

1 **Estimating the size of the COVID-19 outbreak in Italy:**
2 **Application of an exponential decay model to the weighted and**
3 **cumulative average daily growth rate**

4 **Short title:** Estimating the size of the COVID-19 outbreak in Italy with an
5 exponential decay model

6 Authors:

7 Nicola Bartolomeo^{1¶*}, Paolo Trerotoli^{1¶}, Gabriella Serio^{1¶}

8

9

10

11 ¹ Department of Biomedical Science and Human Oncology, University of Bari Aldo Moro, Bari,
12 Italy

13

14

15 * Corresponding author

16 E-mail: nicola.bartolomeo@uniba.it (NB)

17

18

19 [¶] These authors contributed equally to this work.

20

21 **Abstract**

22 To estimate the size of the novel coronavirus (COVID-19) outbreak in Italy, this paper introduces
23 the cumulated and weighted average daily growth rate (WR) to evaluate an epidemic curve, On the
24 basis of an exponential decay model (EDM), we provide estimations of the WR in four-time
25 intervals from February 27 to April 07, 2020. By calibrating the parameters of the EDM to the
26 reported data in Hubei Province of China, we also attempt to forecast the evolution of the outbreak.
27 We compare the EDM applied to WR and the Gompertz model, which is based on exponential
28 decay and is often used to estimate cumulative events. Specifically, we assess the performance of
29 each model to short-term forecast of the epidemic, and to predict the final epidemic size.
30 Based on the official counts for confirmed cases, the model applied to data from February 27 until
31 the 17st of March estimate that the cumulative number of infected could reach 131,280 (with a
32 credibility interval 71,415-263,501) by April 25 (credibility interval April 12 to May 3). With the
33 data available until the 24st of March the peak date should be reached on May 3 (April 23 to May
34 23) with 197,179 cumulative infections expected (130033-315,269); with data available until the
35 31st of March the peak should be reached on May 4 (April 25 to May 18) with 202,210 cumulative
36 infections expected (155.235-270,737); with data available until the 07st of April the peak should
37 be reached on May 3 (April 26 to May 11) with 191,586 (160,861-232,023) cumulative infections
38 expected. Based on the average mean absolute percentage error (MAPE), cumulated infections
39 forecasts provided by the EDM applied to WR performed better across all scenarios than the
40 Gompertz model.
41 An exponential decay model applied to the cumulated and weighted average daily growth rate
42 appears to be useful in estimating the number of cases and peak of the COVID-19 outbreak in Italy
43 and the model was more reliable in the exponential growth phase.

44

45

46 **Introduction**

47 In December 2019, an outbreak of a novel coronavirus (COVID-19) occurred in Hubei Province,
48 China. Since then, COVID-19 has spread around the world, and on March 11, 2020, it was declared
49 a pandemic by the World Health Organization (WHO). As of April 30, 2020, there have been
50 3,090,445 cases and 217,769 deaths, reported globally. Italy has been one of the worst affected
51 countries, and so far has reported 203,951 cases and 27682 deaths (as of April 30, 2020) [1]. The
52 Italian government has implemented a series of social distancing and public health strategies to try
53 and reduce the spread of COVID-19, including closing schools, screening passengers at railway
54 stations and airports, identifying active cases, and isolating suspected cases. On March 9, 2020,
55 Italy became the first country to implement a nation-wide lockdown to try and limit the COVID-19
56 spread [2].

57 While the transmission potential of COVID-19 can be high [3,4], the epidemiological features of
58 COVID-19 are still unclear. Differences in the reporting of cases and deaths complicate any
59 analysis of the COVID-19 epidemic. Governments need to be able to accurately estimate cases in a
60 geographical area so that social distancing and public health measures can be implemented quickly.
61 Besides, it is essential to formulate a plan to ensure that patients with COVID-19 can be cared for in
62 the hospital if required.

63 Many scholars have been attempting to estimate the basic reproduction number (R_0) and predict the
64 future trajectory of the COVID-2019 outbreak. In this paper, we use phenomenological models
65 without detailed microdata, which allows for simple calibrations to be made to the reported
66 empirical data so that quick interpretations can be formulated.

67 Many authors have attempted to use the cumulative number of new infections to estimate the
68 "peak" of the outbreak with the main models used in epidemiology, SIR, SEIR, Log-Linear,
69 Negative-Binomial. However, these forecasting models are limited in that the transmission speed of
70 new viruses, such as SARS-CoV-2 is unknown. As such, many epidemiologists have reported that it
71 is challenging to make medium and long-term predictions regarding the peak.

72 Studying the spread of COVID-19 in a country such as China could be useful. In particular, Hubei
73 Province, which is similar to Italy in terms of resident population numbers, could provide vital
74 information for exploring the dynamics of COVID-19, despite the social contexts between the two
75 locations being very different.

76 The contagion curve of China and/or the Hubei Province could be used to create epidemiological
77 growth models, from which the parameters could be used as a starting point for other populations.
78 However, the absolute number of infections depends on a number of population-based
79 characteristics, such as ethnicity and age structure.

