

Assessment of the weekly fluctuations of the Covid-19 cases in Italy and worldwide

Pier Luigi Bragato

National Institute of Oceanography and Applied Geophysics – OGS

pbragato@inogs.it

May 10, 2020

Abstract

The time series of daily new cases of Covid-19 infection in different zones of Italy, several European countries and New York City were analyzed in order to characterize the large fluctuations contained in them. The comparison shows that the time series share a similar evolution, although with delays of one or two weeks, with Italy that precedes the other countries. Correlation analysis indicates a strict resemblance between the time series of the first differences, which emphasize accelerations and slowdowns of the epidemic (e.g., correlation coefficient $r=0.70$ for Italian vs. worldwide data). According to Fourier analysis the signals have a dominant oscillatory component at a period of seven days, with minima located near the weekends. Some of the possible causes of this time modulation are discussed. In particular, it is hypothesized a delayed or missed recognition of a significant fraction of the new cases during the weekend, with possible consequences for the evolution of the epidemic.

Keywords

Covid-19 pandemic; weekly fluctuations; delayed recognition of new cases; health system situation.

Introduction

The trend of the Covid-19 epidemic in Italy since the end of February 2020 was characterized by large fluctuations in the number of daily new cases, with a series of abrupt accelerations, sometimes preceded by a few quieter days, when the epidemic seemed to slow down. These variations had psychological effects, discouraging people subjected to the heavy restrictions of the lockdown. They also surprised the scientists, which wondered if the dynamic of epidemic or the virus itself had changed¹. During the current decreasing phase of the epidemic the oscillations could be interpreted as the consequence of a less rigorous application of the lockdown. Consequently they have direct implications for the decisions to be taken by authorities on how to restart economic and social activities. The aim of this study was to assess the characteristics of the observed oscillations by considering the time series of the number of new Covid-19 cases per day recognized in different regions of Italy and in groups of regions. To get confirmation of a general epidemic behavior, a similar analysis was conducted for some countries in Western Europe, for New York City (NYC hereafter) as well as at the global level. NYC was preferred to the entire USA as it combines a large population with homogeneous procedures for recognizing and accounting Covid-19 cases, also avoiding the complication due to the time of spreading of the epidemic across a large territory.

Data collection

The quantity taken into consideration was the number new Covid-19 cases recognized each day in the study areas. It is well known that, for various reasons, it is just a biased and delayed sample of the population affected by the virus. Nonetheless it is a clear and simple indicator of the state of evolution of the epidemic, also used by authorities as one of the parameters for emergency management. In the present study no attempt was made for removing or mitigating the bias using auxiliary data. Rather, several noisy time series, referring to different areas, were compared in order to bring out a coherent, useful signal. For the case of Italy, the number of new cases was that published daily by the Dipartimento di Protezione Civile for the entire country and for each of its regions (<https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-regioni/dpc-covid19-ita-regioni.csv>). The European and global data source was the web site of the European Centre for Disease Prevention and Control (ECDC, <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>). For NYC the data were those published and

updated daily by the NYC Health Department through the official web site of the city (<https://www1.nyc.gov/site/doh/covid/covid-19-data.page>). The observations cover the time periods February 24 - April 30 for Italy, March 1-April 30 in the other cases. With the exception of NYC, the data represent the number of laboratory tests with positive outcome produced in each day. The outcome of the test may follow one or more days the collection of the swab, which in turn is delayed in respect to the onset of the symptoms. The delay is neither fixed nor easily quantifiable, as it depends on the directives given by authorities, the severity of cases and the capability of the health system and laboratories to collect and process large quantities of tests. The latter point is very critical in the most acute phase of the epidemic. For NYC a new case is assigned retroactively to the day in which the swab was taken, so that the time from the onset of symptoms is reduced. The time series taken into consideration are depicted in Figs. 1 and 2.

