An improved mathematical prediction of the time evolution of the Covid-19 Pandemic in Italy, with Monte Carlo simulations and error analyses

Ignazio Ciufolini[1] and Antonio Paolozzi[2]

1. Dipartimento di Ingegneria dell'Innovazione, University of Salento, Lecce and Centro Fermi, Rome, Italy
2. Scuola di Ingegneria Aerospaziale, Sapienza, University of Rome

Abstract

Here we present an improved mathematical analysis of the time evolution of the Covid-19 pandemic in Italy and a statistical error analyses of its evolution, including Monte Carlo simulations with a very large number of runs to evaluate the uncertainties in its evolution. A previous analysis was based on the assumption that the number of nasopharyngeal swabs would be constant, however the number of daily swabs has been increasing with an average factor of about five with respect to our previous analysis, Therefore, here we consider the time evolution of the ratio of the diagnosed positive cases to number of swabs, which is more representative of the evolution of the pandemic when the number of swabs is increasing or changing in time. We consider a number of possible distributions representing the evolution of the pandemic in Italy and we test their prediction capability over a period of up to four weeks. The results show that a distribution of the type of Planck's black body radiation law provides very good forecasting. The use of different distributions provides an independent possible estimate of the uncertainty. We then consider five possible cases for the number of daily swabs and we estimate the potential dates of a substantial reduction in the number of diagnosed positive cases. We then estimate the spread in a substantial reduction, below a certain threshold, of the daily cases per swab among the Italian regions. We finally perform Monte Carlo simulations with 25000 runs to evaluate a random uncertainty in the prediction of the date of a substantial reduction in the number of diagnosed daily cases per swab.

1. Introduction

In a previous paper, we estimated the possible dates of a substantial reduction in the daily number of diagnosed positive cases of the Covid-19 pandemic based on the assumption that the number of nasopharyngeal swabs would remain roughly constant[1,2]. At the time of our previous analysis (March 26), the average daily number of swabs from February 15 was about 9000 per day. However, from March 27 up to April 25, the average number of daily swabs was about 45000 (recently on May 1, the number of daily swabs reached the maximum of 74208). Therefore, to study the evolution of the Covid-19 pandemic, we have to consider the analysis of the ratio of daily diagnosed positive cases per swab. To possibly mathematically predict the evolution of the pandemic in Italy we can fit the ratio of cases per swab using several different distributions, however some distributions are more suitable than others for forecasting the future behavior of the pandemic. We consider the following distributions: Gauss, Beta, Gamma, Weibull, Lognormal and two of the type of the Planck's black body radiation law. The number of parameters chosen in some distributions is between two and three. It turns out that Planck's law types distribution with three independent parameters provides the best predictions.

Furthermore, since the number of daily swabs depends on factors that are unknown to us, such as the daily availability of reagents and specialized personnel, we consider five possible cases for the daily number of swabs, we have also assumed some possible time evolution in the number of daily swabs which for brevity we do not report here. We fit the time evolution of the positive cases per unit swab up to April 25, using the two best-fit-prediction distributions, i.e., Planck with three parameters and Gamma, along with a Gauss distribution. After analyzing the time evolution of the positive cases per unit swab using these three distributions and five conceivable number of daily swabs, we estimate the evolution in the number of diagnosed positive cases and the dates of a substantial reduction in such a daily number. Furthermore, in section 3.1 we estimate the spread in the dates of a substantial reduction in the number of daily cases per swab among the regions of Italy, where the conditions are quite different from each other, including the number of swabs per person.

The different distributions that we use provide a possible independent way to estimate the uncertainty. Indeed, a basic problem is to mathematically estimate the uncertainty in the date of a substantial reduction of daily cases. For the purpose of estimating the random uncertainties, in section 3 we report the results of a Monte Carlo simulation with 25000 runs.

2 A mathematical analysis of the ratio of daily positive cases per swab

After analyzing the time trend of the ratio of daily positive cases to the number of daily swabs[3-5], we found[1,2] that this trend can be modeled by a Gauss distribution, however the time trend has also a certain amount of skewness that can better be fitted by choosing a skewed distribution such as the Weibull, Log-normal, Beta and Gamma distributions, and also other distributions of the type of the Planck's black body law. This last one, reported in Fig. 1, shows the best prediction capabilities, with three parameters *a* and *b* and *c*:

$$\frac{a\,t^b}{e^{c\,t}-1}$$

where *t* is the time. In Figs. 2a to 2e are reported the fits of the data with a distributions of the type of the Planck's black body law with two parameters (i.e., with the exponent of the time variable *t* equal to 3 as in the Planck's black body law), and with distributions of the type of the Gamma, Beta, Weibull, Lognormal respectively. The data can also be approximated by a function of the type of a Gauss function with three parameters *a, b* and *c*:

$$ae^{-\frac{(t-b)^2}{c}}$$

as shown in Fig. 2f.

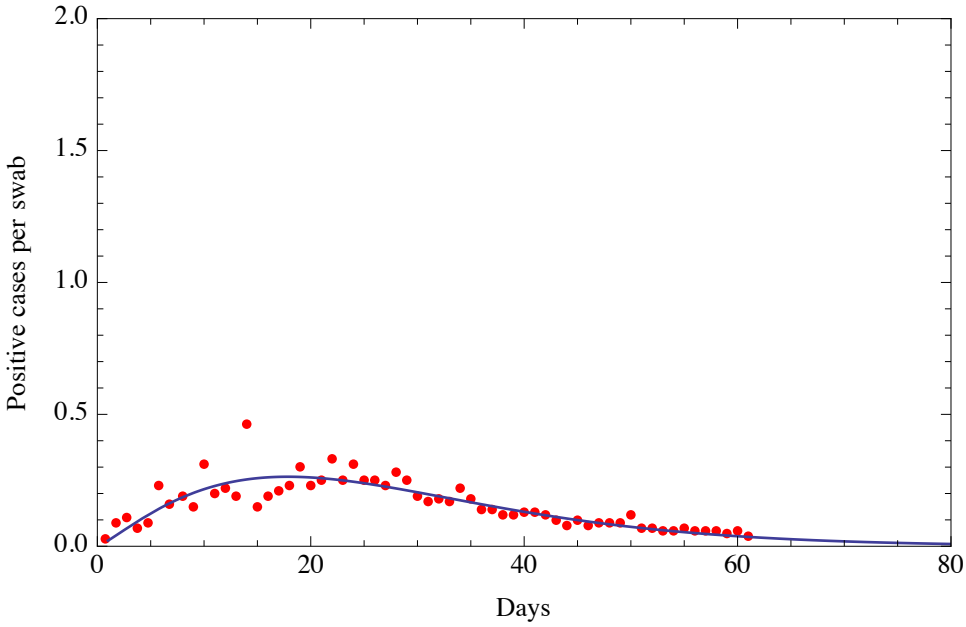


Fig. 1 Fit of the ratio of daily positive cases per swab obtained with a function of the type of the Planck's black body law with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.0446

