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 four with respect to our previous analysis, Therefore, here we consider the time evolution of the ratio of 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 then consider five possible cases for the number of daily swabs and we then estimate the potential dates of a substantial reduction in the number of diagnosed positive cases. We then perform a number of Monte Carlo simulations to evaluate the uncertainty in the prediction of the date of a substantial reduction in the number of diagnosed daily cases. Finally, we present an alternative method to evaluate the uncertainty in our mathematical predictions based on the study of each region of Italy and we present an application of the Central Limit Theorem to display the uncertainty in our mathematical predictions based on the analysis of each region.

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 cases based on the assumption that the number of nasopharyngeal swabs would remain roughly constant[1,2]. At the time of our analysis (March 26), the average daily number of swabs from February 15 was about 9000 per day. However, from March 27 up to April 19, the average number of daily swabs was about 41500. 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 number of swabs. 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 have considered five possible cases for the daily number of swabs and we have also assumed some possible time evolution in the number of daily swabs. We fitted the time evolution of the positive cases per unit of swab up to April 19, using three different distributions: the Gauss, the Planck and the Gamma distribution. After estimating the time evolution of the positive cases per unit of swab using these three distributions and the conceivable number of daily swabs, we were able to estimate the evolution in the number of diagnosed positive cases and the dates of a substantial reduction in such a daily number.

A basic problem is to mathematically estimate the uncertainty in the date of a substantial reduction of daily cases. For such a purpose, in section 4 we report the results of 25000 Monte Carlo simulations. Furthermore, in section 5 we present an alternative way to estimate the uncertainty in the dates of a substantial reduction in the number of daily cases which is based on the study of each region of Italy, where the conditions are quite different from each other, including the number of swabs per unit of person. We finally present an elegant way to display the results of Italy using the analysis of the single regions based on the Central Limit theorem.

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

After analyzing the time trend of the ratio of daily positive cases to the number of daily swabs, we found that this trend can be modeled by a Gauss distribution but this time trend has also a small amount of skewness that can be fitted by choosing a skewed distribution such as the Weibull, Log-normal, Beta and Gamma distributions, and also other functions such as the Planck's law. This last one, for example, with three parameters *a, b* and *c*:

$$\frac{2a}{c^2} t^3 \frac{1}{e^{-\frac{a}{b}t} - 1}$$

where t is the time, is reported in Fig. 1a. In Figs. 1b to 1e are reported the fits of the data with the Weibull, Lognormal, Beta and Gamma distributions, respectively. However, the data can also be well 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. 2.