Analysis

The investigation was performed separately for each territorial unit (region or country) and for groups of them. For each territorial unit it was analyzed the time series of daily new cases

$$X = \{x_t: t = 1, 2, \dots, n\}. \quad (1)$$

In order to make them comparable, each time series was normalized to its maximum value (e.g. red dots in Figs. 3a and 3c for the Italian regions Lombardy and Veneto). Furthermore, they were smoothed for removing the short-term, possibly random variability (blue lines in Figs. 3a and 3c). Smoothing was performed by kernel regression² using a Gaussian smoothing function with $\sigma=1$ day. From each normalized and smoothed time series

$$X_{NS} = \{x_{NS_t}: t = 1, 2, \dots, n\}, \quad (2)$$

it was obtained the time series of the first differences

$$dX_{NS} = \{dx_{NS_t}: dx_{NS_t} = x_{NS_t} - x_{NS_{t-1}}, t = 2, 3, \dots, n\}, \quad (3)$$

used in correlation analysis. By matching the time series of the first differences, the interest was posed on the common accelerations and decelerations of the epidemic. The first differences also remove the long-term trends, so reducing the potential for spurious correlation³. As a last step, the power spectral density (PSD) of dX_{NS} was computed using the function "spectrum" of the R software⁴. The inspection of the PSD provided information on the oscillatory components that dominate the signals. Groups of territorial

units (regions or countries) were processed similarly to the individual units. In those cases smoothing kernel regression was performed on the union of the normalized daily values (Figs. 3e and 3g), so that each area of the group was given the same weight in the estimation.

The results of the analysis are illustrated in Figs. 4-8 for some combinations of regions and countries of particular interest. Figure 4 concerns two regions of Italy, Lombardy and Veneto, compared with the average trend of all the regions (data in Fig. 3i). Lombardy is the region most affected by the epidemic, with about 30% of the overall number of cases in Italy. Its health system was heavily stressed by the epidemic, perhaps over its capabilities⁵, so that only patients with severe symptoms were tested⁶, with consequent large underestimation of the symptomatic cases (not to say of the asymptomatic ones). Even in Veneto the epidemic hit hard, but within the management capabilities of the health system, so that it was also possible to test patients with weaker symptoms and, in some occasions, asymptomatic people⁶. In terms of recognized cases (smoothed normalized curves in Fig. 4a) the time evolution of the epidemic was similar in the two regions during the increasing period, with a main peak and a secondary peak reached around March 21 and March 28, respectively. Later, in the decreasing part of the epidemic, they diverged, with, in proportion, more cases in Veneto, perhaps due to the possibility to perform a larger number of tests in the less critical period. The time series of the first differences (Fig. 4b) are positively correlated, with Pearson correlation coefficient $r=0.58$. Also the correlation with the average Italian time series is high ($r=0.79$ and $r=0.67$ in the comparison with Lombardy and Veneto, respectively). According to Fourier analysis (Fig. 4c) each time series is dominated an oscillation at the period of 7 days. Similar results have been obtained by comparing the group of most affected region of northern Italy with the group of the largest regions of southern Italy (Fig. 5, $r=0.58$).

In several countries of continental Western Europe the epidemic emerged one week behind Italy. Fig. 6a shows the average time series of the Covid-19 cases computed for the group formed by the four most populated of these countries (France, Germany, Netherlands and Spain, FGNS group hereafter) that adopted similar lockdown measures. After the adjustment for the time lag, the time evolution of the epidemic appears very similar to the Italian one, especially in the comparison of the first differences ($r=0.87$ in Fig.6b), both dominated by the oscillation with the time period of 7 days (Fig.5c). In the same figure, the curves of FGNS are compared with those of the worldwide cases. The two time series had similar trend during the month of March, when the global data were

dominated by the Covid-19 cases in Europe, while they diverged after the first week of April (day 38 in the graph), when the global data became dominated by the cases in USA. Nonetheless, the times series of the first differences are similar ($r=0.75$), denoting fluctuations with the same weekly rhythm. NYC (Fig. 7) is representative of a mixed public/private health system (it is mainly public in Europe) heavily stressed by the epidemic (a situation similar to that of Lombardy in Italy), where the lockdown started on March 16, together to many European countries (Fig.2). The normalized-smoothed time series of daily new cases in NYC resembles that of FGNS (Fig. 7a), with FGNS two days behind NYC. Likely this time lag is an effect of the choice made in NTYC to assign the new cases to the day of execution of the swab. With the appropriate time shift, the time series of NYC also matches those of Italy. In particular, there is strong positive correlation between the time series of the first differences ($r=0.77$ and $r=0.79$ for FGNS and Italy, respectively), which are all dominated by a weekly period of oscillation. It is to note that in NYC the minima of the fluctuations are located exactly on Saturday and Sunday (see also Fig. 2).