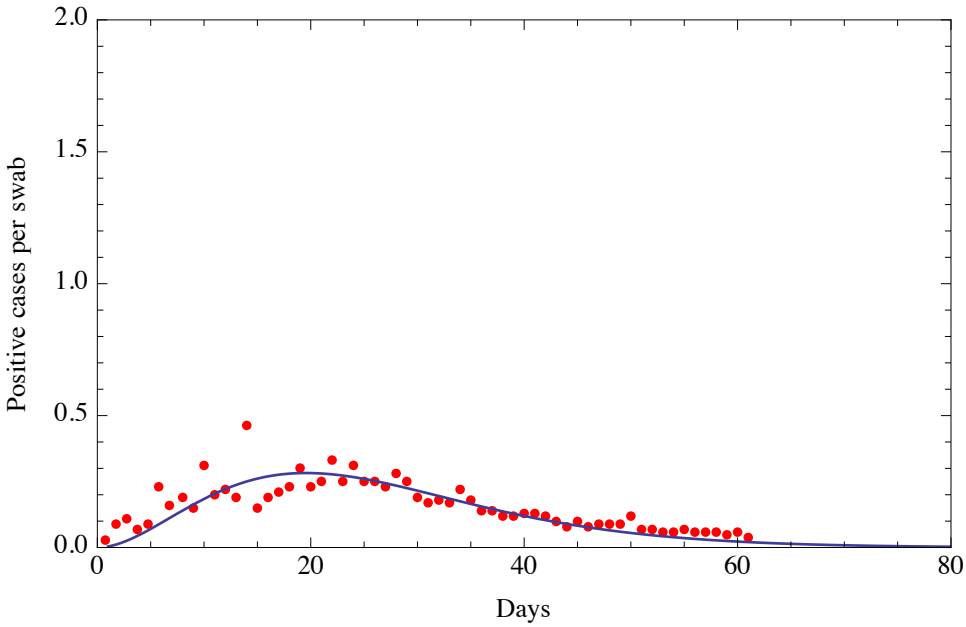


Fig. Fit of the ratio of daily positive cases per swab obtained with a function of the type of the Planck's black body law with two parameters, with the exponent of the time variable *t* equal to 3 as in the Planck's law. The beginning date is February 25. Root Mean Square of the residuals is 0.050

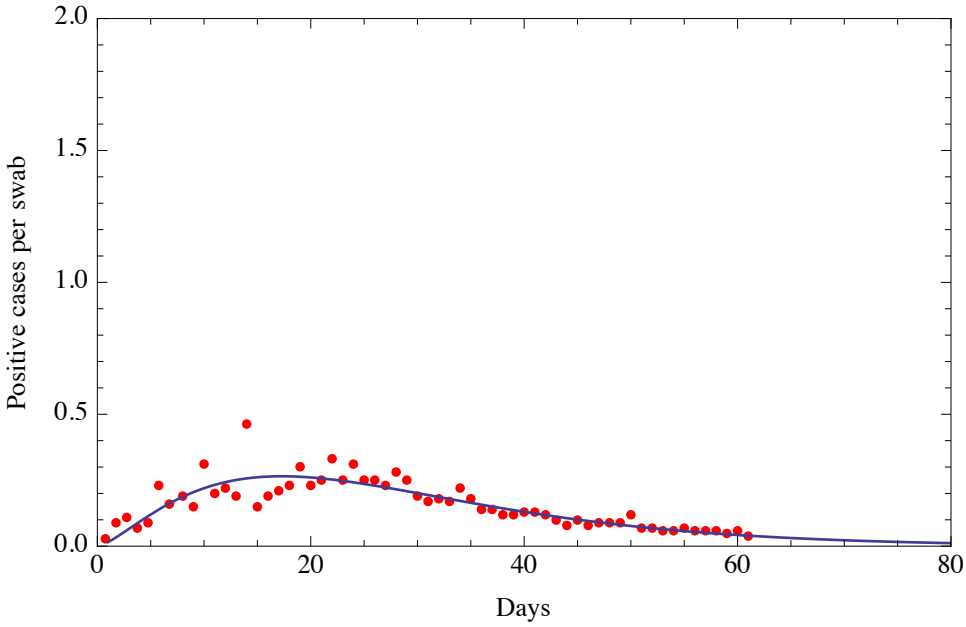


Fig. 2b Fit of the ratio of daily positive cases per swab obtained with a Gamma distribution with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.0450

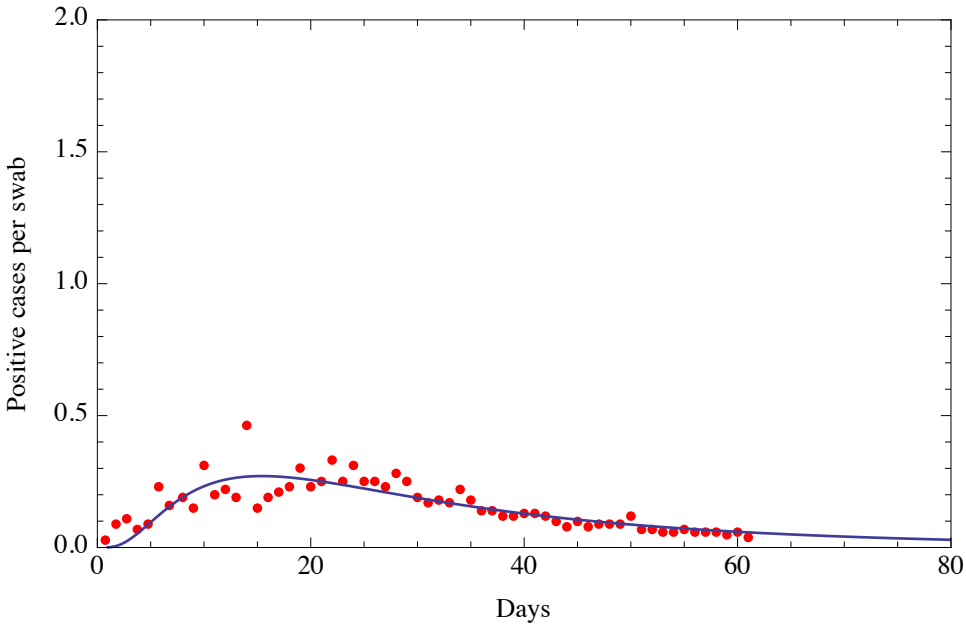


Fig. 2c Fit of the ratio of daily positive cases per swab obtained with a Beta distribution with two parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.046