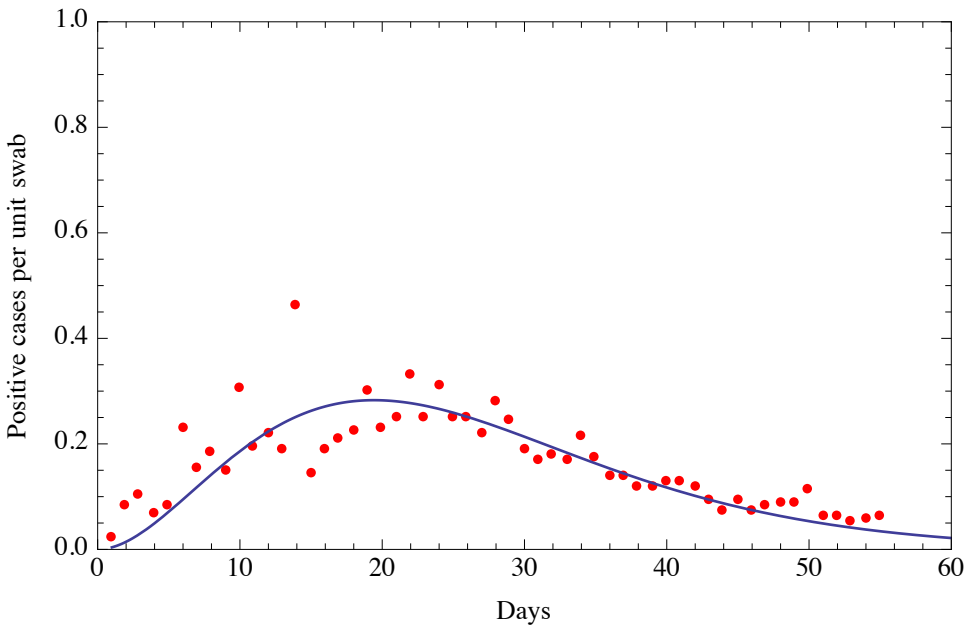

Fig. 1a Fit of the ratio of daily positive cases per unit swab obtained with a function of the type of the black body Planck's law with three parameters. The beginning date is February 25.

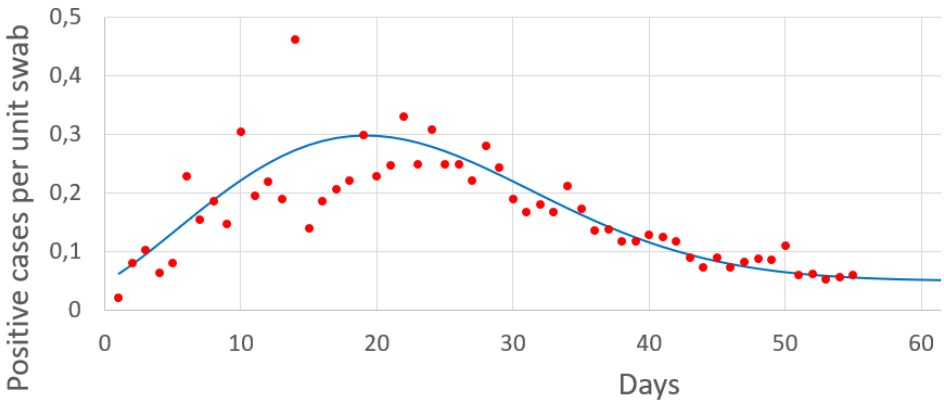

Fig. 1b Fit of the ratio of daily positive cases per unit swab obtained with a Weibull distribution. The beginning date is February 25.

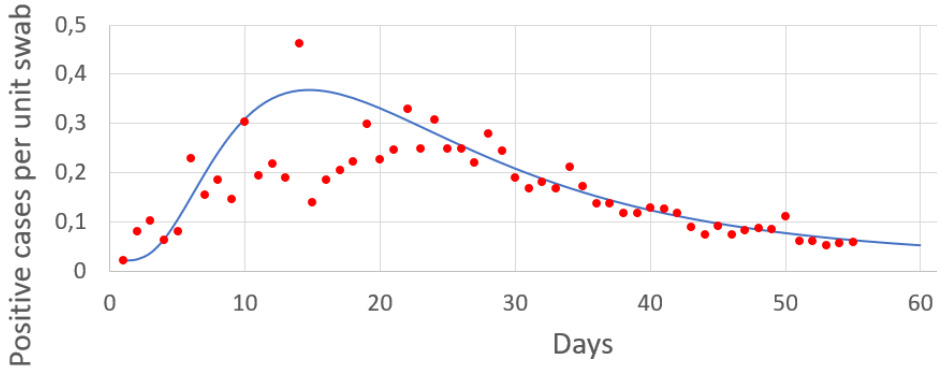

Fig. 1c Fit of the ratio of daily positive cases per unit swab obtained with a Lognormal distribution. The beginning date is February 25.

