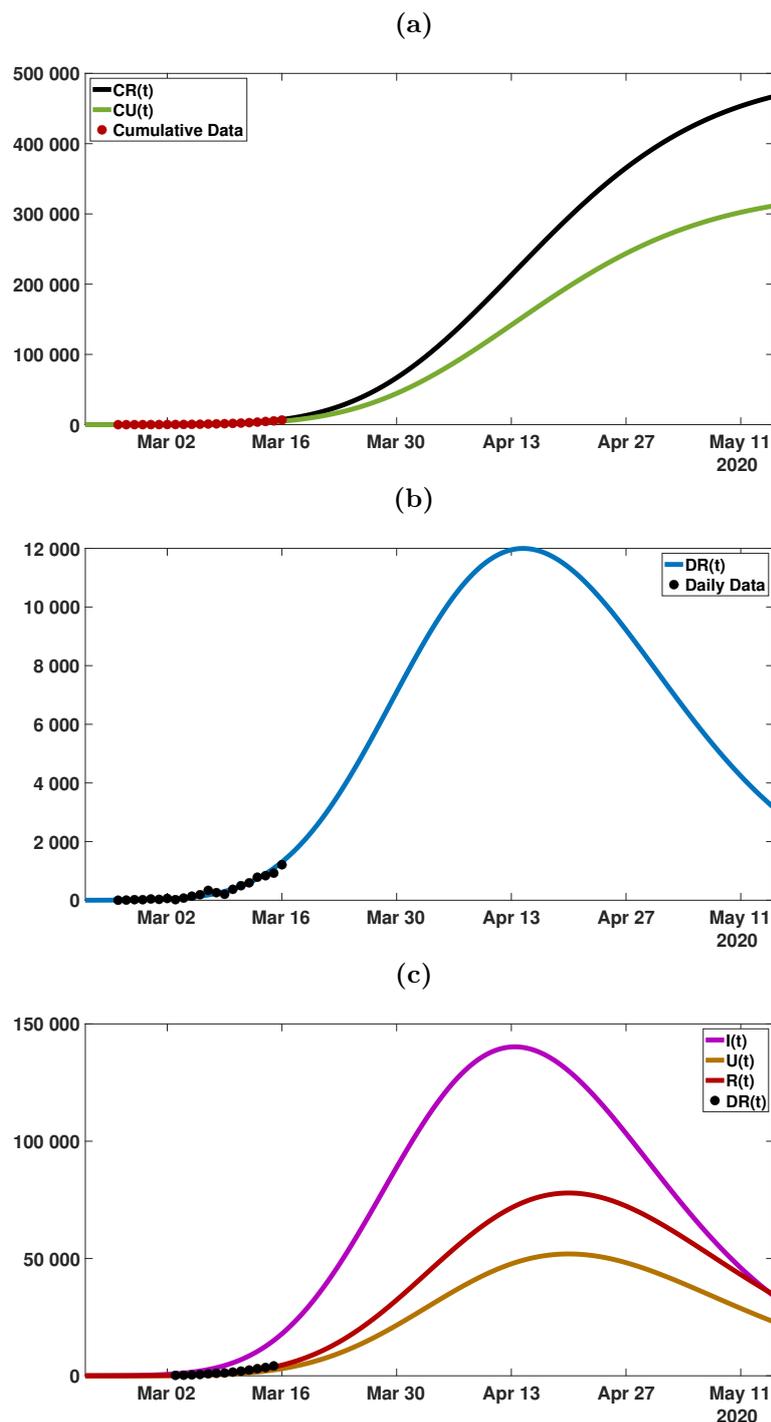
Figure 8: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.1$ (10% of the cases are reported). $I_0 = 19.5$, $U_0 = 6.0$, $\mathcal{R}_0 = 4.30$. The final size of the epidemic is approximately $600,000$ reported cases, and $5,500,000$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately April 7. (c) Plot of $I(t)$ (purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately April 11.*