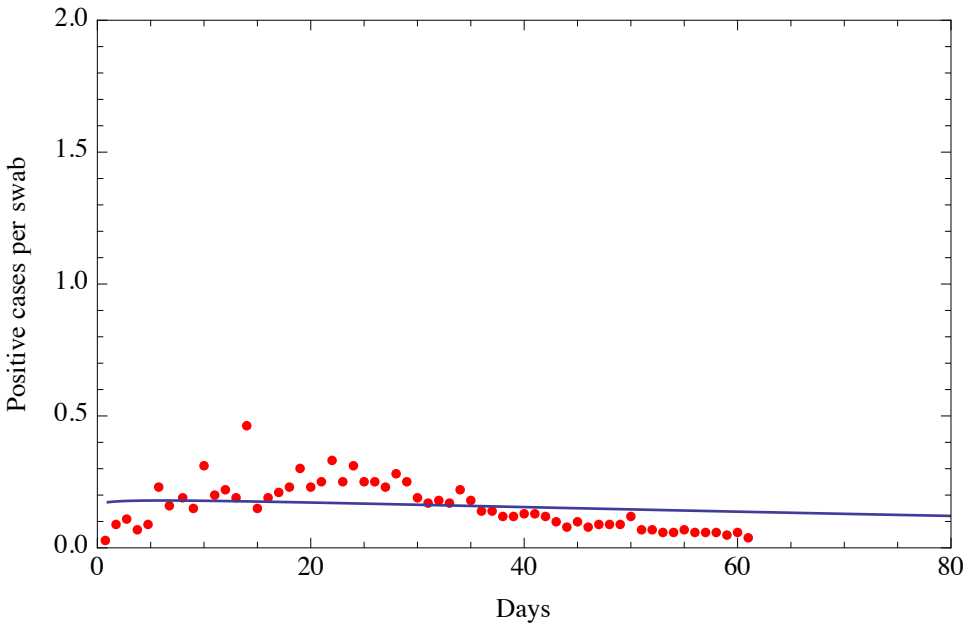


Fig. 2d Fit of the ratio of daily positive cases per swab obtained with a Weibull distribution with two parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.082

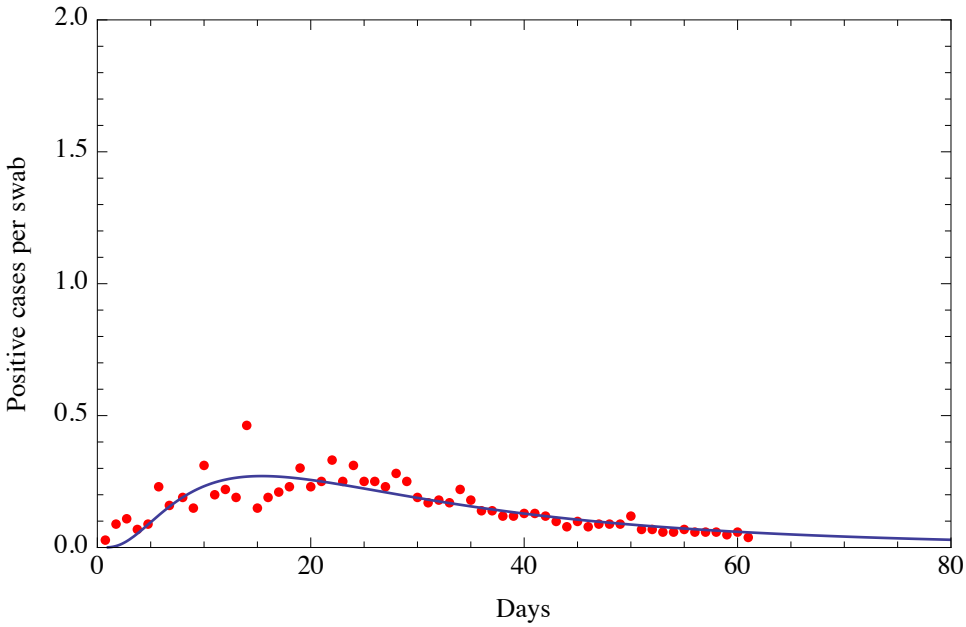


Fig. 2e Fit of the ratio of daily positive cases per unit swab obtained with a Lognormal distribution with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.049