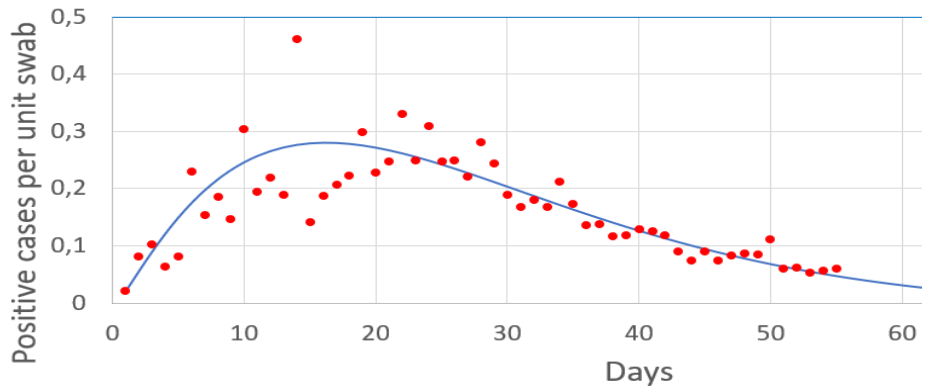

Fig. 1d Fit of the ratio of daily positive cases per unit swab obtained with a Beta distribution. The beginning date is February 25.

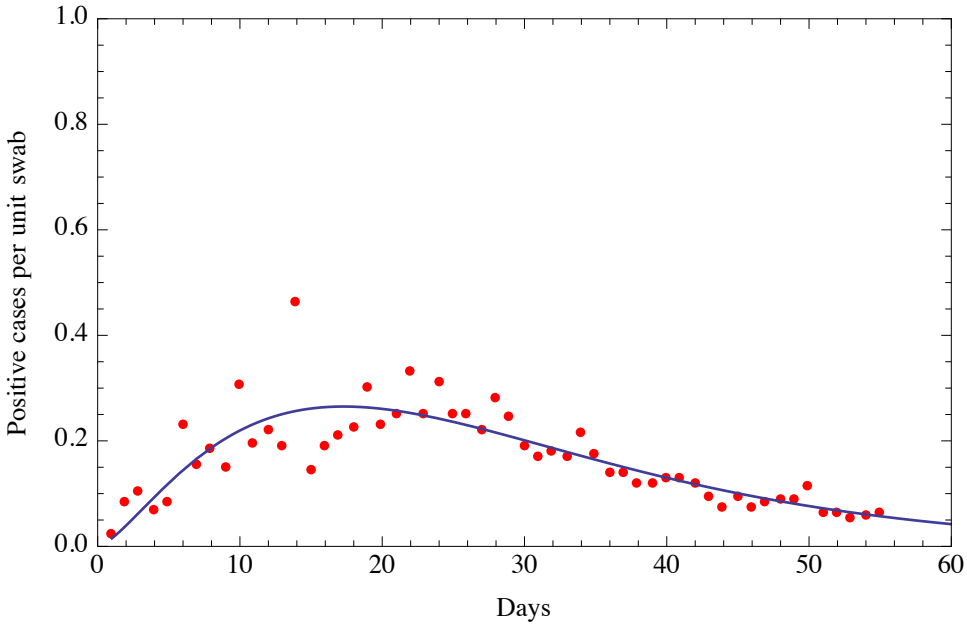

Fig. 1e Fit of the ratio of daily positive cases per unit swab obtained with a Gamma distribution. The beginning date is February 25.