5.3 Predicting the number of cases for France

(a)

(b)

(c)

Figure 9: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.6$ (60% of the cases are reported). $I_0 = 4.2$, $U_0 = 0.58$, $\mathcal{R}_0 = 4.45$. The final size of the epidemic is approximately $470,000$ reported cases, and $310,000$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately April 16. (c) Plot of $I(t)$ (purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately April 20.*
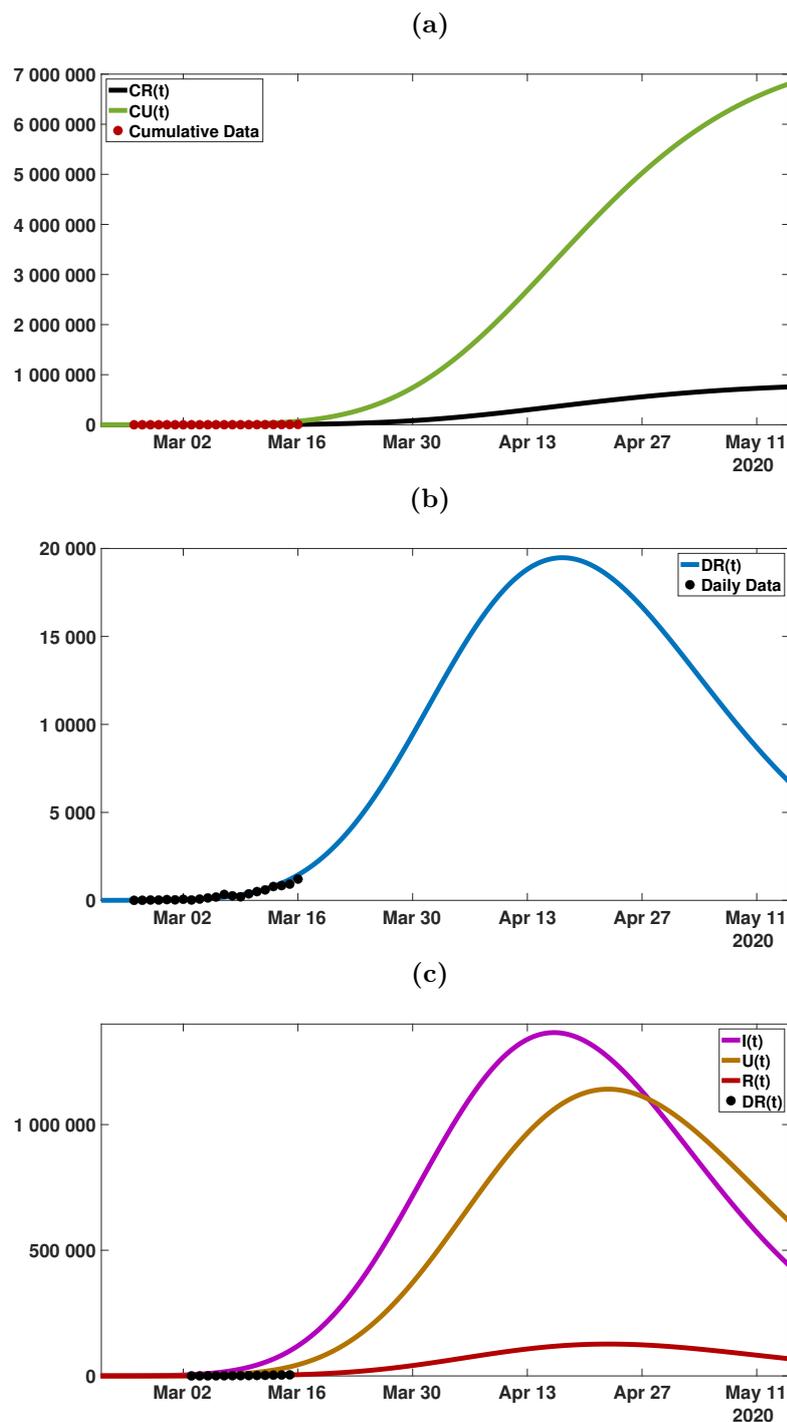
Figure 10: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.1$ (10% of the cases are reported). $I_0 = 20.6$, $U_0 = 6.1$, $\mathcal{R}_0 = 4.49$. The final size of the epidemic is approximately $700,000$ reported cases, and $6,800,000$ unreported cases. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. The turning point is approximately April 17. (c) Plot of $I(t)$(purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier). The turning point of the weekly data and $CR(t)$ and $CU(t)$ is approximately April 20.*

