

# Assessing the impact of data-driven limitations on tracing and forecasting the outbreak dynamics of COVID-19

Giulia Fiscon<sup>1,2</sup>, Francesco Salvatore<sup>3</sup>, Valerio Guarrasi<sup>4</sup>, Anna Rosa Garbuglia<sup>5\*</sup>, Paola Paci<sup>1,4</sup>

1. *Institute for Systems Analysis and Computer Science "Antonio Ruberti", National Research Council, Rome, Italy*
2. *Fondazione per la Medicina Personalizzata, Via Goffredo Mameli, 3/1 Genova, Italy*
3. *CINECA, HPC Department, Rome Office, Italy*
4. *Department of Computer, Control and Management Engineering "A. Ruberti" (DIAG), Sapienza University of Rome, Rome, Italy*
5. *Laboratory of Virology, Lazzaro Spallanzani National Institute for Infectious Diseases, IRCCS, Rome, Italy*

\* Corresponding author: [argarbuglia@iol.it](mailto:argarbuglia@iol.it)

## Supplementary Information

This file reported the results obtained by applying the standard SIR model and the A-SIR model to study the COVID-19 epidemic dynamic in other countries with similar trends to Italy (i.e. Spain, Germany, and France). Our outcomes confirmed what has already observed for Italy, that is the SIR model failed to properly reproduce data in the presence of undetected asymptomatic infectives (first peak of infection), overestimating certain very relevant parameters and underestimating others; whereas we found that the SIR model well-fitted with the wavy behaviour characterizing the infected population curve when detected asymptomatic infectives were part of the total pool of infectives (second peak of infection).

## Spain

The initial conditions for the SIR model are reported in Table S1, whereas the first and second wave-modelling resulting values for the optimized parameters are reported in Table S2.

**Table S1. Initial configuration for the SIR model.**

Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
<b>Initial condition</b>		
N	46'754'778	
S	$N - I_0$	
I	$I_0$	
R	0	

N is equal to the population of Spain in January 2020 according to <https://www.worldometers.info/demographics/spain-demographics/>;  $I_0$  is the number of infectives at time zero, i.e. at the 7<sup>th</sup> of March 2020 for the first wave modelling (i.e., 460 infectives) and at the 8<sup>th</sup> of August 2020 for the second wave modelling (i.e., 135'483 infectives).

**Table S2. Parameters for the SIR model obtained through the fit of I(t) for the first wave and second wave.**

	First wave	Second wave
<b>Fitting time window</b>		
start	7 <sup>th</sup> of March 2020	8 <sup>th</sup> of August 2020
end	30 <sup>th</sup> of April 2020	30 <sup>th</sup> of September 2020
<b>Estimated Parameters</b>		
$\beta$	1	0.2
$\gamma$	0.89	0.17
$R_0$	1.13	1.2

The initial conditions for the A-SIR model are reported in Table S3, whereas the first and second wave-modelling resulting values for the optimized parameters are reported in Table S4.