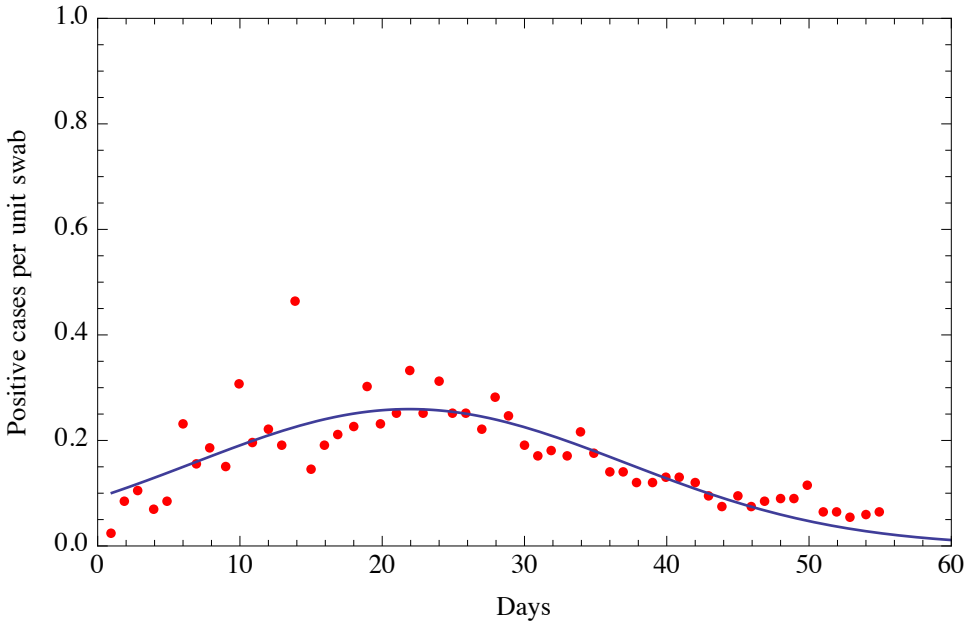

Fig. 2 Fit of the ratio of daily positive cases per unit swab obtained with function of the type of a Gauss function with three parameters. The beginning date is February 25.

3. Modelling the daily nasopharyngeal swabs

Since we do not know the future capabilities of the Italian laboratories to perform the daily swabs, we have selected five cases to represent the number of swabs which will be performed daily. We have also modelled the number of daily swabs using a Gaussian distribution, Planck's law and a linear monotonic increasing distribution. In Fig. 3, we report the corresponding fits.

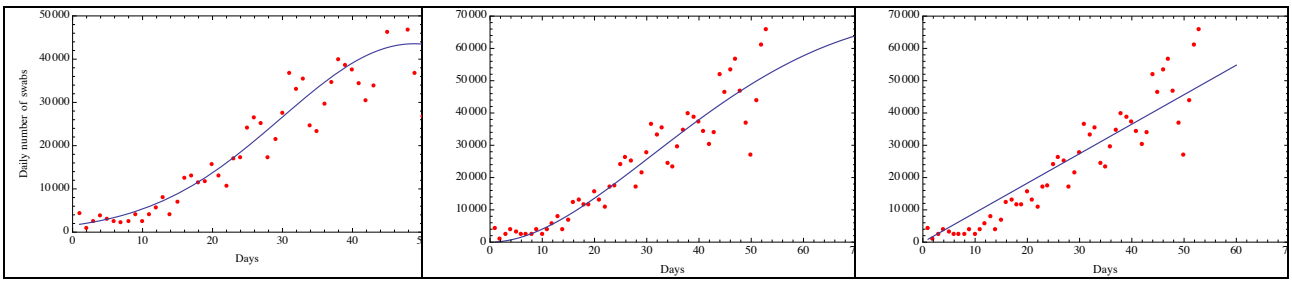

Fig. 3 Daily number of swabs and fitting functions of the type of Gaussian (left) distribution, Planck's law (center) and linear fit. The beginning date is February 25.

Now we have both the distributions fitting the daily positive cases per unit swab and the distributions fitting the number of daily swabs. In order to model the number of daily positive cases we can simply multiply these two distributions obtaining a total of nine cases. Just as an example we report here only one case. In Fig. 4 we report both the number of daily positive cases modeled with the product of two Gaussians (4a) and the cumulative number of positive cases modelled with the integral of such a product (4b). Since the product of two Gaussian is still a Gaussian, it is possible in this case to easily obtain the primitive which is the Error Function.

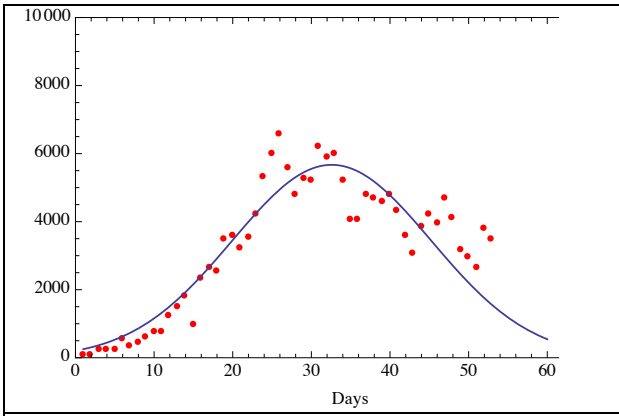

Fig. 4a. Fit of the daily positive diagnosed cases with a function obtained as the product of the two Gaussian functions used to fit both the positive cases-to-swab ratio and the number of daily swabs. The y-axis reports the number of daily positive cases.