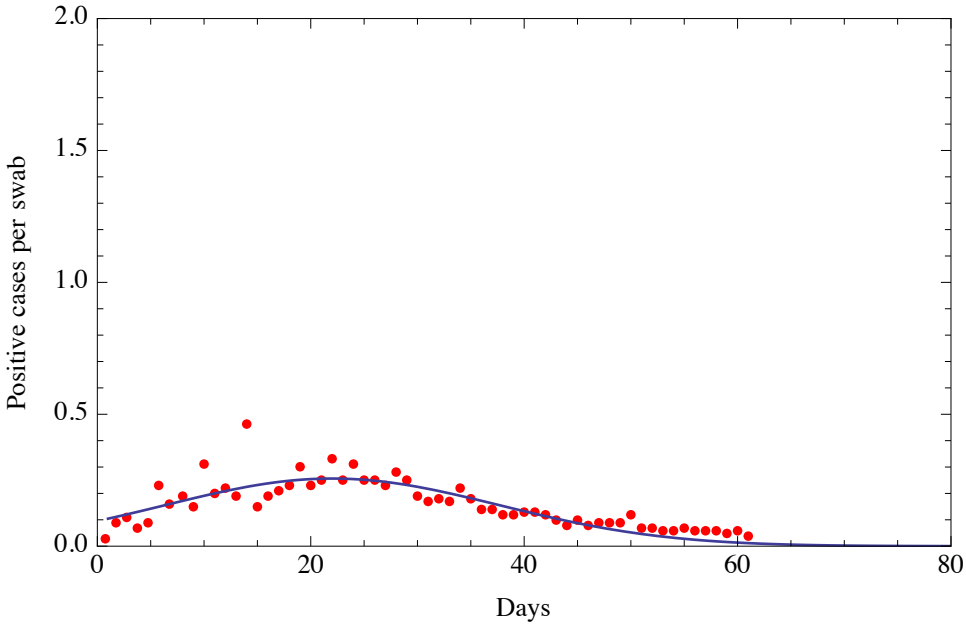
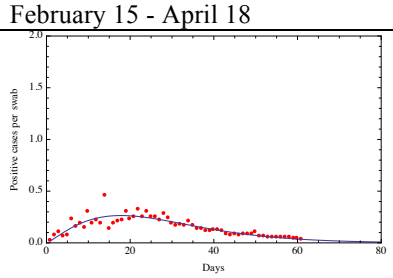
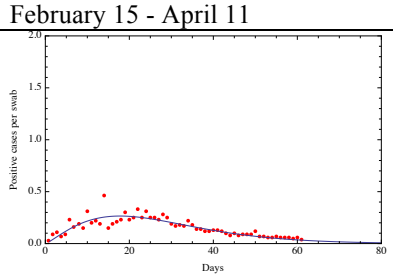
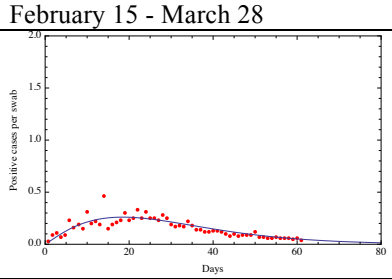
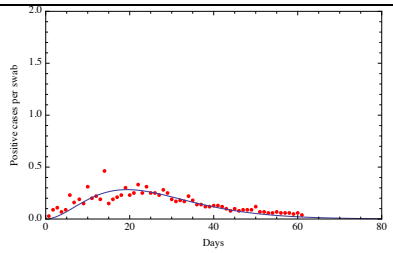
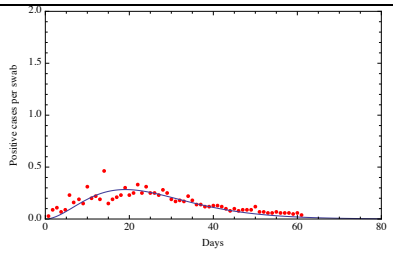
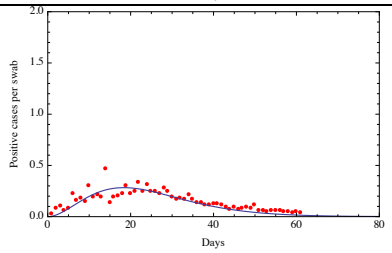
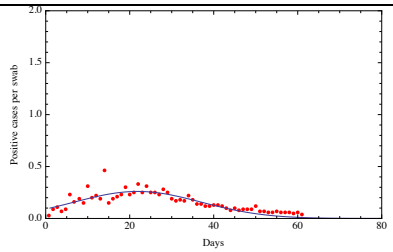
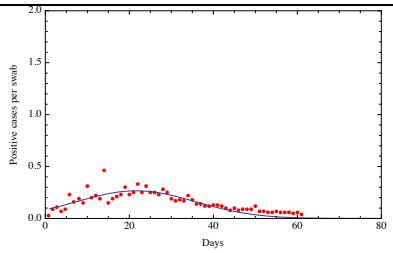
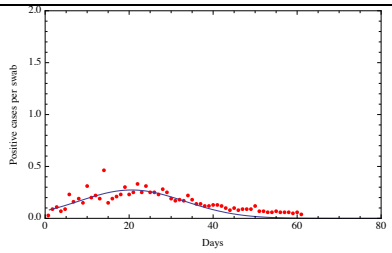
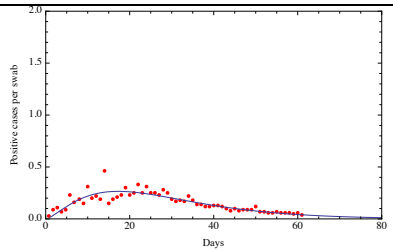
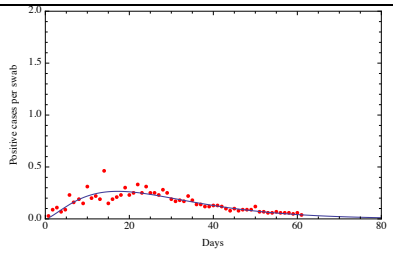
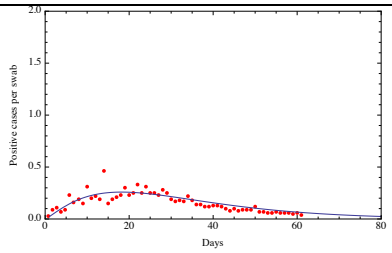
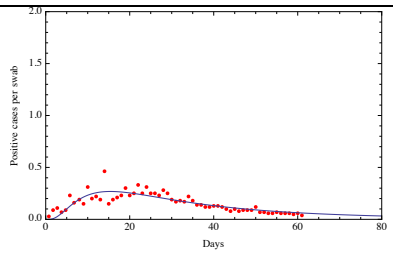
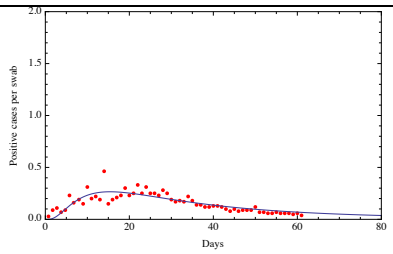
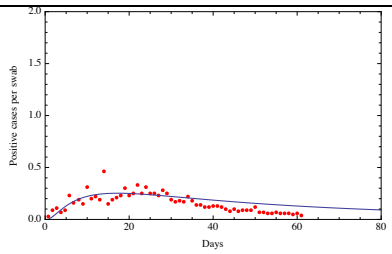
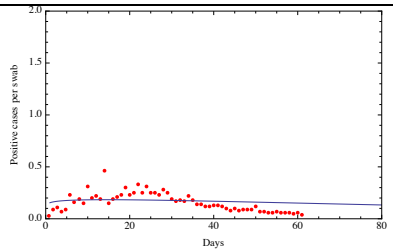
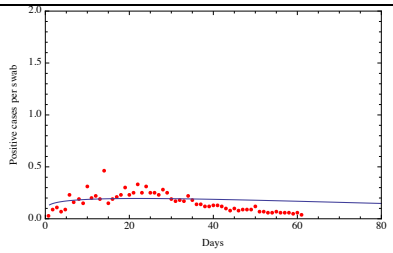
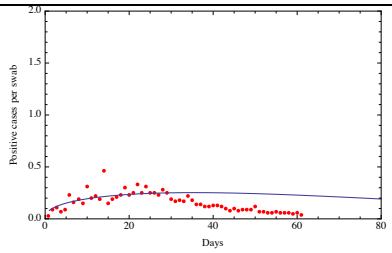


Fig. 2f Fit of the ratio of daily positive cases per swab obtained with function of the type of a Gauss function with three parameters. The beginning date is February 25. Root Mean Square of the residuals is 0.049

2.1 Prediction capabilities of the distributions fitting the data

As a relevant test to select the distribution which is more suitable to predict the evolution in time of the daily positive cases per swab, we consider for each of the above distributions, three different intervals of daily ratios: from February 25 until April 18, from February 25 until April 11 and from February 25 until March 28, i.e., one week before the date of April 25, two weeks before April 25 and four weeks before April 25, respectively. We then test which one gives the best predictions for the measured ratios during the last five days (from April 21 until April 25) and the lowest absolute value of the mean and Root Mean Square (RMS) of the residuals (difference between the measured daily data and the fitting function) up to April 25, from April 18, April 11 and March 28, respectively. As shown in Tables 1 and 2, over a

long period of about one month the distribution providing the best prediction is of the type of the Planck's black body law with three parameters. Over a short period of up to two weeks the Gamma distribution shows a slightly better prediction behavior but less stable than the Plank's law with three parameters. The Beta and Plank with two parameters are also distributions with good prediction capability. The Weibull distribution does not fit well the data.