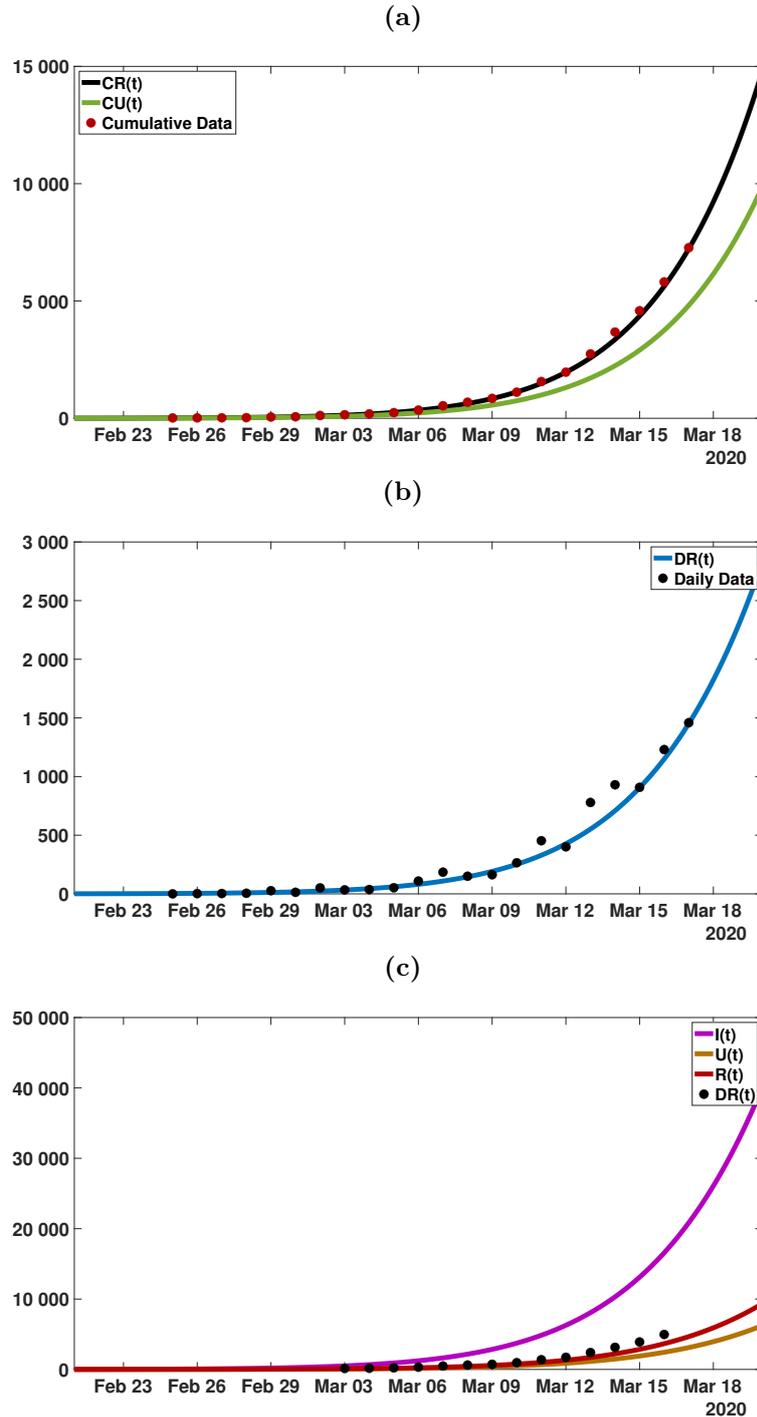## 5.4 Predicting the number of cases for Germany