At last, the graphs in Fig. 8 illustrate the trend of the epidemic in the United Kingdom and Sweden, where the epidemic was one week behind FGNS (two weeks behind Italy). Furthermore, Sweden applied no lockdown measures⁷. Even in these countries the evolution was similar to FGNS in the rhythm of accelerations and slowdowns (Fig.8b and 8c) with $r=0.56$ and $r=0.53$ in the two cases. The correlation was much higher for the couple United Kingdom – Sweden ($r=0.73$ in Fig.8b)

Discussion and conclusions

Despite the differences in the typology the health systems (public/private), their level of stress and the strategies and constrains in the identification of the new cases, the time evolution of the Covid-19 epidemic has been quite similar (although with delays and different intensities) in Italy, in the cluster of European countries FGNS and in NYC. Considering the time series of the first differences, which emphasizes accelerations and decelerations of the epidemic, the similarity extends to the United Kingdom, to Sweden (in absence of lockdown) and at the global level. The main common feature that emerges from Fourier analysis is the dominance of an oscillation with a period of one week. In NYC, where the new cases are attributed to the day of collection of the swab and very near to the onset of symptoms, the minima of the fluctuations occur on Saturday and Sunday. In the other countries the minima occurs about two days later, which might be explained as the

average time of swab processing. It would be of interest to find the origin of the variations and to assess their consequences for the evolution of the epidemic. At first glance, in Italy they might depend on the series of lockdown measures with increasing restrictions introduced almost weekly between February 24 and March 22, with effects that, based on a rough estimation of the incubation period of the virus, were awaited within 14 days (horizontal segments in Fig. 2a). This hypothesis should be rejected based on the comparison with the other countries, in particular with Sweden, where no lockdown was imposed. In a specular manner, the Italian fluctuations in the decreasing period of the epidemic should not be due to a less rigorous application of the lockdown. The example of NYC shows that the periodicity is not related with the rhythm swab processing. Furthermore, the fluctuations also occurred in Lombardy during the peak of the epidemic, when the laboratories worked almost continuously 24 hours a day seven days a week to process as many swabs as possible. Another hypothesis is that some symptomatic people delayed a medical examination during the weekend. This explanation should be the simplest one in normal times, especially for weak symptoms. It sounds unlikely during the peak of an epidemic whose symptoms and great risks are well known. This especially in a public health system, which is fully operative during the weekend (behind its limits in Lombardy) with no extra costs for the patients, even with telephonic advice and (for not-saturated health systems) home medical examinations. Anyway, in case of delay it should be awaited a rebound of cases on Monday. According to Fig.2g this did not happen in NYC, where, with the exclusion of March 30, the number of infected people recognized on the first day of the week agrees with the trend in the following few days, both in the increasing period (March 16 and 23) and in decreasing period (e.g. April 6 and 13, days 37 and 44 in the figure, respectively). The lack of rebound suggests that some people simply avoided any medical consultation, perhaps due to the disappearance of the symptoms in less-serious cases. Independently on the motivation, any delayed or missed detection had consequences for the epidemic. In fact, people that were clearly contagious were not promptly isolated, so contributing to the spread of the disease. The effect was modeled in a previous study⁸ with reference to some pandemics of the past (e.g. the Spanish Flu of 1917): it was shown that delays of a few days in the recognition of contagious people can undermine the effectiveness of any isolation and quarantine measure. It will be possible to try a similar analysis for Covid-19 in the near future, as the data of the pandemic will be consolidated.