80 Another useful indicator to evaluate the speed of the virus spread, which can be used comparatively
81 between two populations, is the daily growth rate and its associated trend. The daily growth rate,
82 which is the number of new cases at time t divided by the total number of cases recorded up to time
83 $t-1$, is a useful indicator for assessing the daily spread and speed of infection. The peak of an
84 epidemic is considered to be when the growth rate is zero. This rate can be comparatively used as it
85 is affected by the intrinsic characteristics of the virus. However, it may be affected by the particular
86 detection conditions in a single day (number of tests carried out, time of detection, delay in the
87 processing of tests by laboratories, partial communication of daily data by regions, and endogenous
88 situations, such as climatic events that influence the movements of people). Indeed, the daily
89 growth rate can have significant variability, particularly if the rate is only tracked over a short
90 period of time. Besides, the daily growth rate and its variability depend on the total number of
91 infections. At the beginning of the outbreak, the daily rates are more unstable, but they appear to
92 stabilize when the cumulative number of infections increases. The excessive variability of daily
93 growth rates adversely affects the robustness of forecasting models.

94 The average of the daily growth rate could be used to reduce this high variability in the daily growth
95 rate, as it has a smoothing effect on the rates cited.

96 This study aimed to define and evaluate the daily growth rate and apply the cumulative and
97 weighted average of the crude daily growth rate to an epidemic curve.

98

99 **Material and Methods**

100 Italian epidemic data registered between February 26, 2020 and April 18, 2020 by Italian Protezione
101 Civile [5] and Chinese (Hubei Province) epidemic data, between January 21, 2020 and March 18,
102 2020, were taken from Hubei Province government website and used in this study [6].

103 The first step was to evaluate the growth rate: given the daily infections at day t (DI_t) and the
104 cumulative infections at day t (CI_t), respectively. The daily growth rate (DGR) was calculated by:
105 DI_t/CI_{t-1} .

106 Second, the average of the DGR was evaluated. These were determined using the highest values of
107 DI_t and CI_t , weighted for CI_{t-1} , and calculated progressively, using the cumulative number of cases
108 from the second day of the series:

109

$$110 \quad (1) \quad WR_t = \frac{\sum_{i=2}^t (DGR_i * CI_{i-1})}{\sum_{i=2}^t CI_{i-1}} \quad \text{where } t > 1.$$

111

112 $DGR_t = DI_t/CI_{t-1}$ and $DRG_t * CI_{t-1} = DI_t$, (1) becomes

113

$$114 \quad (2) \quad WR_t = \frac{\sum_{i=2}^t (DGR_i * CI_{i-1})}{\sum_{i=2}^t CI_{i-1}} = \frac{\sum_{i=2}^t (DI_i)}{\sum_{i=2}^t CI_{i-1}}, \text{ where } t > 1.$$

115

116 Third, the weighted and cumulated average of the daily growth rate (WR_t) was determined by
117 analyzing the trends in Hubei, from which the results were used to model an epidemic in Italy.

118 The series of WR_t observed in Hubei approach a negative exponential model. Therefore, a
119 nonlinear regression was applied to the exponential decay model (EDM) to evaluate its reliability.

120 In this model, if a function $Y(t)$ is monotonically decreasing with a velocity parameter $r > 0$, then

121 $Y(t)$ the exponential decay model with the form:

122

123 $(3) Y(t) = a * e^{-bt}$,

124

125 where t is time in days, a is the starting value $Y(0)$, and b the decay rate; in our application, a is
126 WR_t when $t=0$.

127 To take into account the plateau, which is expected at the end of the epidemic in (3), we added a
128 term k that is the rate at the end of the curve:

129

130 $(4) Y(t) = k + (a - k) * e^{-bt}$.

131

132 When the number of cases $\rightarrow 0$ then zero, $WR_t \rightarrow k$.

133

134 It was thought that the epidemic in Italy had peaked when this model was applied. Therefore, it was
135 assumed that a conceivable k term in the exponential decay model for the Italian epidemic could be
136 the same as that observed in the Hubei epidemic that was signed as \bar{k}_H , so the model applied was:

137

138 $(5) WR_t(t) = \bar{k}_H + (WR_t(0) - \bar{k}_H) * e^{-bt}$.

139

140 To evaluate the variability of the forecast, we used credibility intervals [7], determined through a
141 first-order Taylor expansion and Monte Carlo simulation [8,9].

142 Expected occurrences, \overline{DI}_t , and expected cumulated occurrences \overline{CI}_t , were determined using the
143 inverse of (2) once estimated $WR_t(t)$.

144

145 ***Qualitative evaluation of forecast model***

146 A comparison with the Gompertz model, which is often used to estimate cumulative events, was
147 made to evaluate the fitting of the model based on the weighted and averaged growth rates. This

148 was done because the two-parameter Gompertz model (GoM) is based on exponential decay, and is
149 given by the ordinary differential equations:

$$150 \quad (6) \quad dC(t)/dt = C'(t) = rC(t) \exp(-bt),$$

151 where t is time, $C'(t)$ describes the incidence curve over time, and $C(t)$ is the cumulative number of
152 cases at

153 time t , and $b > 0$ describes the exponential decay of the growth rate r (with $r > 0$). If $C(0)$ is the
154 initial number of cases, then the solution of (6) is

$$155 \quad (7) \quad C(t) = C(0) \exp((r/b)(1-\exp(-bt))),$$

156 This model was applied to the cumulative infection data in Italy during the same period.

157 To compare the mathematical models and assess their predictive accuracy at different stages of the
158 outbreak, we estimated the two models (5 and 7) using the observed values in four-time intervals:

159 P1 February 27 to March 17, 2020; P2 February 27 to March 24, 2020; P3 February 27 to March
160 31, 2020; P4 February 27 to April 07, 2020. We used the estimated parameters at each time period
161 to determine both the rates and the cumulative cases expected, with credibility intervals until May
162 31, 2020.