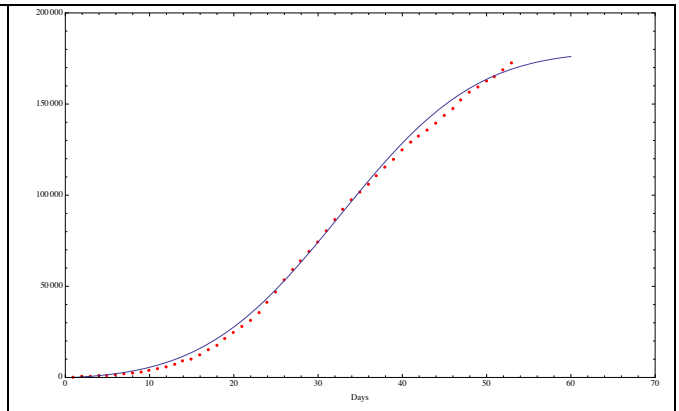

Fig. 4b. Fit of the cumulative positive diagnosed cases with a function of the type of a Gauss error function, obtained as the integral of the product of the two Gaussian functions used to fit both the positive cases-to-swab ratio and the number of daily swabs. The y-axis reports the cumulative number of positive cases.

Since the number of daily swabs depends on factors that are difficult to predict, such as the daily availability of reagents and specialized personnel, we have considered five possible cases for the daily number of swabs corresponding to some relevant situations. The five cases are: 9000, i.e., 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]). 21000, i.e., the mean between February 15 and April 19. 41000, i.e., the mean between March 26 and April 19. 66000, i.e., the maximum number of swabs up to April 19. 100000, i.e., an estimated possible upper bound to the number of daily swabs, see Table 1. In Table 1 are reported those five cases for the Gauss distribution and the Planck's law.

Daily swabs ↓ Distribution→	Gauss	Planck
9000 mean up to 26 March	62 (26 April)	68 (2 May)
21000 mean up to 19 April	66 (30 April)	77 (11 May)
41000 mean from 26 Mar to 19 Apr	70 (4 May)	83 (17 May)
66000 max value up to 19 April	72 (6 May)	87 (21 May)
100000 upper bound	74 (8 May)	91 (25 May)

Tab.1 Days in which the number of positive cases becomes lower than 100. The beginning date is February 25 (included)

Tab. 1 shows a variability in the dates. That is partly due to the large variability in the number of daily swabs, from 9000 to 100000, which will obviously introduce a large variability in the number of positive cases. Therefore, the threshold value of 100 cases per day that was used should be renormalized for the number of daily swabs.

Since the number of daily swabs was increasing by a factor of about four 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]. A better indication of the evolution of the pandemic is reported in Figs. 1 and 2 which report the positive cases per unit swab. The curves show that the pandemic is significantly reducing during the end of April.

As an example, in Fig. 4c, we report a 3D representation of the number of daily cases as a function of the number of daily swabs for each day for a Gaussian distribution of the number of cases per swab.