Figure 11: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.6$ (60% of the cases are reported). $I_0 = 3.9$, $U_0 = 0.47$, $\mathcal{R}_0 = 4.21$. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. (c) Plot of $I(t)$(purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier).*
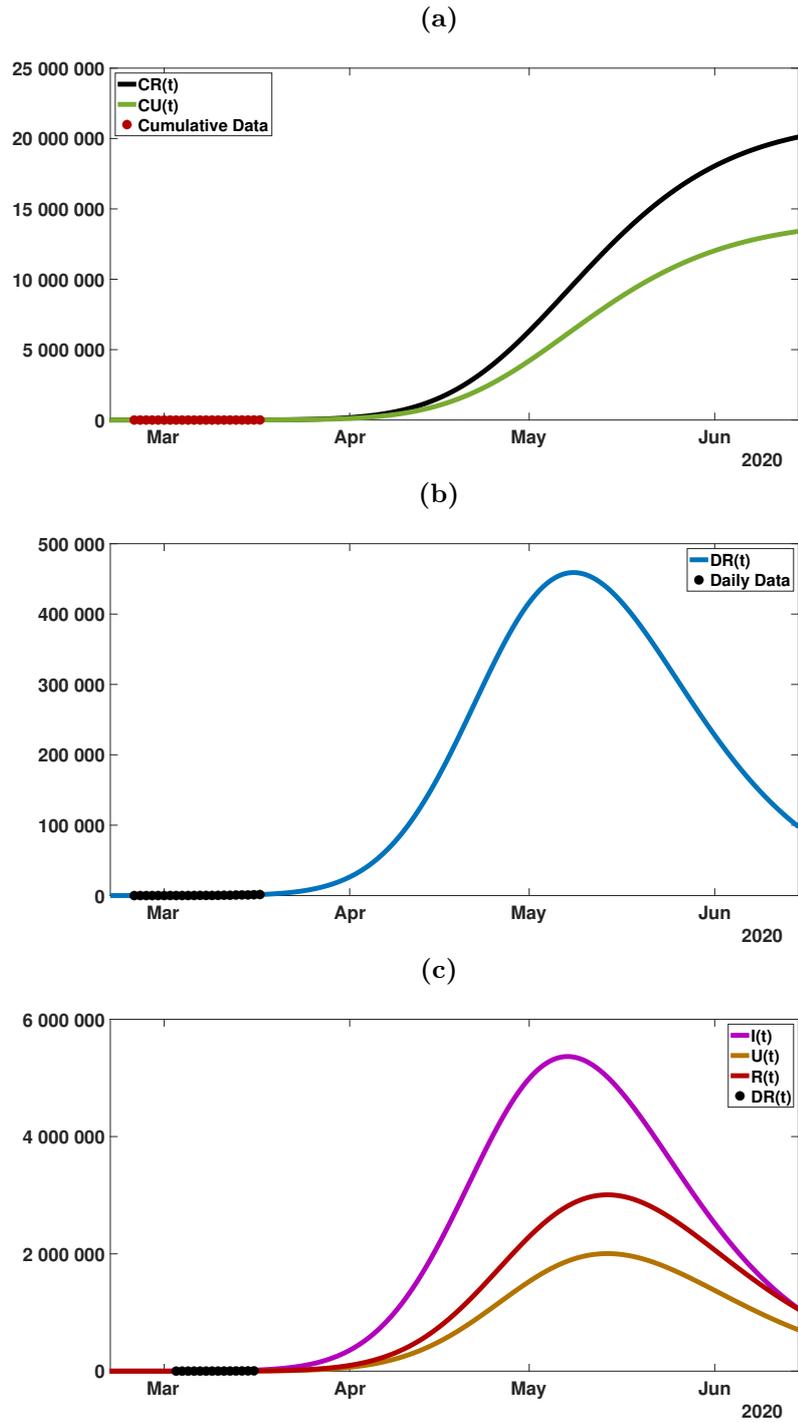
12