163 It is possible to quantify the error of the model fit to the data using performance metrics [10]. The
164 most widely used performance metrics are the Pearson's correlation coefficient, root mean squared
165 error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), given by
166 the respective expressions:

$$167 \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\bar{y}_{t_i} - y_{t_i})^2} \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |\bar{y}_{t_i} - y_{t_i}| \quad \text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\bar{y}_{t_i} - y_{t_i}}{y_{t_i}} \right|$$

168

169 MAPE, together with the correlation coefficient, are performance indicators independent of the
170 outcome size. For this reason, they were used to compare model (5) applied to WR_t model (7)

171 applied to CI. RMSE and MAE were used to compare the same model in relation to the different

172 time phases of the estimate.

173 These four metrics are also useful for quantifying the error associated with a forecast and were used
 174 to evaluate the goodness of fit of the models and to measure the reliability of the model in
 175 predicting the values observed after the date of the estimate.

176 The analysis was carried out using R software (version 3.6.3) [11] with the packages “nls2” [12]
 177 and “propagate”[5].

178

179 Results

180 Starting from February 27, 2020 until April 07, 2020 (last day before the model fitting), the number
 181 of infections had grown from 650 to 135,606. The daily growth rate ranged from 2.3% to 62.5%,
 182 with a median of 13.4% (Inter Quartile Range IQR 6.9% to 22.8%). The WR_t ranged from 7.0%-
 183 45.4%, with a median of 17.3% (IQR, 11.3%-26.0%) (Fig 1).

184 **Fig 1.** CI, DGR, and WR_t observed and calculated in Italy during the reference period (27 February
 185 2020 - 07 April 2020).

186 Table 1 shows the cumulative cases, daily growth rates, and WR_t determined at each time point
 187 chosen for the evaluated models.

188 **Table 1.** Descriptive statistics of the CI, DGR, and WR for each time phase

Time Model (end Date)	N° Days (from 27/02)	CI at end Date (from 650)	DGR (%)		WR_t (%)	
			Range	Median [IQR]	Range	Median [IQR]
P1 (17/03)	20	31,506	10.7-62.5	22.8 [19.1-25.7]	18-45.4	26.0 [22.1-31.4]
P2 (24/03)	27	69,176	8.1-62.5	20.2 [13.6-24.6]	13.6-45.4	22.5 [17.7-28.3]
P3 (31/03)	34	105,792	4.0-62.5	15.9 [8.8-23.3]	9.5-45.4	20.7 [14.7-26.2]
P4 (07/04)	41	135,606	2.3-62.5	13.4 [6.9-22.8]	7.0-45.4	17.3 [11.3-26.0]

189 DGR, daily growth rate; CI, cumulative infections; IQR, inter quartile range; P1–P4, time intervals;
 190 WR_t , weighted and cumulated average of the daily growth rate.

191

192 To estimate \bar{k}_H the exponential decay model was applied to the Hubei Province data from January
 193 29, 2020 to March 13, 2020. The initial parameters were $k=0$, $a=0.388$, and $b=0.05$. The algorithm

194 for the estimation of the model converged in three iterations, and all parameters were statistically
 195 significant ($a = 0.406 \pm 0.006$, $p < .0001$; $b = 0.070 \pm 0.003$, $p < .0001$; $k = 0.011 \pm 0.004$, $p = 0.019$) (Fig
 196 2). The fitting of the curve was remarkably high in relation to the observed values, and the residual
 197 standard error was 0.0084.

198 **Fig 2.** CI, DGR, and WR_t observed and estimated with model 4) in Hubei Province (China)
 199 between 29 January 2020 and 13 March 2020.

200 We used $\bar{k}_H = 0.011$ in (5) to estimate WR_t models for the Italian COVID-19 outbreak in the
 201 following four-time periods: P1 February 27 to March 13, P2 February 27 to March 17, P3
 202 February 27 to March 31, and P4 March 27 to April 07, 2020. The parameter a showed a range
 203 from 0.414 to 0.415, which was statistically significant at all time intervals. The parameter b ranged
 204 between 0.046 and 0.047 and was also statistically significant. The value of RMSE decreased from
 205 0.019 for P1 to 0.014 for P4.

206 The cumulative number of cases (CI) was estimated at the same time intervals P1-P4 with the
 207 model (6): parameters decreased from 602.2 in P1 to 118.2 in P4, whereas exponential decay b and
 208 growth rate increased (Table 2). In these models, the RMSE increased from 256.7 in P1 to 106 in
 209 P2, and the MAE was opposite to the growth rate models (Table 2).