Distrib. ↓	Fit Interval→data of the whole period minus the last 1 week, i.e.: February 15 - April 18	Fit Interval→data of the whole period minus the last 2 weeks, i.e.: February 15 - April 11	Fit Interval→data of the whole period minus the last 4 weeks, i.e.: February 15 - March 28
Planck 3 param.			
Planck 2 param.			
Gauss 3 param.			
Gamma 3 param.			
Beta 2 param.			
Weibull 2 param.			

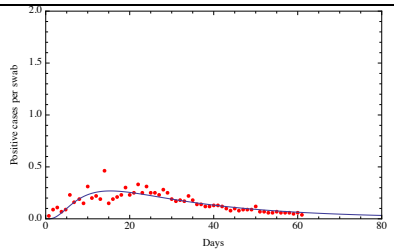
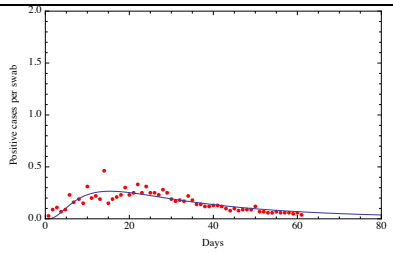
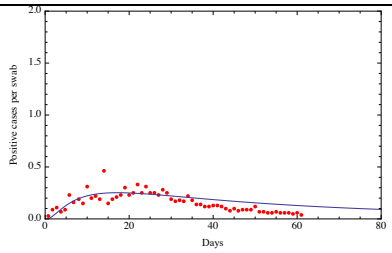
Log Normal 3 param.			

Table 1. Plots of the seven distributions (rows) obtained by fitting the daily ratios over three different time intervals (columns): from February 25 up to one week before the date of April 25, two weeks before April 25 and four weeks before April 25, respectively.

Distribution ↓	Fit up to April 25 minus the last 7 days. Mean and RMS between April 21 and April 25	Fit up to April 25 minus the last 14 days. Mean and RMS between April 21 and April 25	Fit up to April 25 minus the last 28 days. Mean and RMS between April 21 and April 25
Planck 2 par.	-0.022 0.0055	-0.024 0.0056	-0.029 0.0059
Planck 3 par.	-0.0061 0.0052	-0.0078 0.0052	0.0087 0.0050
Gauss 3 par.	-0.033 0.0057	-0.038 0.0061	-0.043 0.007
Gamma 3 par.	-0.0011 0.0052	-0.0016 0.0052	0.023 0.0050
Beta 2 par	0.0096 0.0050	0.011 0.0050	0.013 0.0050
Weibull 2 par	0.106 0.0067	0.124 0.0066	0.184 0.0060
Lognormal 3 par	0.0199 0.0055	0.026 0.0054	0.085 0.0054

Table 2. Mean and Root Mean Square of the residuals, calculated over the last five days, i.e. from April 21 to April 25, corresponding to the seven distributions fitted over the three different periods specified in the three columns.

Distribution ↓	Mean and RMS between April 18 and April 25	Mean and RMS between April 11 and April 25	Mean and RMS between March 28 and April 25
Planck 2 par.	-0.022 0.0046	-0.026 0.011	-0.026 0.015
Planck 3 par.	-0.0056 0.0044	-0.0073 0.011	0.018 0.016
Gauss 3 par.	-0.033 0.0047	-0.0398 0.011	-0.043 0.019
Gamma 3 par.	-0.00065 0.0044	-0.0014 0.011	0.029 0.016
Beta 2 par	0.011 0.0046	0.013 0.011	0.022 0.016
Weibull 2 par	0.10 0.0065	0.112 0.017	0.15 0.038
Lognormal 3 par	0.0195 0.0045	0.023 0.012	0.072 0.023

Table 3. Mean and Root Mean Square of the residuals calculated over the last week, the last two weeks and the last four weeks (i.e., from April 18, April 11 and March 28 up to April 25), corresponding to the seven fitting distributions.

3. Analysis of daily diagnosed cases per swab of Covid-19 in Italy and Monte Carlo simulations

In the present section, we first analyze the expected dates of a substantial reduction in the ratio of the daily diagnosed cases per swab in Italy and we then perform a Monte Carlo simulation with 25000 runs to evaluate the random uncertainties in these dates.

Using the best fitting distribution, i.e., a Planck-type distribution with three parameters (see previous section 2) and Gamma-type and Gauss-type distributions, we find the dates when we expect that the ratio of daily diagnosed cases per swab will be below certain thresholds. We choose these thresholds to be equal to the minimum measured ratio of cases per swab (in the range February 25 to April 30), equal to 0.02163, incidentally corresponding to the first available datum of such ratio occurred on February 25; one fifth of it and one tenth of it. These dates are summarized in Table 4.

	Threshold	Planck 3 param.	Gamma 3 param.	Gauss 3 param.
Historical minimum from February 25 to April 30	0,02163	68 (May 2)	71 (May 5)	57 (April 21)
5 times smaller than the historical minimum	0,004326	89 (May 23)	94 (May 28)	67 (May 1)
10 times smaller than the historical minimum	0,002163	97 (May 31)	104 (June 7)	70 (May 4)

Table 4. Dates of a substantial reduction in the ratio of cases per swab according to the three fitting distributions: Planck-type with three parameters, Gamma-type and Gauss-type.

However, several uncertainties can influence the number of the diagnosed positive cases of the Covid-19 pandemic, in addition to the number of nasopharyngeal swabs that have increased with time as explained in the previous section. Another source of uncertainty is related to the fact that the number of swabs is associated both to new cases and to the repeated ones of persons already tested as positive. However if we assume that the percentage of re-tested people is approximately constant such error should be small.