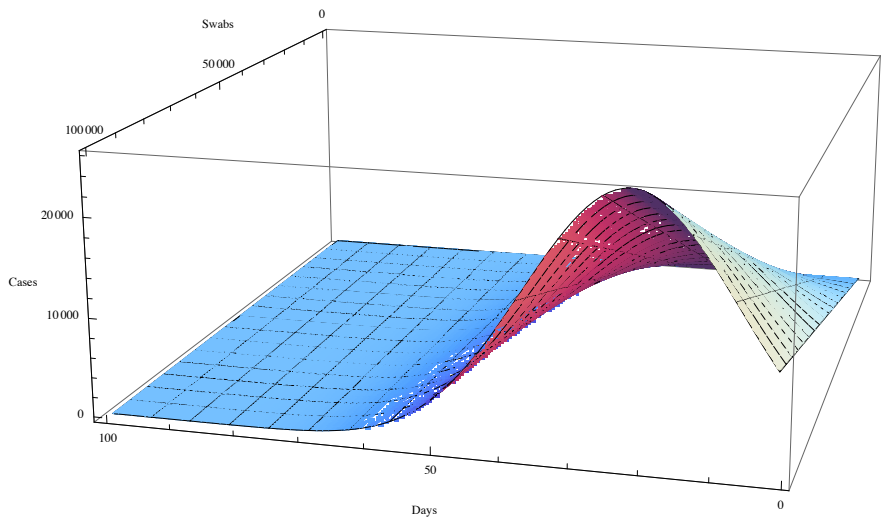

Fig 4c. 3D representation of the number of daily cases as a function of the number of daily swabs for each day for a Gaussian distribution of the number of case per swab.

In the next sections 4 and 5 the uncertainties associated with the mathematical predictions reported above are evaluated.

4. Monte Carlo simulations of cumulative positive cases of Covid-19 in Italy

Several uncertainties can influence the diagnosed 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 3. To possibly estimate the uncertainties in the number of positive cases, we have used two methods: Monte Carlo simulations[1,2,3], similarly to what done in previous works[4,5] and a new method, described in the next section 5, using the study of each region of Italy. The uncertainty we consider in the Monte Carlo simulation is <u>not</u> the difference between the total number of the actual positive cases (which is unknown) and the 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. The Monte Carlo simulations were performed for the number of positive cases.

For convenience to the reader, we summarize here the procedure used previously[4,5], the only difference being the number of simulations that have been largely increased from 150 to 25000 in this section, and to 100000 in the next section. We have assumed a measurement uncertainty in the total number of positive cases equal to 20% of each daily number (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, we generated n Gaussian distributions with 25000 outcomes, with means equal to the n nominal values and with 20% standard deviation. Then, for each of the 25000 simulations, those n values (corresponding to the cumulative positive cases, of n days) were fitted with a one parameter function of the type of the Gauss Error Function (see section 2) and we then determined the date of the flex with such fitted function for each simulation. Using the fitted function we also determined the date at which the number of daily positive cases will be less than a certain threshold that, for example, we have chosen to be 100 for the diagnosed positive cases. Finally, we calculated the standard deviation of the 25000 simulations. The value of the standard deviation is about one day.

In Fig. 5, is reported the histogram of the frequencies versus the day of a substantial reduction in the number of daily positive cases which has been chosen to be 100. The histogram approaches a Gaussian with mean equal to day $\cong 67$, approximately corresponding to what reported in Figs. 1 and 2 for a substantial reduction in the number of positive cases per unit swab. The standard deviation in Fig. 5 is approximately 1 day.