210 **Table 2.** Estimated parameters with the exponential decay model applied to WR_t Eq. (5)) and the
 211 Gompertz model applied to the CI (Eq. (7)) for each time phase

Model	N°	k	a	b			r	RMSE	MAE
	Days	Fixed = \bar{k}_H	Mean±Std.Err	Mean±Std.Err	Mean±Std.Err	Mean±Std.Err			
WR_t -P1	20	0.011	0.421±0.011 ***	0.050±0.003 ***	-	-	0.019	0.016	
WR_t -P2	27	0.011	0.414±0.009 ***	0.047±0.002 ***	-	-	0.018	0.014	
WR_t -P3	34	0.011	0.414±0.007 ***	0.046±0.002 ***	-	-	0.016	0.011	
WR_t -P4	41	0.011	0.415±0.006 ***	0.047±0.001 ***	-	-	0.014	0.010	

CI-P1	20	-	602.2±63.6 ***	0.044±0.004 ***	0.309±0.019 ***	***	256.7	192.6	

CI-P2	27	-	498.7±70.8	***	0.049±0.003	***	0.338±0.020	***	564.6	410.3
CI-P3	34	-	200.3±40.9	***	0.068±0.002	***	0.480±0.028	***	943.5	804.5
CI-P4	41	-	118.2±21.2	***	0.075±0.001	***	0.559±0.023	***	1016.3	832.2

212 *** $p < .0001$ CI, cumulative infections; P1–P4, time intervals; WR_t , weighted and cumulated

213 average of the daily growth rate; RMSE, root mean squared error; MAE, mean absolute error

214

215 The results of the models for WR_t and for CI in the four periods, P1-P4 are shown in Fig 3.

216 **Fig 3.** Fitted models based on the exponential decay model (Eq. (5); left column) and Gompertz
 217 model (Eq. (7); right column) using epidemic data for the four successive time phases P1, P2, P3,
 218 and P4, respectively. The expected values (continuous red line) and the 95% credibility interval (red
 219 dotted line) for the fitted models are displayed along with the observed data of the outbreak (empty
 220 black circles).

221 The correlation between the expected and observed values is high in all models, with the CI models
 222 having coefficients of nearly 1. The MAPE result was lower only in P1 for the CI models. In
 223 contrast, in time intervals P3 and P4, the measure of fit was higher for WR_t than for CI (Table 3a).
 224 Table 3b shows the results of the measure of fit between the forecasted expected values and the
 225 observed cases until April 21, 2020. The forecast was made with data already known at intervals
 226 P1, P2, P3, and P4, so there were four forecasts made to evaluate what could happen if the models
 227 were applied during each period. The MAPE of the exponential decay models applied to WR_t was
 228 lower than the Gompertz models for CI, and the correlation coefficient was high for all forecast
 229 models. This suggests a strong positive relationship between the expected values forecasted and the
 230 observed values.

231

232 **Table 3.** Fitting (a) and forecasting (b) performance statistics for the exponential decay model
 233 applied to WR_t (Eq. (5)) and the Gompertz model applied to the CI (Eq. (7)). For each time
 234 interval, we highlight the lowest MAPE (gray) for the size of the error.

a) Model metrics

Model	N° Days	WR		CI	
		MAPE	r_{Pers}	MAPE	r_{Pers}
P1 - 17/03	20	0.056	0.968	0.036	1.000
P2 - 24/03	27	0.051	0.978	0.058	1.000
P3 - 31/03	34	0.044	0.986	0.141	1.000
P4 - 07/04	41	0.044	0.990	0.151	1.000

b) Post-model metrics (until 21 April 2020)

Model	N° Days until 21/04	WR		CI	
		MAPE	r_{Pers}	MAPE	r_{Pers}
P1 - 17/03	35	0.075	0.998	0.371	0.984
P2 - 24/03	28	0.029	0.997	0.322	0.996
P3 - 31/03	21	0.037	0.999	0.040	0.999
P4 - 07/04	14	0.023	1.000	0.028	0.999

235 CI, cumulative infections; P1–P4, time intervals; MAPE, mean absolute percentage error ;

236 r_{Pers} , Pearson coefficient correlation.

237 The forecasts from all models are shown in Fig 4. According to the model with WR_t and the
 238 exponential decay with available data for P1, the peak of infections should be reached on April 25
 239 (credibility interval April 12 to May 3) with 131,280 cumulative infections expected (credibility
 240 interval 71,415-263,501); with the data available for P2 the peak date should be reached on May 3
 241 (April 23 to May 23) with 197,179 cumulative infections expected (130033-315,269); with data
 242 available for P3 the peak should be reached on May 4 (April 25 to May 18) with 202,210

243 cumulative infections expected (155,235-270,737); with data available for P4 the peak should be
244 reached on May 3 (April 26 to May 11) with 191,586 (160,861-232,023) cumulative infections
245 expected.

246 **Fig 4.** Epidemic forecasts based on the exponential decay model (Eq. (5); left column) and
247 Gompertz model (Eq. (7); right column) calibrated using the epidemic data for scenarios 1, 2, 3, and
248 4, respectively. The expected values (continuous red line) and the 95% credibility interval (red
249 dotted line) for the calibrated models are displayed along with the observed data of the epidemic,
250 both those used for the estimation of the models (empty black circles) and those after modeling
251 (filled black circles).

252 All models for CI based on Gompertz have forecast peak dates beyond the May 31 with cumulative
253 infections expected using the data available for P1, P2, P3 and P4 of 585,836 (227,674-1,027,058),
254 437,665 (309,556-564,412), 222,690 (190,579-243,492) and 195,472 (178,952-204,502),
255 respectively.