(a)

(b)

(c)

Figure 12: *Graphs of the model simulation and data as in Figure 11, for a longer time period.*
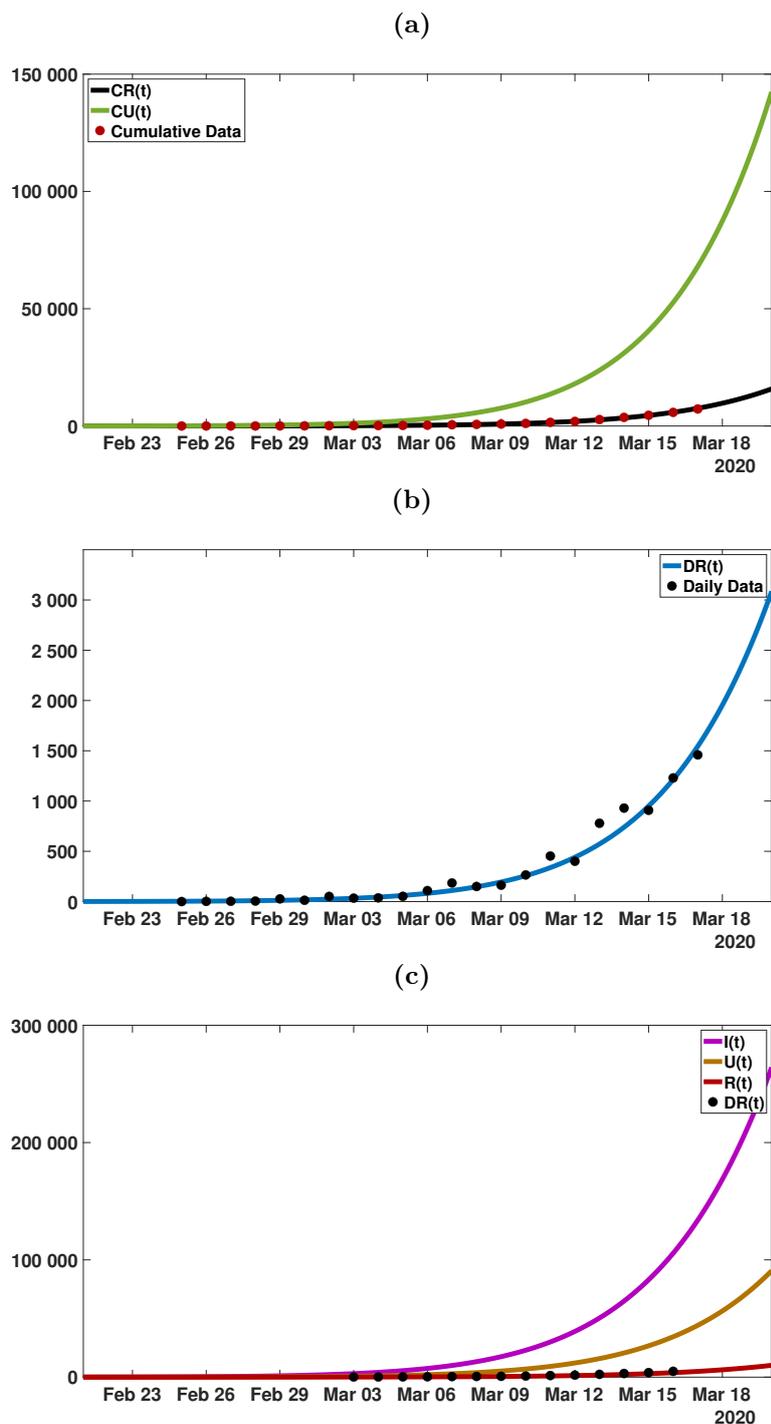
(a)