**Table S3. Initial configuration for the A-SIR model.**

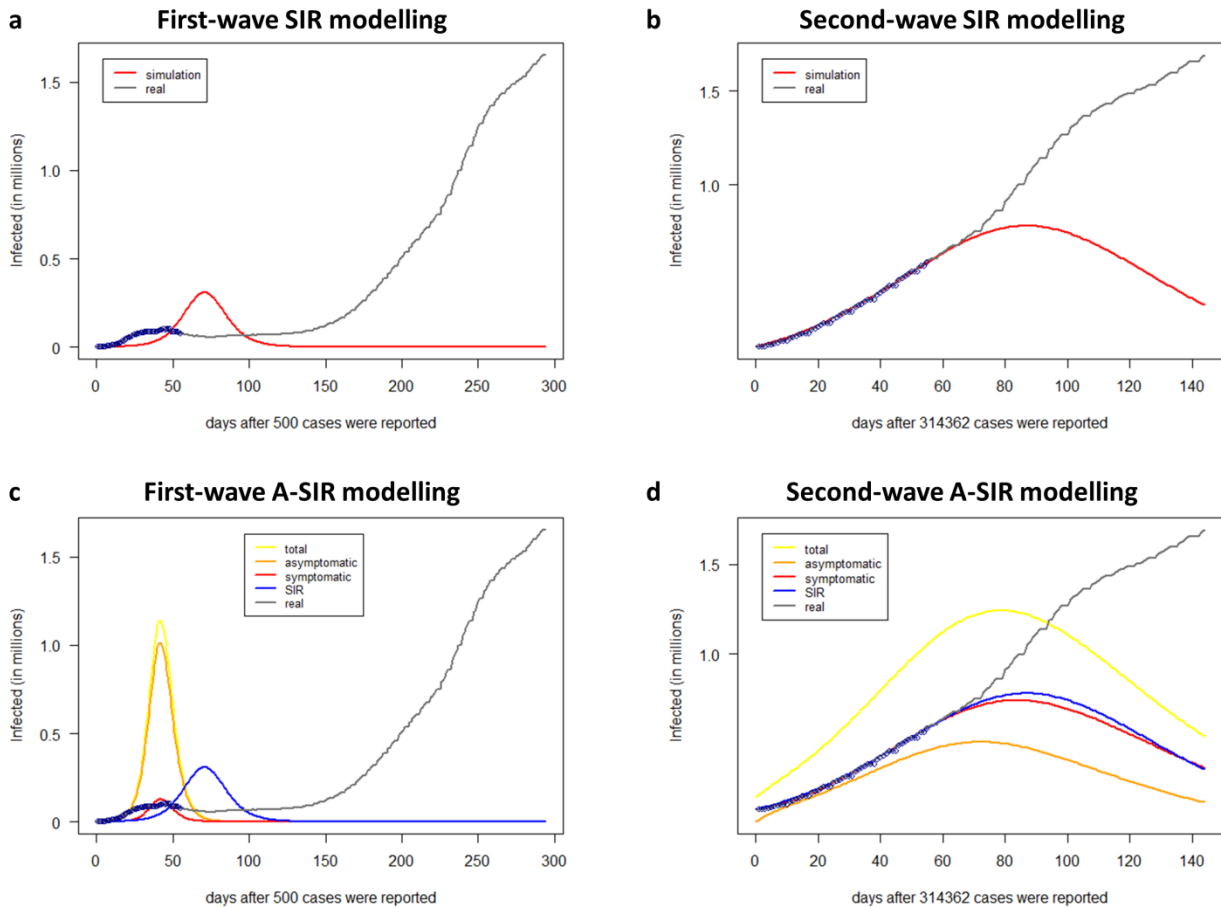
Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
$\eta$	0.5	(0,1)
$\xi$	0.15	(0.1,0.3)
<b>Initial condition</b>		
N	46'754'778	N
S	$N - I_0 - \sigma I_0$	
I	$I_0$	
J	$\sigma I_0$	
R	0	
U	0	

N is equal to the population of Spain in January 2020 according to <https://www.worldometers.info/demographics/spain-demographics/>;  $I_0$  is the number of infectives at time zero, i.e., at the 7<sup>th</sup> of March 2020 for the first wave modelling (equal to 460 infectives) and at the 8<sup>th</sup> of August 2020 for the second wave modelling (equal to 135'483 infectives);  $\sigma = 0.5$ .

**Table 4. Parameters for the A-SIR model obtained through the fit of I(t) for the first wave and the second wave.**

	First wave	Second wave
<b>Fitting time window</b>		
start	7 <sup>th</sup> of March 2020	8 <sup>th</sup> of August 2020
end	30 <sup>th</sup> of April 2020	30 <sup>th</sup> of September 2020
<b>Estimated Parameters</b>		
$\beta$	1	0.22
$\gamma$	0.7	0.06
$\eta$	0.79	0.31
$\xi$	0.1	0.23
$R_0$	1.4	2.42

The epidemic dynamics of the infected class of individuals predicted by the SIR and A-SIR model for both the first and second wave of SARS-CoV-2 infection in Spain are shown in Figure S1.