Acknowledgements

Correlation and spectral analyses were performed using R software version 3.3.3⁴.

All the figures were made using Generic Mapping Tools version 5.1.1⁹ available at <http://www.soest.hawaii.edu/gmt>.

References

1. Olgiati, P. Coronavirus, Gismondo: "Forse è mutato, troppi morti in Lombardia", ADSKronos. https://www.adnkronos.com/fatti/cronaca/2020/03/21/virologia-forse-virus-mutato-troppi-morti-lombardia_rkfGj7XVYSoucJM10GIGyL.html. Published March 21, 2020. Accessed May 10, 2020.
2. Bowman, AW, Azzalini, A. Applied Smoothing Techniques for Data Analysis, 193 pp., Oxford Univ. Press, Oxford, U. K. Published 1997.
3. Johansen S. The analysis of nonstationary time series using regression, correlation and cointegration, Contemporary Economics. 2012: 6:40-57.
4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>. Published 2013. Accessed May 10, 2020.
5. Winfield, N, and Barry C. Italy's health system at limit in virus-struck Lombardy, AP News. <https://apnews.com/837274f1bab9af1aab12f1b9481b2d62>. Published March 2, 2020. Accessed May 10, 2020.
6. Misculin, L. A chi stiamo facendo il tampone?, Il Post. <https://www.ilpost.it/2020/03/20/tampone-test-coronavirus/>. Published March 20, 2020. Accessed May 10, 2020.
7. Milne, R. Swedish ministers defend resisting coronavirus lockdown, Financial Times. <https://www.ft.com/content/5eb0a90b-ceb5-4441-9456-e30f9a2a7028>. Published April 16, 2020. Accessed May 10, 2020.
8. Young, L-S, Ruschel, S, Yanchuk, S, Pereira, T. Consequences of delays and imperfect implementation of isolation in epidemic control, Scientific Reports. 2019:9:3505.
9. Wessel, P, Smith, WHF, Scharroo, R, Luis, J, Wobbe, F. Generic Mapping Tools: Improved version released, Eos Trans. AGU. 2013:94:409–410.

Figure Captions

- Figure 1.** Daily new Covid-19 cases in Italy and in some of its regions. The vertical grid marks the Monday of each week.
- Figure 2.** Daily new Covid-19 cases in some European countries and in New York City. The horizontal segments show the beginning of the lockdown measures and the following two weeks, representative of the incubation period of the virus. The vertical grid marks the Monday of each week.
- Figure 3.** Time series considered in this study for some Italian regions and groups of them. Panels on the left: daily new Covid-19 cases (normalized values, red points) and the corresponding smoothed curve estimated by kernel regression (X_{NS} in equation 2). Panels on the right: time series of the first differences of the smoothed curve on the left (dX_{NS} in equation 3).
- Figure 4.** Time evolution of the Covid-19 epidemic in the Italian regions Lombardy and Veneto compared with the average behavior of all the regions. a) smoothed normalized time series of daily new cases (equation 2); b) time series of the first differences of the smoothed curves (equation 3); c) power spectra density of the time series in (b). In (b) the Pearson correlation coefficient for couples of time series is also reported. The vertical grids marks the Monday of each week.
- Figure 5.** Time evolution of the Covid-19 epidemic for groups of regions in northern and southern Italy compared with the average behavior of all the regions. The meaning of the graphs is the same as in Fig.4.
- Figure 6.** Comparison between the time evolution of the Covid-19 epidemic in Italy, a group of European countries and worldwide. The meaning of the graphs is the same as in Fig.4. The time series of Italy is shifted forward one week.
- Figure 7.** Comparison between the time evolution of the Covid-19 epidemic in Italy, a group of European countries and New York City. The meaning of the graphs is the same as in Fig.4. The time series of Italy and NYC are shifted forward 7 and 2 days, respectively.
- Figure 8.** Comparison between the time evolution of the Covid-19 epidemic in a group of European countries, in the United Kingdom and in Sweden. The meaning of the graphs is the same as in Fig.4. The time series of the group of European countries is shifted forward 7 days.