To possibly estimate the uncertainties in the number of positive cases, we use a Monte Carlo simulations[7-9] with 25000 runs, similarly to what done in previous works[1,2], and the differences in the predictions among the best distributions considered previously. The Monte Carlo simulation should account mainly for random uncertainties. The uncertainty we consider in the Monte Carlo simulation does not take into account the difference between the real number of cases (which is unknown) and the daily diagnosed ones which can be one order of magnitude higher, or even more, than the actual cases. However, it is usual in statistics to use a sample as being representative of the population under study and the present Monte Carlo simulation is performed for the number of diagnosed daily cases per swab.

For convenience to the reader, we summarize here the procedure used previously[1,2] for the total number of cases where instead here is applied to the ratio of cases per swab, furthermore the number of runs have been largely increased from 150 to 25000. We assume a measurement uncertainty in the daily number of diagnosed cases per swab equal to 20% of each daily ratio (Gaussian distributed). Then, a random matrix $(m \times n)$ is generated, where n (columns) is the number of observed days and m (rows) is the number of random outcomes, which we have chosen to be 25000. Each number in the matrix is part of a Gaussian distribution with mean equal to 1 and sigma equal to 0.2 (i.e., 20% of 1), either row-wise and column-wise. So, starting from the n nominal values of the daily data per swab, we generated n Gaussian distributions with 25000 outcomes, with mean equal to the n nominal values and with 20% standard deviation. Then, for each of the 25000 simulations, those n values (corresponding to the daily cases per swab of n days) are fitted with a three parameter function of the type of a Planck function (see section 2) and we then determine the date at which the number of daily positive cases will be less than a certain threshold that, for example, we choose to be equal to the minimum measured ratio of cases per swab (in the range February 25 to April 30), equal to 0.02163, one fifth of it and one tenth of it.

Finally, we calculate the mean and the standard deviation of the 25000 simulations for these three cases, the results are reported in Table 5.

		Planck type distribution with three paramenters	
Threshold description	Threshold	Mean value of day number and corresponding date	Standard deviation
Historical minimum from February 25 to April 30	0,02163	68 (May 2)	1,8
5 times smaller than the historical minimum	0,004326	89 (May 23)	2,9
10 times smaller than the historical minimum	0,002163	97 (May 31)	3,3

Table 5. Results of the Monte Carlo simulation with 25000 runs.

In Fig. 3, is reported the histogram of the frequencies versus the day of a substantial reduction in the number of daily cases per swab corresponding to less than 0.004326, i.e., to less than one fifth of the minimum value of cases per swab measured during the period February 25 to April 30. The histogram approaches a Gaussian with mean equal to day 89, approximately corresponding to what reported in Fig. 1. The standard deviation is approximately 3 days.

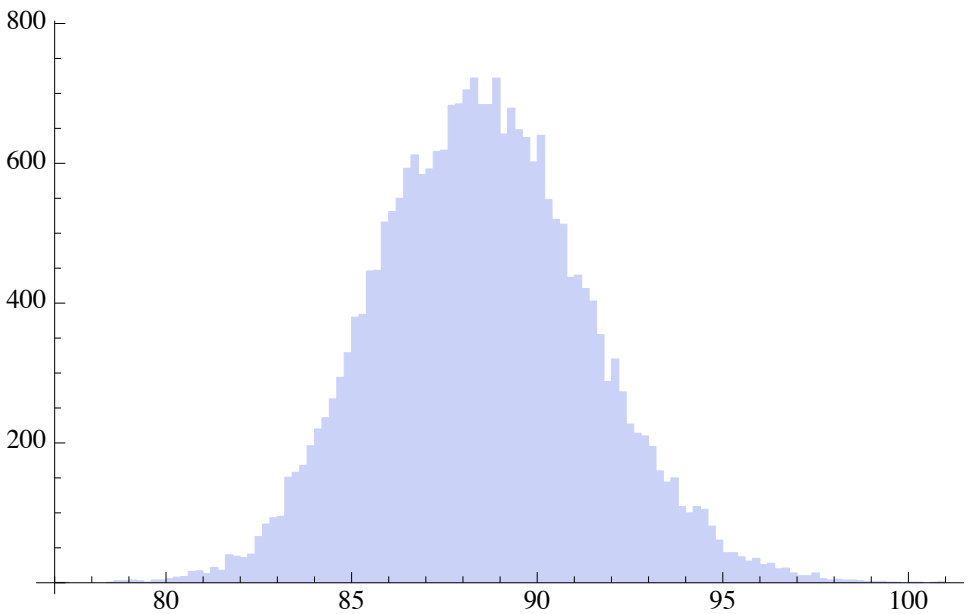


Fig. 3. Histogram of the 25000 runs of the Monte Carlo simulation for Italy: frequencies versus the day in which a substantial reduction in the daily cases per swab is less than 0.004326, i.e., less than one fifth of the minimum value of cases per swab measured during the period February 25 to April 30.

3.1 Analysis of each region of Italy

We independently analyze each of the twenty Italian regions from February 25, 2020 (included) until April 25, 2020 (included). As in the previous sections also here we consider the ratio of daily cases over the number of daily swabs. Indeed, the number of daily swabs and other relevant conditions vary quite differently from one region to the other so that we can get an indication of the spread between different regions. To calculate such spread, we evaluate the date of the reduction of the daily cases per swab in each region below a certain threshold which is chosen to be the minimum value of cases per swab in each region in the range under study (25 February- 25 April). We then fit the daily ratios of each region using a function of the type of the Planck's law with three parameters and finally, for each region, we obtain the date at which the number of daily cases per swab reduces below the given threshold for each region. In Fig. 4 we report the 20 dates corresponding to the 20 regions. The calculation of the spread of the 20 dates provides a 1-sigma standard deviation of 11.4 days.

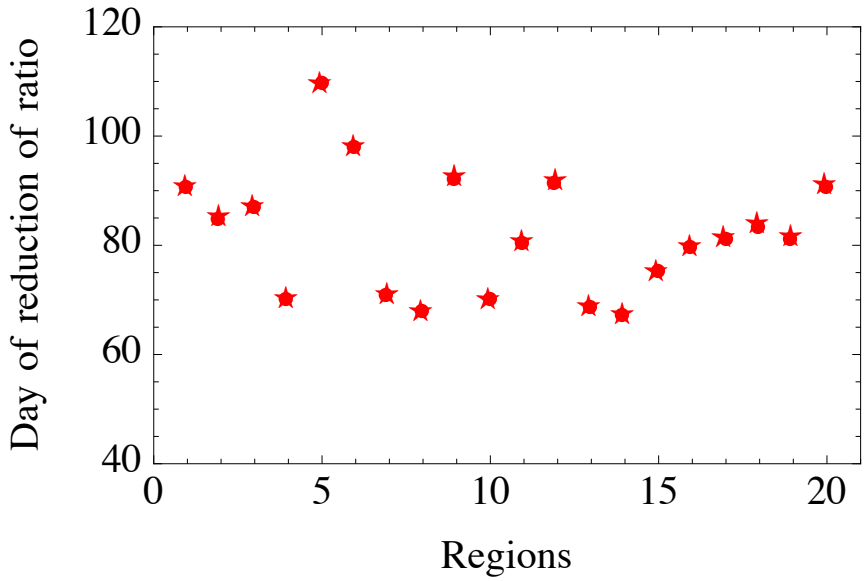


Fig. 4 Days of a substantial reduction in the number of daily cases per swab for each of the 20 Italian regions. The 1-sigma standard deviation is 11.4 days