**Figure S1. Predictions for the infectives  $I(t)$  provided by the SIR model (a-b) and the A-SIR model (c-d) for the first and second wave in Spain.** (a-b). The infected individuals predicted by the SIR model (red line) are plotted as a function of days together with the real observed infected people (grey line). (c-d) The total number of infected individuals  $I+J$  (yellow lines), symptomatic infected  $I$  (red lines), and asymptomatic infected  $J$  (orange lines) predicted by the A-SIR model are plotted as a function of days, together with the infected individuals predicted by the SIR model (blue lines) and the real observed infected people (grey lines). The dark blue points represent real data used to estimate the optimal parameters for the best fitting, i.e. the cumulative number of individuals reported to be infected in Spain up to the 30<sup>th</sup> of April 2020 after the day that 500 cases were reported (the 7<sup>th</sup> of March 2020) for the first peak (a-c) and up to the 30<sup>th</sup> of September 2020 after that 314362 cases were reported (the 8<sup>th</sup> of August 2020) for the second peak (b-d).

## Germany

The initial conditions for the SIR model are reported in Table S5, whereas the first and second wave-modelling resulting values for the optimized parameters are reported in Table S6.

**Table S5. Initial configuration for the SIR model.**

Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
<b>Initial condition</b>		
N	83'783'942	
S	$N - I_0$	
I	$I_0$	
R	0	

N is the initial population equal to the population of Germany in January 2020 according to <https://www.worldometers.info/demographics/germany-demographics/>.  $I_0$  is the number of infectives at time zero, i.e. at the 6<sup>th</sup> of March 2020 for the first wave (equal to 653 infectives) and at the 7<sup>th</sup> of October 2020 for the second wave modelling (equal to 31'833 infectives).

**Table S6. Parameters for the SIR model obtained through the fit of I(t) for the first wave and second wave.**

	First wave	Second wave
<b>Fitting time window</b>		
start	6 <sup>th</sup> of March 2020	7 <sup>th</sup> of October 2020
end	30 <sup>th</sup> of April 2020	1 <sup>st</sup> of December 2020
<b>Estimated Parameters</b>		
$\beta$	1	0.15
$\gamma$	0.91	0.12
$R_0$	1.1	1.18

The initial conditions for the A-SIR model are reported in Table S7 and the first and second wave-modelling resulting values for the optimized parameters are reported in Table S8.