Figures

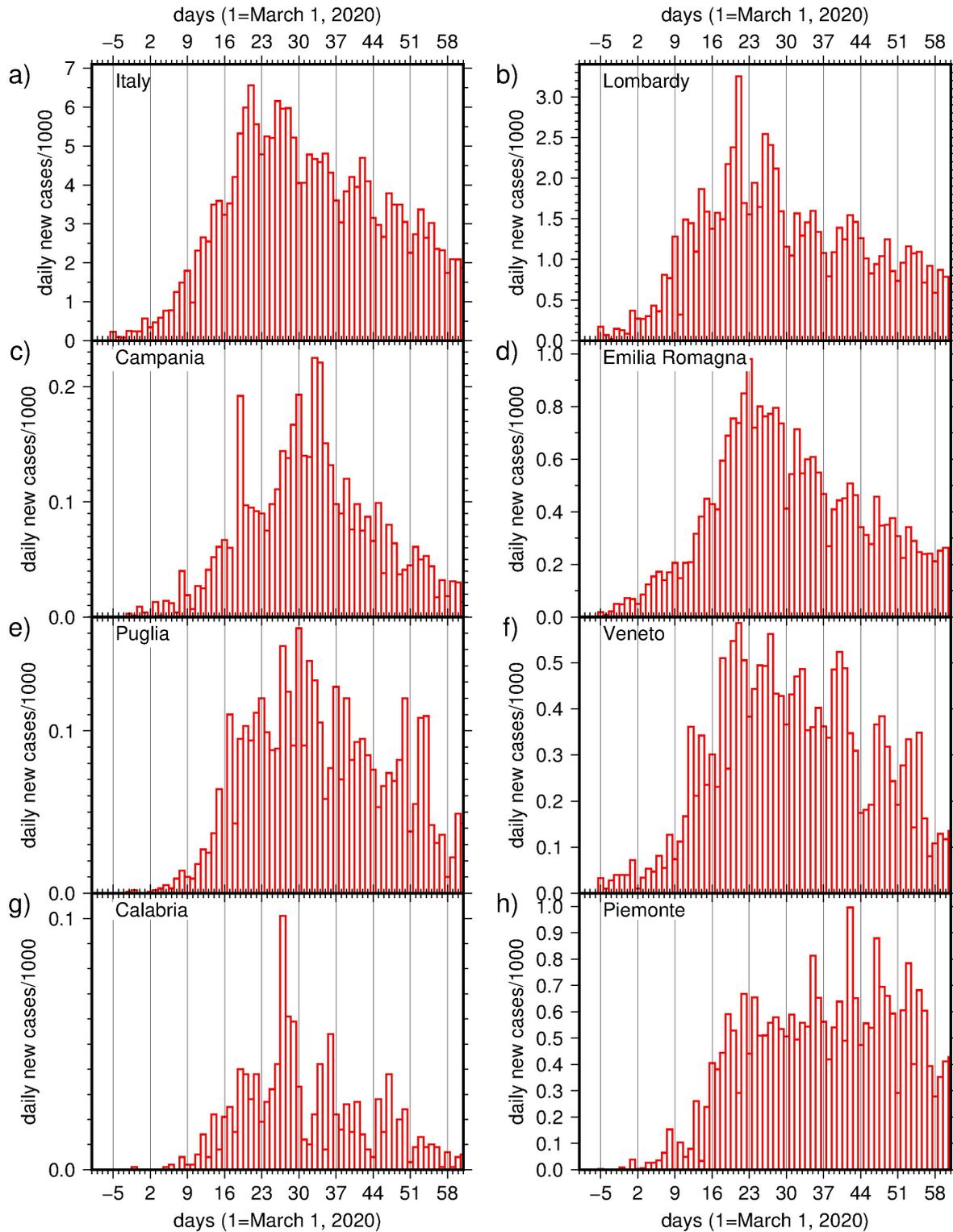


Figure 1. Daily new Covid-19 cases in Italy and in some of its regions. The vertical grid marks the Monday of each week.