256

257 **Discussion**

258 Dynamic growth models provide an important quantitative framework for predicting epidemic
259 trajectories, generating estimates of key transmission parameters, assessing the impact of control
260 interventions, gaining insight into the contribution of different transmission pathways, and
261 producing short and long-term forecasts [13].

262 To forecast the evolution of the outbreak, we used phenomenological models with an empirical
263 approach, as such models are useful for reproducing patterns observed in time series data [14]. The
264 result is a reasonably simple temporal description of the epidemic growth patterns [13].

265 Observation of the epidemic curve in other locales that have already experienced the same outbreak
266 is fundamental in the application of phenomenological models. In this study, we used the epidemic
267 curve of the Chinese province of Hubei to estimate one of the parameters of the exponential decay

268 model that was then applied to the Italian data. It was necessary to find the time point at the
269 beginning of the outbreak in which the curves of the weighted and cumulative average rates were
270 superimposable between the two populations to enable this transposition. WR_t proved to be a stable
271 and useful indicator for estimating the epidemic model based on Chinese data. At the beginning of
272 the epidemic, with few cases, the Gompertz model on the cumulative cases had a better fit and
273 delivered a more precise forecast, supported by the MAPE statistics. However, this does not mean
274 that the forecasted peak date and the number of infections are necessarily correct. The Gompertz
275 model does not have a parameter related to the size of the outbreak. Using data available on 17th of
276 March, the Gompertz model overestimated both the duration and number of cases. The exponential
277 decay model on the same date underestimated the duration and number of cases, but the observed
278 cases were within the credibility limits. Wu et al. showed the same level of inaccuracy using a
279 logistic growth model, generalized logistic growth model, generalized growth model, and
280 generalized Richards model on March 10 [15]. The generalized logistic model provided the best
281 scenario, predicting that the peak would be reached in 10 days. This scenario would have led to a
282 total cumulative number of infected cases of 59,199 (95% CI: [16,796, 107,097]) at 15 days, and
283 91,719 (95% CI: [16,944, 183,440]) at the end of the pandemic[15]. In contrast, a pessimistic
284 scenario provided by the logistic model predicted that there could be up to 2 million infected cases
285 by the end of May 2020.

286 The exponential decay model applied to the weighted and averaged growth rates WR , seemed to
287 perform better than that shown by Fanelli and Piazza on the 15th of March [16]. They analyzed the
288 temporal dynamics of the COVID-19 outbreak in China, Italy, and France between January 22 and
289 March 15, 2020 and applied a susceptible-infected-recovered-deaths model. They estimated the
290 peak in Italy to be about March 21, 2020, forecasting a maximum number of 26,000 cases at the
291 epidemic peak.

292 Using data until 24th March, the Gompertz model overestimated the cumulative number of cases.
293 The expected cases in the exponential decay model with *WR* were close to those observed, with a
294 low value of MAPE (0.029).
295 When the exponential model was applied to data until the 31st of March, the cases forecast was
296 202,690, whereas the Gompertz model forecast more than 222,690 cases.
297 Chintalapudi et al. [17] used an ARIMA model, assuming that the lockdown would last for 60 days.
298 According to this study, the 60-day forecast of infected cases could increase to 105,732-182,757.
299 The MAPE for the ARIMA model was 6.25, which is higher compared to the exponential decay
300 model with *WR*.
301 The exponential model with *WR*, fitted with data available between February 27 and April 07
302 appears to be close to the observed data, and the MAPE shows a very good fit, with a lower
303 difference between the expected and observed cases than those in the previous model. In contrast,
304 the Gompertz model has a sudden deviation from the observed line and overestimated the number
305 of cases and duration.
306 It is still unclear what effect the number of swabs performed has on the reported number of positive
307 subjects in the last week of the epidemic compared to at the beginning, and this could significantly
308 affect any modeling of the COVID-19 outbreak.

309

310 **Conclusion**

311 An exponential decay model applied to the growth rate of infections weighted with *WR* appears to
312 be useful in estimating the number of cases and peak of the COVID-19 outbreak in Italy.
313 The model was more reliable in the exponential growth phase due to the statistics used in the model.
314 The epidemic trend information from the COVID-19 outbreak in Hubei Province, China, proved to
315 be essential for estimating the parameters to use in the Italian model.

316

317 **Acknowledgments:** We would like to thank Editage (www.editage.com) for English language
318 editing

319 **References**

320 1) World Health Organization. Novel Coronavirus (2019-nCoV) Situation Reports 103. Available
321 online from: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>
322 (accessed on 30 April 2020).

323 2) Sylvers E, Legorano G. (9 March 2020). "As Virus Spreads, Italy Locks Down Country". Wall
324 Street Journal. Retrieved 10 March 2020. Italy is the world's first country to place its entire territory
325 under quarantine.

326 3) Mizumoto K, Kagaya K, Chowell G. Early epidemiological assessment of the transmission
327 potential and virulence of 2019 Novel Coronavirus in Wuhan City: China, 2019–2020. medRxiv
328 2020.

329 4) Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. The Novel Coronavirus,
330 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. *Popul. Evol.* 2020.