**Table S7. Initial configuration for the A-SIR model.**

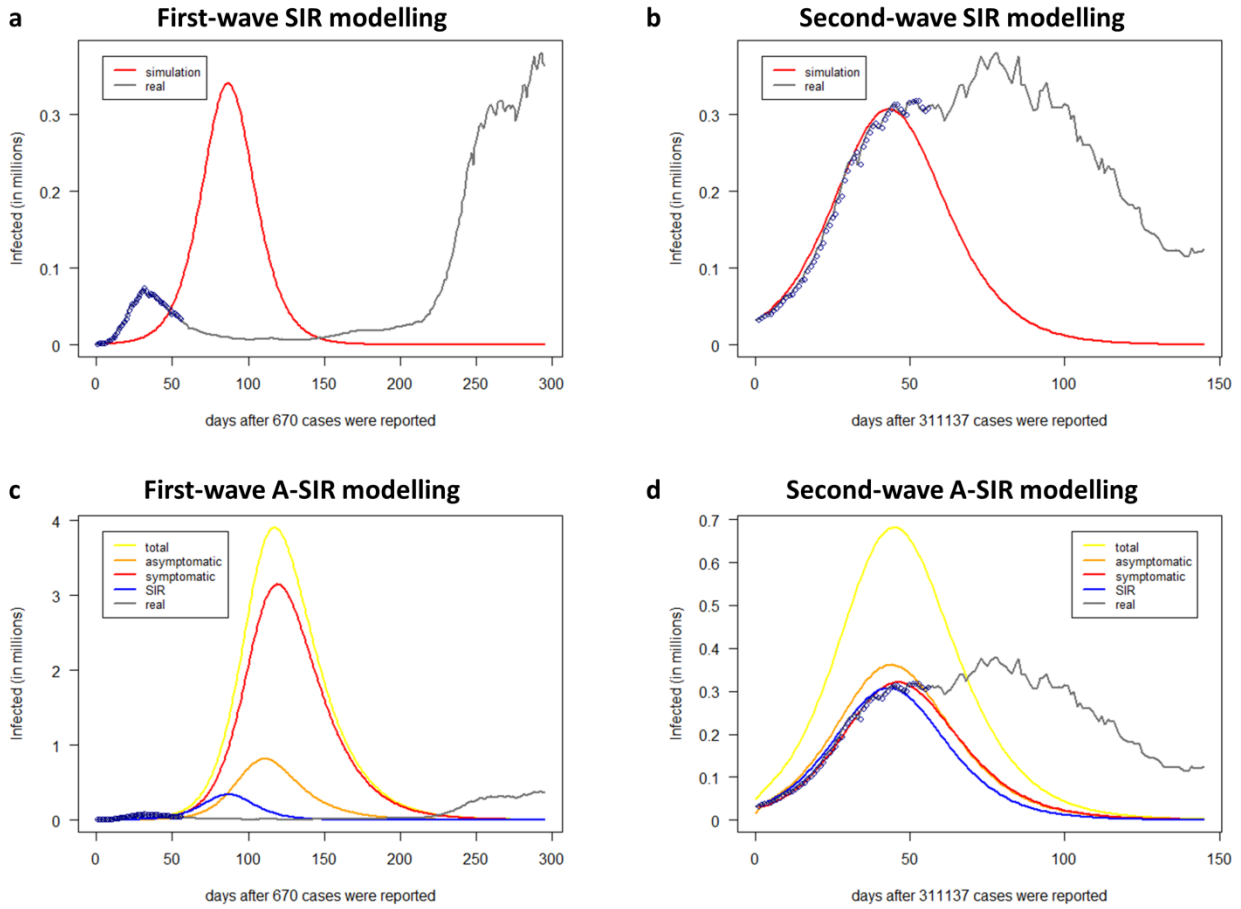
Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
$\eta$	0.5	(0,1)
$\xi$	0.15	(0.1,0.3)
<b>Initial condition</b>		
N	83'783'942	
S	$N - I_0 - \sigma I_0$	
I	$I_0$	
J	$\sigma I_0$	
R	0	
U	0	

N is the initial population equal to the population of Germany in January 2020 according to <https://www.worldometers.info/demographics/germany-demographics/>.  $I_0$  is the number of infectives at time zero, i.e. at the 6<sup>th</sup> of March 2020 for the first wave (equal to 653 infectives) and at 7<sup>th</sup> of October 2020 for the second wave modelling (equal to 31'833 infectives);  $\sigma = 0.5$ .

**Table S8. Parameters for the A-SIR model obtained through the fit of I(t) for the first wave and the second wave.**

Estimated Parameters	First wave	Second wave
$\beta$	0.41	0.79
$\gamma$	0.88	0.32
$\eta$	0.88	1
$\xi$	0.3	0.22
$R_0$	4.67	2.49

The epidemic dynamics of the infected class of individuals predicted by the SIR and A-SIR model for both the first and second wave of SARS-CoV-2 infection in Germany are shown in Figure S2.



**Figure S2. Predictions for the infectives  $I(t)$  provided by the SIR model (a-b) and the A-SIR model (c-d) for the first and second wave in Germany.** (a-b). The infected individuals predicted by the SIR model (red line) are plotted as a function of days together with the real observed infected people (grey line). (c-d) The total number of infected individuals  $I+J$  (yellow lines), symptomatic infected  $I$  (red lines), and asymptomatic infected  $J$  (orange lines) predicted by the A-SIR model are plotted as a function of days, together with the infected individuals predicted by the SIR model (blue lines) and the real observed infected people (grey lines). The dark blue points represent real data used to estimate the optimal parameters for the best fitting, i.e. the cumulative number of individuals reported to be infected in Germany up to the 30<sup>th</sup> of April 2020 after the day that 670 cases were reported (the 6<sup>th</sup> of March 2020) for the first peak (a-c) and up to the 1<sup>st</sup> of December 2020 after that 311137 cases were reported (the 7<sup>th</sup> of October 2020) for the second peak (b-d).