This result may also be displayed using the Central Limit Theorem. We choose a sample of 30 regions, with repetition, out of the 20 regions, for 1200000 times. We then calculate the mean of each sample and we report the frequencies of the mean of the 1200000 samples in the histogram of Fig. 5. We obtain a Gaussian with a standard deviation of 2.03 days, that, multiplied for the square root of 30, gives back a standard deviation of approximately 11.4 days as shown in Fig. 4. This represents the fact that each region has quite different conditions for what regards the intensity and phase of the pandemic, the number of daily swabs and other factors.

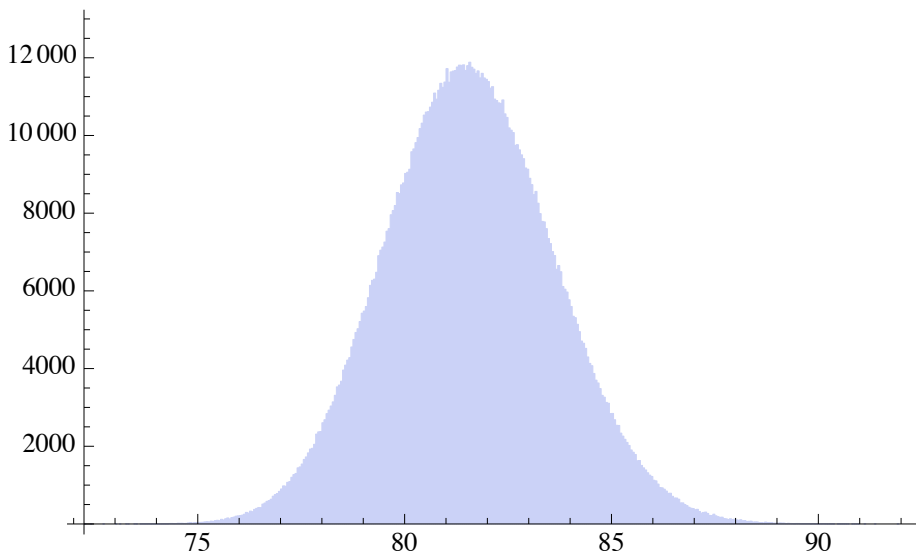


Fig. 5 Result of 1200000 simulations using the 20 Italian regions. The figure is the histogram of the mean of each sample of the regions. The standard deviation is 11.13

4. Modelling the daily swabs

Since the actual number of positive cases is much higher than the measured ones[6], by increasing the number of daily swabs (which depends on factors that are difficult to predict, such as the daily availability of reagents and specialized personnel), the number of diagnosed positive daily cases would also increase. Therefore, since the number of daily swabs is significantly changing in time, the ratio of daily positive cases per swab is more meaningful than the number

of daily cases. Nevertheless, in this section we consider five possible cases for the daily number of swabs corresponding to some relevant situations. We also tried to model the number of daily swabs using a Gaussian distribution, a distribution of the type of the Planck's law and a linear monotonic increasing distribution, however for the sake of brevity we do not report this part here. The five cases we consider are: 9000, i.e., about the number of daily swabs equal to the mean of daily swabs between February 15 and March 26 (the date of our first analysis[1,2]); 24000, i.e., about the mean between February 15 and April 25; 45000, i.e., about the mean between March 26 and April 25; 67000, i.e., about the maximum number of swabs up to April 25 and 100000, i.e., an estimated possible upper bound to the number of daily swabs. In Table 6 are reported those five cases for distributions of the type of Planck's law and the Gamma one, i.e., the best distributions for what regards the prediction capability as reported in Tables 1-3, and for a Gauss distribution as a lower bound for the predicted dates of a substantial reduction in the number of daily cases. In the case of the Planck-type distribution, we have also obtained the result of Table 6, using a Monte Carlo simulation with 25000 runs, as described in the previous section 3, with the same mean (corresponding to May 23) with a standard deviation of 2.25.

Normalized threshold for the number of daily cases	Distribution→ Daily swabs ↓	Day number and (date) with Gauss 3 parameters	Day number and (date) with Planck 3 parameters	Day number and (date) with Gamma 3 parameters
100 cases	**9000** mean up to 26 March	62 (26 April)	77 (11 May)	81 (15 May)
267 cases	**24000** mean up to 25 April			
500 cases	**45000** mean from 27 Mar to 25 Apr			
744 cases	**67000** max value up to 25 April			
1111 cases	**100000** upper bound			

Table 6. Day when the number of positive cases becomes lower than a normalized threshold according to the predictions of the three distributions reported here. The beginning date is February 25 (included).

In our previous work[1,2], a substantial reduction of the daily cases was chosen to occur when the number of cases reduced to about 100. However, since the number of daily swabs is significantly changing and thus the number of cases increases as the number of swabs, we normalize 100 by multiplying it by the ratio: "number of swabs divided by 9000". The number 9000, as mentioned earlier, is approximately the average number of swabs up to March 26. Table 6 shows the dates of a substantial reduction of positive cases predicted with the three different distributions. The date of a substantial reduction of cases in the population should be independent of the number of swabs since it describes the real evolution of the pandemic at a certain stage. For this reason we introduce the normalization just described. In other words, the different numbers of daily cases reported in the first column of Table 6 represent the same actual condition of the pandemic independently on the number of daily swabs.

Since the number of daily swabs was increasing after March 26 by a factor of about five with respect to our previous analysis, a substantial decrease in the number of daily positive cases is reached later with respect to our previous mathematical prediction[1,2]. As mentioned at the beginning of this section, a better indication of the evolution of the pandemic is provided by the behavior of positive cases per swab as reported in Figs. 1 and 2 and in section 3. Indeed. the maximum value of the number of positive cases per swab was 0.462 reached on March 9, whereas the minimum measured value (in the range February 25 to April 30) was equal to 0.02163. According to our previous work[1,2], the pandemic should have significantly reduced during the end of April. Indeed, on May 1, the ratio of daily cases per swab was 0.0265 (i.e., nearly its measured value at the beginning of the pandemic in Italy) and on May 5 the ratio was even smaller, reaching the historical minimum of 0.0195.