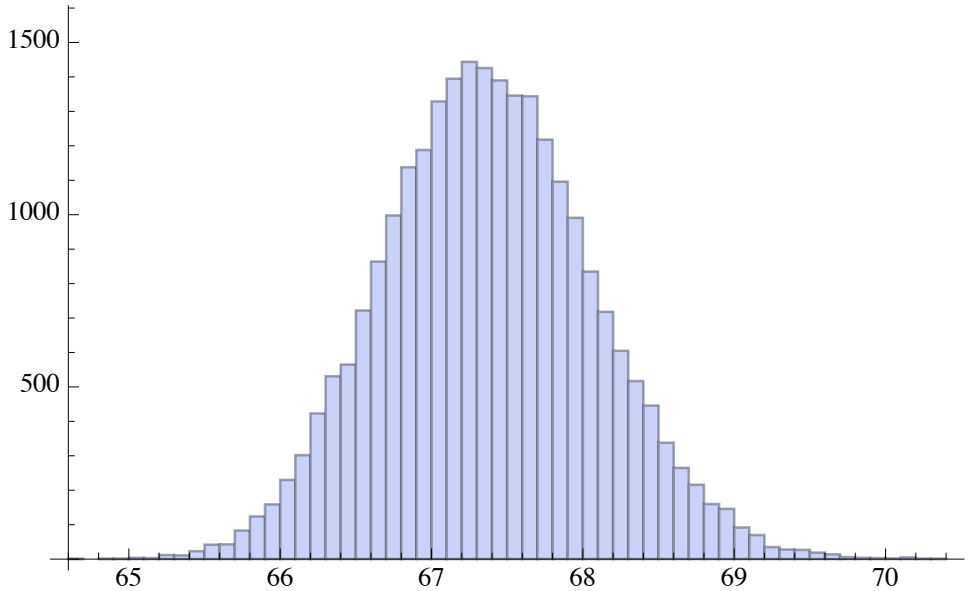

Fig. 5. Histogram of the 25000 Monte Carlo simulations for Italy: frequencies versus the day in which a substantial reduction in the number of daily cases occurs.

5. Analysis of uncertainty using each region of Italy

As an alternative approach to estimate the uncertainty in the date of a substantial reduction in the positive cases, we independently analyzed each of the twenty Italian regions from February 22, 2020 (included) until April 6, 2020 (included). Indeed, the number of daily swabs and other relevant conditions vary quite differently from one region to the other. We then evaluated the date of a reduction of the positive cases in each region below a certain threshold. The national threshold was chosen to be 100 cases, however, for each region, we normalized 100 for the number of positive cases in each region at the date of April 6 divided by the total number of national cases at the date of April 6 (i.e., 132547 cases). We then fitted the cumulative number of cases of each region using a function of the type of an Error Function including four free parameters and finally, for each region, we obtained the date at which there is a substantial reduction of the diagnosed positive cases below the given threshold for each region. In Fig. 6 we report the 20 dates for each region. We then calculated the mean and the standard deviation of the 20 dates and we obtained a 1-sigma standard deviation of 9 days.

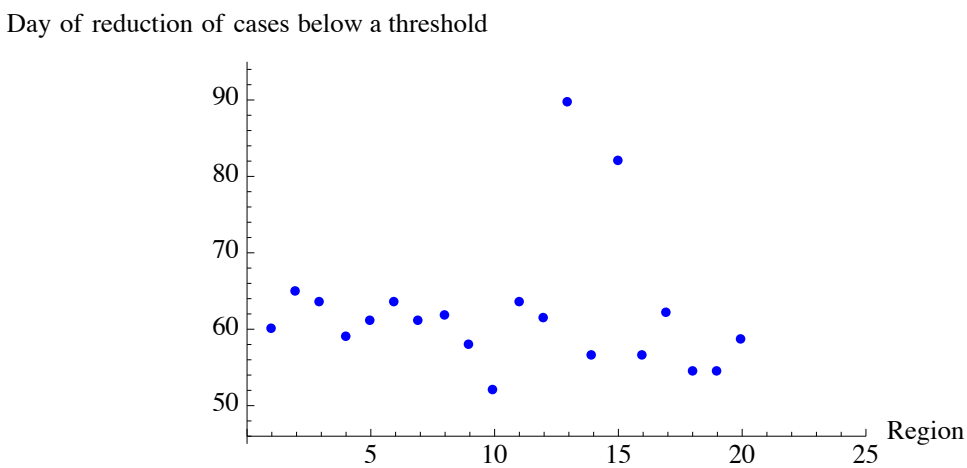

Fig. 6. Dates of a substantial reduction in the number of daily diagnosed positive cases for each of the 20 Italian regions. The 1-sigma standard deviation is about 9 days.

This result may also be displayed using the Central Limit Theorem. We have randomly chosen a sample of 30 regions, with repetition, out of the 20 regions, for 100000 times. We then calculated the mean of the day of a

substantial reduction for each sample and we reported the mean of each of the 100000 samples in the histogram of Fig. 7. We obtained a Gaussian with a standard deviation of 1.62 days, that, multiplied for the square root of 30, gives back a standard deviation of about 9 days. Since each region has a different number of daily swabs and quite different conditions, this figure might very roughly represent a kind of systematic error in the estimate of a substantial reduction in the number of daily positive cases in Italy. In this analysis we didn't take into account the systematic increasing number of swabs since we just wanted to evaluate the uncertainty in the date of a substantial reduction in the positive cases. This uncertainty seems to be consistent with the large variability in the date of a substantial reduction in the number of cases due to the assumed fitting distributions of the ratio of cases per swab and to the number of daily swabs, as reported in Table 1.