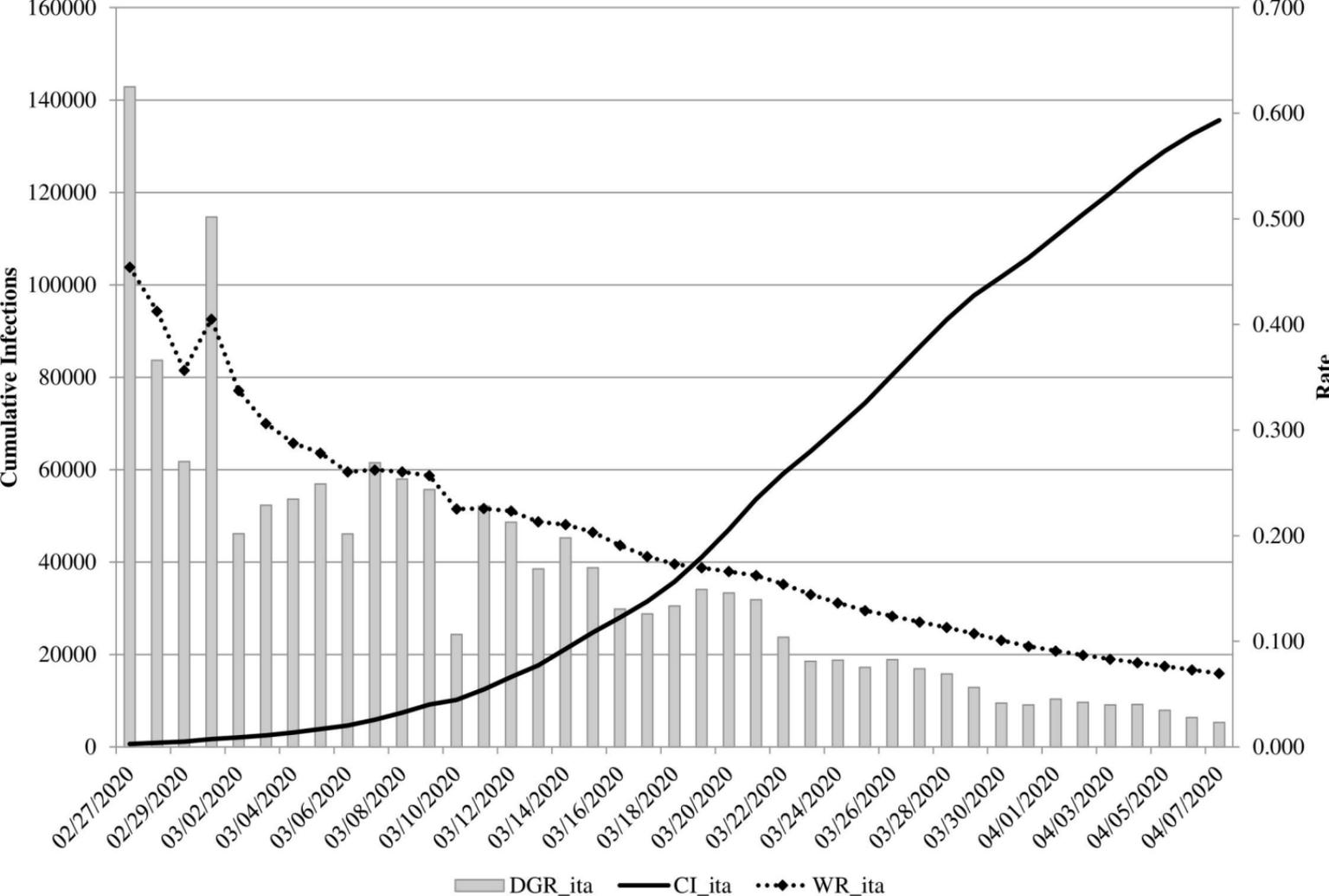
331 5) Protezione Civile Italiana. Coronavirus Emergency - Situation Map. Available online from:
332 [http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d](http://opendatadpc.maps.arcgis.com/apps/opsdashboard/index.html#/b0c68bce2cce478eaac82fe38d4138b1)
333 4138b1 (last access 21 April 2020).