In Fig. 6 we report a 3D representation of the number of daily positive cases as a function of the number of daily swabs, from 0 to 100000, and of the day, from February 25, using a distribution of the type of Planck's law.

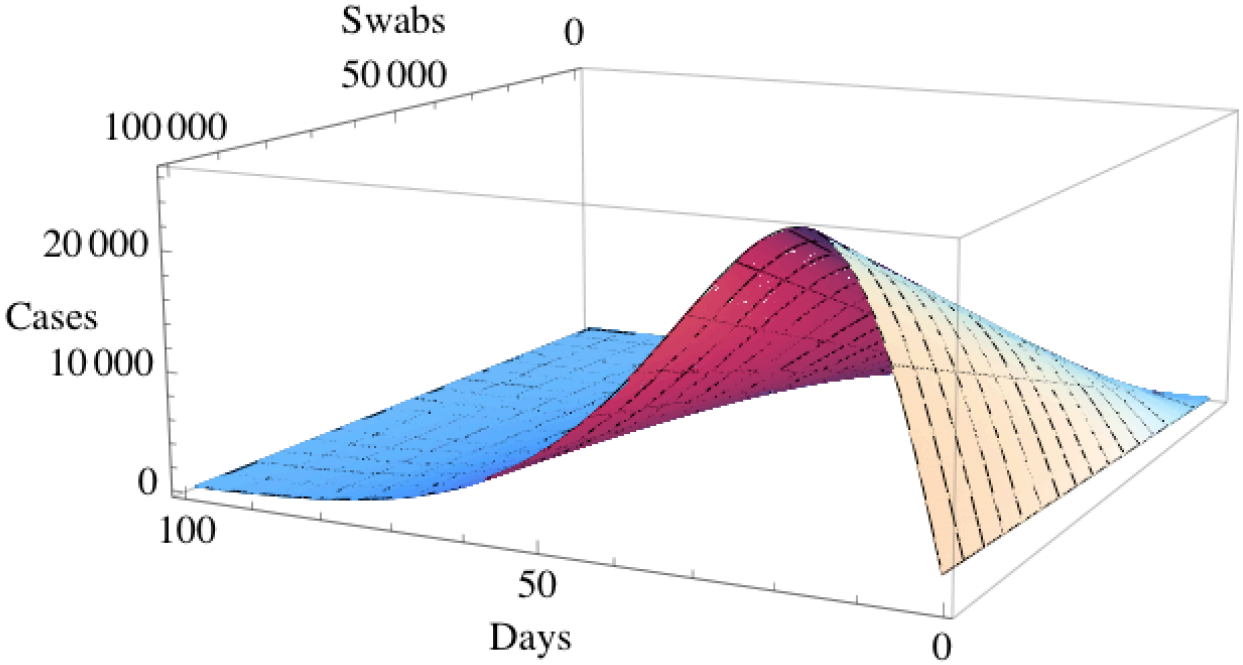


Fig 6. Three dimensional representation of the number of daily cases as a function of the number of daily swabs (from 0 to 100000) and of the day, from February 25, for a distribution of the type of Planck's law with three parameters used to describe the positive cases per swab.

5. Conclusions

Since the number of daily swabs was in Italy highly increasing from March 26 to April 25, we fit the ratio of the positive cases to daily swabs using several functions, including the Gaussian, Weibull, Lognormal, Beta and Gamma distributions and a function of the type of the Planck's law, incidentally this last one fits well the number of daily positive cases in China too. Indeed, using the best fitting distribution, i.e., the Planck-type distribution with three parameters, and depending on the chosen threshold for the daily cases per swab (historical minimum from February 25 to April 30, or one fifth of it, or one tenth of it), we found that a substantial reduction in the daily cases per swab is between May 2 and May 31.

Furthermore, considering that is practically impossible to predict of the evolution of the number of daily swabs, we consider five possible relevant cases for the number of daily swabs. By considering these five cases and the Gauss, Gamma and Planck-type distributions, the range of a substantial decrease in the number of daily positive cases goes from April 26 (in agreement with our previous findings) to May 15, depending on the chosen distribution. To characterize a substantial decrease of the pandemic, we have assumed a threshold of 100 cases per day when an average of 9000 swabs per day are used. However, if for instance, the number of daily swabs is instead 67000 (the maximum number of swabs per day so far reached), the corresponding indication of that substantial decrease in the pandemic is given by a much higher number of cases, i.e., 744. Indeed, since the actual number of positive cases is much higher than the measured ones, by increasing the number of daily swabs, the number of diagnosed positive daily cases would also increase.

To estimate the random uncertainty in the dates of a substantial reduction of the pandemic in Italy, we used a Monte Carlo simulation, with 25000 runs, which provides a 1-sigma random uncertainty of approximately three days (calculated for a threshold of one fifth and one tenth of its historical minimum between February 25 and April 25). In addition, we also estimated the spread in a substantial reduction below a certain threshold of the daily cases per swab among the Italian regions.

Acknowledgments

We gratefully thank Richard Matzner (University of Texas at Austin) for helpful suggestions, Alessandro Paolozzi, and Claudio Paris (Centro Fermi).

References

1. Ciufolini, I., and Paolozzi A., Prediction of the time evolution of the Covid-19 Pandemic in Italy by a Gauss Error Function and Monte Carlo simulations. Submitted to BioRxiv on 03.26.2020 and transferred on 03.27.2020 to MedRxiv, doi: https://doi.org/10.1101/2020.03.27.20045104.

2. Ciufolini, I., and Paolozzi A., Prediction of the time evolution of the Covid-19 Pandemic in Italy by a Gauss Error Function and Monte Carlo simulations, Eur Phys J Plus. 2020; 135(4): 355. doi: 10.1140/epjp/s13360-020-00383-y

3. 1. http://www.salute.gov.it/portale/home.html

4. https://www.worldometers.info/coronavirus/country/italy/

5. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

6. Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang and Jeffrey Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2), *Science*, 16 Mar 2020, eabb3221 DOI: 10.1126/science.abb32214.

7. Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes* (Vol. 3). Cambridge: Cambridge University Press.

8. Bevington, Philip R., and D. Keith Robinson. *Data reduction and error analysis for the physical sciences* (McGraw-Hill, New York, 1969).

9. Ciufolini, I., Monge, B. M., Paolozzi, A., König, R., Sindoni, G., Michalak, G., & Pavlis, E. C. (2013). Monte Carlo simulations of the LARES space experiment to test General Relativity and fundamental physics. *Classical and Quantum Gravity*, *30*(23), 235009.