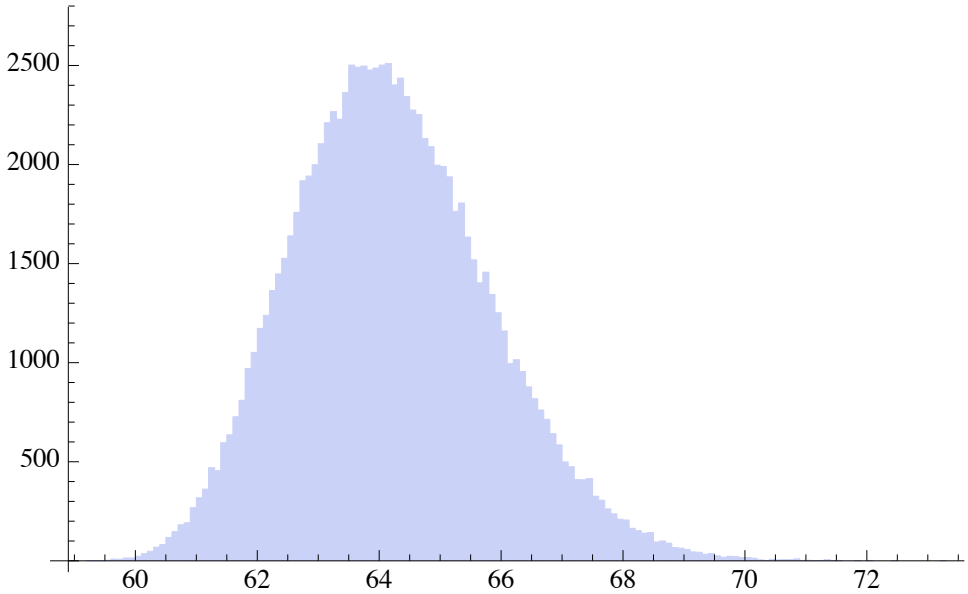

Fig. 7 Result of 100000 simulations using the 20 Italian regions for an estimate of the uncertainty in the date of a substantial reduction in the number of positive cases.

Conclusions

By considering the largely increasing number of daily swabs, from March 26 to April 19, we fitted 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 well fits the number of daily positive cases in China. Considering the difficulty in the prediction of the evolution of the number of daily swabs we only marginally analyzed possible fitting functions of the daily swabs, however we considered five possible relevant cases for the number of daily swabs. By considering these five cases and the Gauss and the Planck functions, the range of a substantial decrease in the number of daily positive cases goes from April 26 to May 25. By taking the mean number of swabs from February 25 to April 19, we obtained the range April 30 to May 11. By using the mean in the number of daily swabs, from February 25 to March 26 (the period of analysis used in our previous paper), we obtained the range from April 26 to May 2, in agreement with our previous findings. To estimate the uncertainties in these dates, we used 25000 Monte Carlo simulations with the Italian data which provided a random uncertainty of about one day. However, to possibly estimate some of the systematic uncertainties affecting our results, we also used the spread in a substantial reduction of the positive cases of each region below a certain threshold. Using this second method, we found an uncertainty of about 9 days about the mean date, in agreement with the estimates given above by changing the fitting functions of the positive cases per swab and the number of daily swabs.

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, to appear in European Journal of Physics Plus.

3. 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.

4. Lessler, J., Reich, N. G., Brookmeyer, R., Perl, T. M., Nelson, K. E., & Cummings, D. A. (2009). Incubation periods of acute respiratory viral infections: a systematic review. *The Lancet infectious diseases*, *9*(5), 291-300.

5. IHME COVID-19 Health Service Utilization Forecasting Team, University of Washington (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator- days and deaths by US state in the next 4 months. Preprint submitted to MedRxiv, tracking ID MEDRXIV/2020/043752.