## France

The initial conditions for the SIR model are reported in Table S9, whereas the first and second wave-modelling resulting values for the optimized parameters are reported in Table S10.

**Table S9. Initial configuration for the SIR model.**

Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
<b>Initial condition</b>		
N	65*273*511	
S	$N - I_0$	
I	$I_0$	
R	0	

N is the initial population equal to the population of France in January 2020 according to <https://www.worldometers.info/demographics/france-demographics/>.  $I_0$  is the number of infectives at time zero, i.e. at the 6<sup>th</sup> of March 2020 for the first wave (equal to 592 infectives) and at the 2<sup>nd</sup> of September 2020 for the second wave modelling (equal to 211'554 infectives).

**Table S10. Parameters for the SIR model obtained through the fit of I(t) for the first wave and second wave.**

	First wave	Second wave
<b>Fitting time window</b>		
start	6 <sup>th</sup> of March 2020	2 <sup>nd</sup> of September 2020
end	30 <sup>th</sup> of April 2020	25 <sup>th</sup> of October 2020
<b>Estimated Parameters</b>		
$\beta$	0.55	0.05
$\gamma$	0.45	0.03
$R_0$	1.23	1.97

The initial conditions for the A-SIR model are reported in Table S11 and the first and second wave-modelling resulting values for the optimized parameters are reported in Table S12.



**Table S11. Initial configuration for the A-SIR model.**

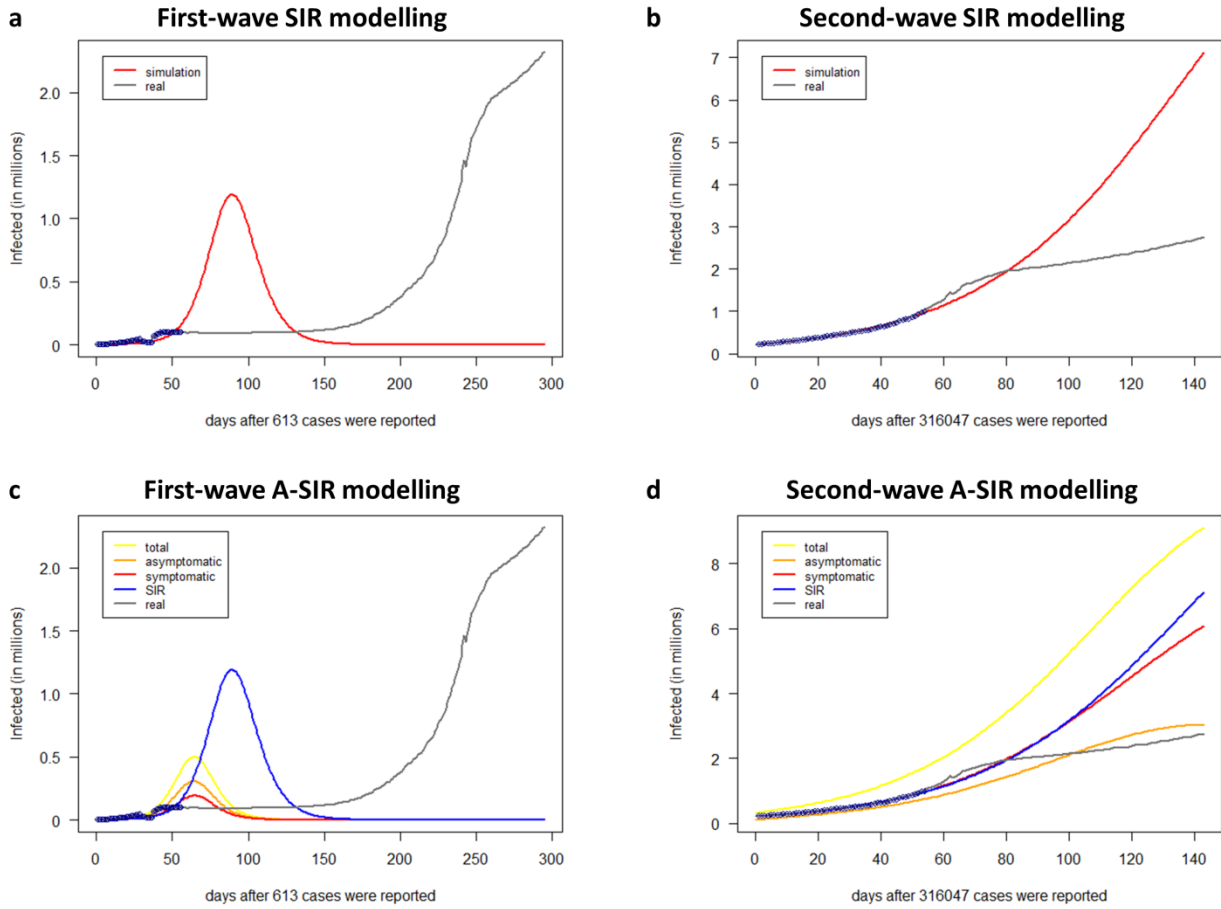
Initial parameters	Value	Constraints interval
$\beta$	0.5	(0,1)
$\gamma$	0.5	(0,1)
$\eta$	0.5	(0,1)
$\xi$	0.15	(0.1,0.3)
<b>Initial condition</b>		
N	65'273'511	
S	$N - I_0 - \sigma I_0$	
I	$I_0$	
J	$\sigma I_0$	
R	0	
U	0	