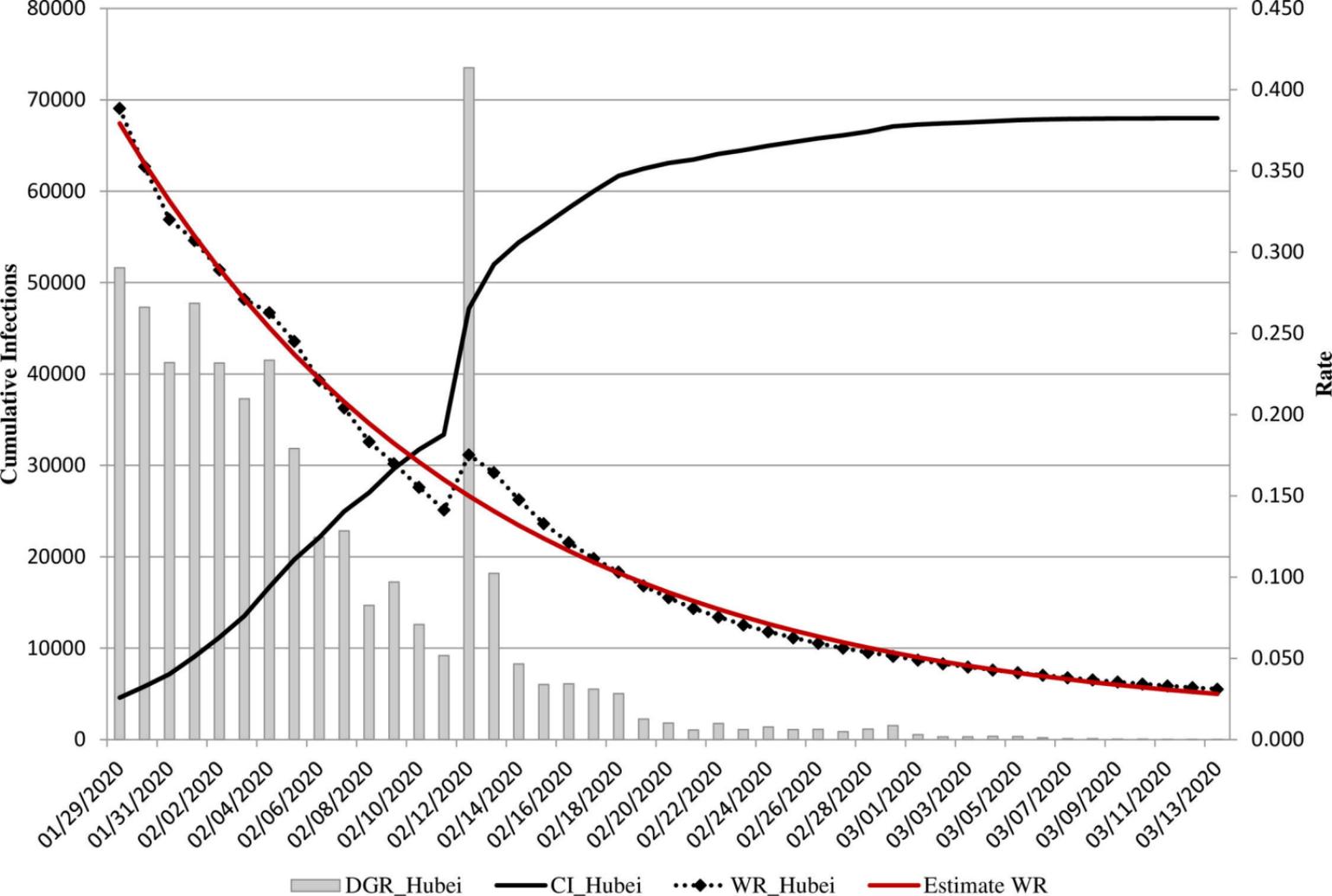
334 6) Hubei provincial government website. Authority bulletin. Available online from:
335 <http://www.hubei.gov.cn/zhuanti/2020/gzxxgzbd/zxtb/> (last access 18 April 2020).