(b)

(c)

Figure 13: *(a) Plot of the data for cumulative reported cases (red dots), $CR(t)$ (black), and $CU(t)$ (green) from the model, for $f = 0.1$ (10% of the cases are reported). $I_0 = 23.6$, $U_0 = 6.3$, $\mathcal{R}_0 = 5.03$. (b) Plot of the data (black dots) for the daily number of cases and $DR(t)$ (blue) from the model. (c) Plot of $I(t)$(purple), $U(t)$ (orange), $R(t)$ (red) and the weekly data (blue dots) (each day minus 7 days earlier).*
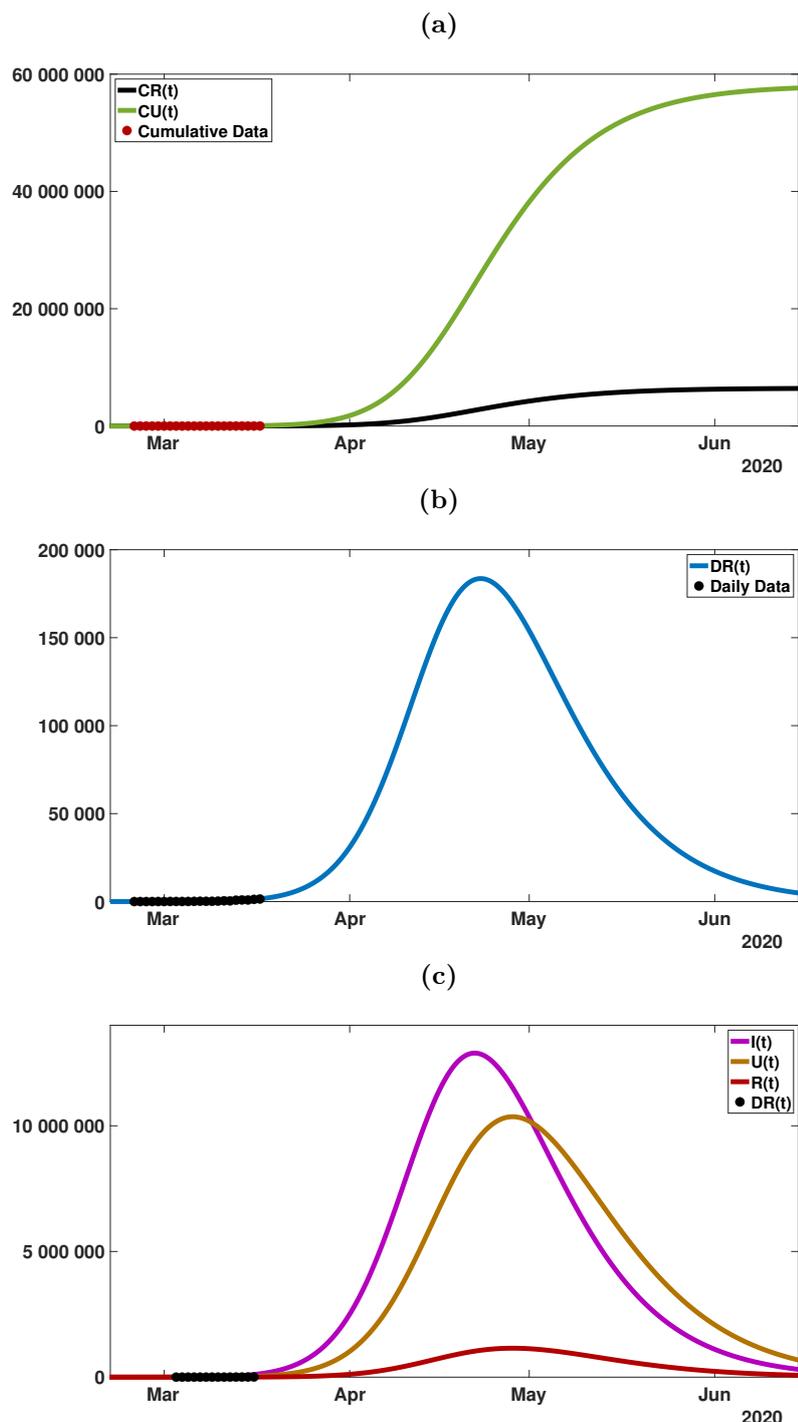
(a)