N is the initial population equal to the population of France in January 2020 according to <https://www.worldometers.info/demographics/france-demographics/>;  $I_0$  is the number of infectives at time zero, i.e. at the 6<sup>th</sup> of March 2020 for the first wave (equal to 592 infectives) and at the 2<sup>nd</sup> of September 2020 for the second wave modelling (equal to 211'554 infectives);  $\sigma = 0.5$ .

**Table S12. Parameters for the A-SIR model obtained through the fit of I(t) for the first wave and the second wave.**

Estimated Parameters	First wave	Second wave
$\beta$	1	0.07
$\gamma$	0.68	0.008
$\eta$	1	0.09
$\xi$	0.29	0.3
$R_0$	1.47	8.48

The epidemic dynamics of the infected class of individuals predicted by the SIR and A-SIR model for both the first and second wave of SARS-CoV-2 infection in France are shown in Figure S3.



**Figure S3. Predictions for the infectives  $I(t)$  provided by the SIR model (a-b) and the A-SIR model (c-d) for the first and second wave in France.** (a-b). The infected individuals predicted by the SIR model (red line) are plotted as a function of days together with the real observed infected people (grey line). (c-d) The total number of infected individuals  $I+J$  (yellow lines), symptomatic infected  $I$  (red lines), and asymptomatic infected  $J$  (orange lines) predicted by the A-SIR model are plotted as a function of days, together with the infected individuals predicted by the SIR model (blue lines) and the real observed infected people (grey lines). The dark blue points represent real data used to estimate the optimal parameters for the best fitting, i.e. the cumulative number of individuals reported to be infected in France up to the 30<sup>th</sup> of April 2020 after the day that 613 cases were reported (the 6<sup>th</sup> of March 2020) for the first peak (a-c) and up to the 25<sup>th</sup> of October 2020 after that 316047 cases were reported (the 2<sup>nd</sup> of September 2020) for the second peak (b-d).

Finally, the smoothest versions of the estimates of  $R_0$  by considering  $R_0(t | 0, 25)$  for all the above-discussed countries were calculated and shown in Figure S4, in comparison with respect to those ones obtained for Italy.



**Figure S4. Estimation of the basic reproduction number  $R_0$ .** The smoothed versions of the estimates of  $R_0$  for  $R_0(t | 0, 25)$  are plotted as function of time  $t$  (days) for Spain (a), Germany (b), and France (c) (red curves), together with the ones obtained for Italy (green curves). We assumed as  $t=0$  the 22nd of January 2020, typical assessment of the start date of the pandemic outbreak in Wuhan. The maximum and the minimum value of the estimated  $R_0$  for each country in the analysed time windows were highlighted.