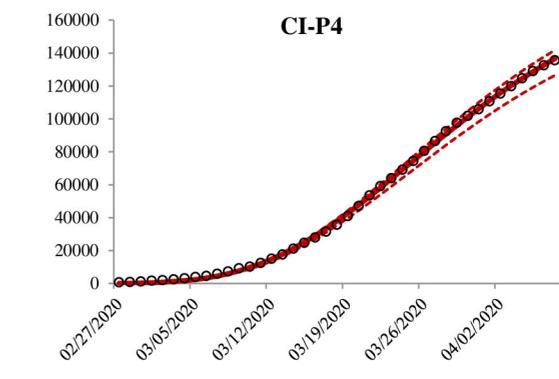
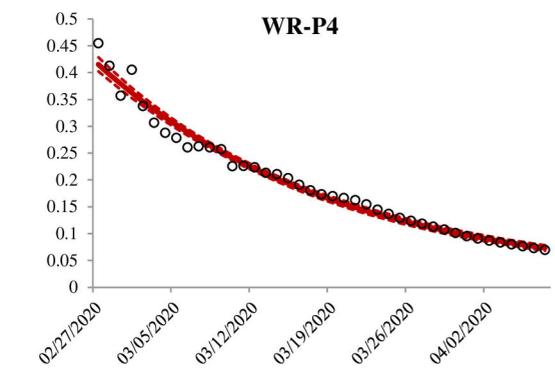
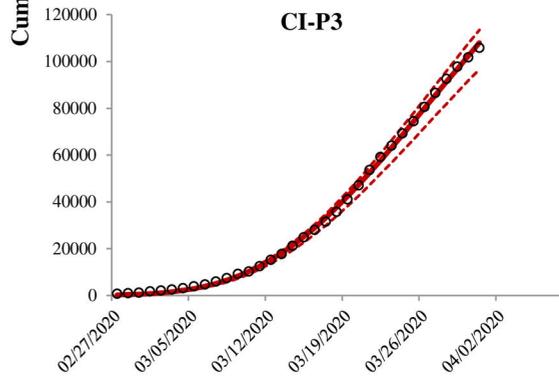
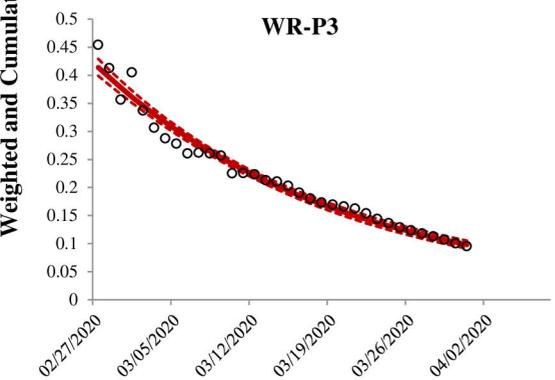
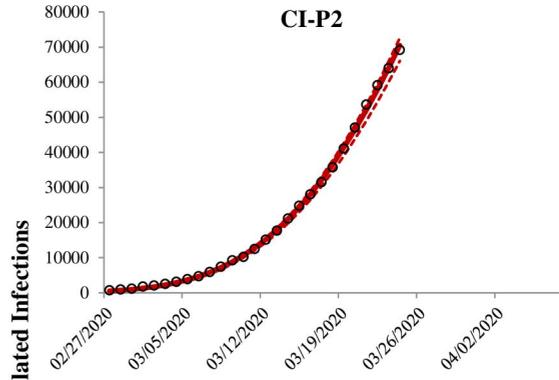
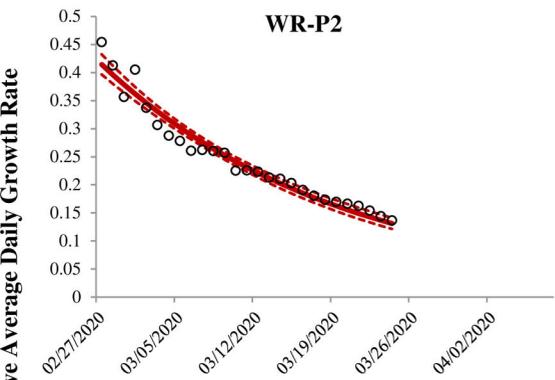
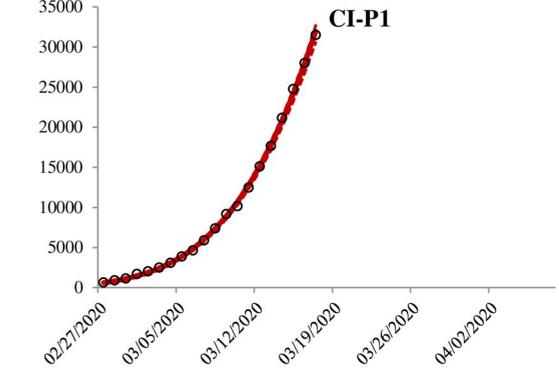
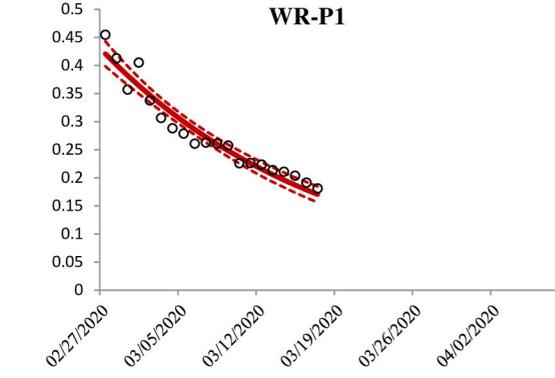
336 7) Andrej-Nikolai Spiess (2018). propagate: Propagation of Uncertainty. R package version 1.0-6.
337 <https://CRAN.R-project.org/package=propagate>.

338 8) Mekid S, Vaja D. Propagation of uncertainty: Expressions of second and third order uncertainty
339 with third and fourth moments.. *Measurement* (2008), 41: 600-609.

- 340 9) Wang CM, Iyer HK. On higher-order corrections for propagating uncertainties. *Metrologia*
341 (2005), 42: 406-410.
- 342 10) Kuhn M, Johnson K. *Applied Predictive Modeling*, Springer, New York, 2013.
- 343 11) R Core Team (2020). R: A language and environment for statistical computing. R Foundation
344 for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- 345 12) Grothendieck G. nls2: Non-linear regression with brute force. R package version 0.2. (2013).
346 <https://CRAN.R-project.org/package=nls2>
- 347 13) Chowell G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A
348 primer for parameter uncertainty, identifiability, and forecast, *Infect. Disease Model.*, 2 (2017),
349 379–398.
- 350 14) Viboud C, Simonsen L, Chowell G. A generalized-growth model to characterize the early
351 ascending phase of infectious disease outbreaks, *Epidemics*, 15 (2016), 27–37.
- 352 15) Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19
353 outbreak in 29 provinces in China and in the rest of the world. medRxiv 2020.03.11.20034363 (pre-
354 print); doi: <https://doi.org/10.1101/2020.03.11.20034363>
- 355 16) Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France.
356 *Chaos Solitons Fractals*. 2020 May;134:109761. doi: 10.1016/j.chaos.2020.109761. Epub 2020 Mar
357 21.
- 358 17) Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and
359 recovered cases after sixty day lockdown in Italy: A data driven model approach. *J Microbiol*
360 *Immunol Infect.* 2020 Apr 13. pii: S1684-1182(20)30098-0. doi: 10.1016/j.jmii.2020.04.004. [Epub
361 ahead of print]







○ Observed WaCADGR
— Predicted WaCADGR
- - - CL 95%

○ Observed CI
— Predicted CI
- - - CL 95%