(b)

(c)

Figure 14: *Graphs of the model simulation and data as in Figure 13, for a longer time period.*

6   Conclusion

We have applied a method developed in [1] and [2] to predict the evolution of a COVID-19 epidemic in a geographical region, based of reported case data in that region. Our method uses early data, when the epidemic is in its second phase and growing exponentially, corresponding to a constant transmission rate. In [1] we demonstrated a method to identify this constant transmission rate. When public measures are begun in order to ameliorate the epidemic, a third phase begins, which we model with a time-dependent exponentially decreasing transmission rate in [2]. In [2] we applied this method to mainland China, and demonstrated the ability of our model to predict the forward time-line of the epidemic. In Figure 4 in [2], we showed how the prediction unfolded week by week, with increasing agreement with reported case

15

data, in mainland China.

In this work we use the cumulative number of reported cases, the daily number of reported cases, and the weekly number of reported cases for four countries. With this data, we project the future number of cases, both reported and unreported, in each country. We assume a fraction f of total cases are unreported. We simulate our model with $f = 0.6$ and $f = 0.1$ for each country. The fractions $f = 0.6$ and $f = 0.1$ correspond to lesser and greater final sizes, respectively. For South Korea, the epidemic has subsided, because of the major measures that were implemented to restrict public distancing. For Italy, France, and Germany, the epidemics are still rising, because these measures have been implemented only recently.

Our model incorporates social distancing measures through the time dependent transmission rate $\tau(t)$ in the third phase. This rate involves μ (the rate of exponential decay corresponding to these measures) and N the day at which they become effective. It is evident that these measures should start as early as possible, and should be as strong as possible. The consequences of late public interventions may have severe consequences for the epidemic outcome (as illustrated in Figure 6 in [2]).

In the case of South Korea, the peak of the epidemic occurred approximately February 29. In Figures 5 and 6, we see that our model agrees very well the data for South Korea. Accordingly to our model, the daily number of cases reaches a maximum of approximately 700 cases near the turning point February 29.

Compared to South Korea, the public interventions in Italy, France, and Germany were relatively late. The peak of the epidemic occurs in Italy around April 9, in France around April 14, and in Germany around May 1. The maximum daily number of cases is between 10 000 and 15 000 in Italy, between 12 000 and 20 000 in France, and between 200 000 and 500 000 in Germany. These numbers will likely remain high for an extended period of several months for these three countries, without further major public measures.

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

References

[1] 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*, 2020, 9(3), 50.

[2] 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, *medRxiv*, 2020.

[3] R. Verity, L.C. Odell, I. Dorigatti, *et al.*, Estimates of the severity of COVID -19 disease, *medRxiv*, 2020.

[4] A. Kucharski, T.W. Russell, C. Diamond, *et al.*, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet*, March 11, 2020.

[5] https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[6] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_France

[7] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_Germany

[8] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_Italy

[9] https://en.wikipedia.org/wiki/2020_coronavirus_outbreak_in_South_Korea