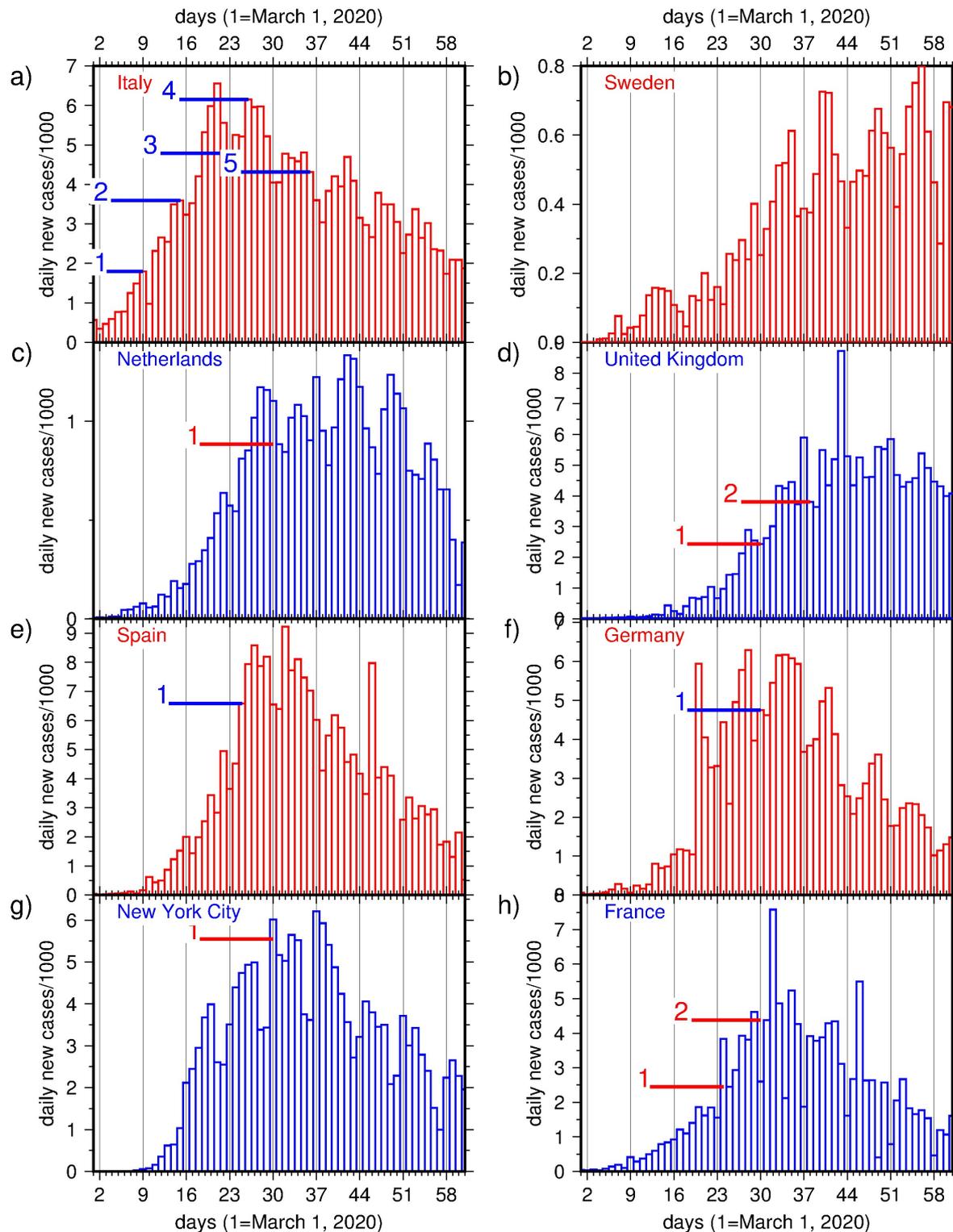


Figure 2. Daily new Covid-19 cases in some European countries and in New York City. The horizontal segments show the beginning of the lockdown measures and the following two weeks, representative of the incubation period of the virus. The vertical grid marks the Monday of each week.

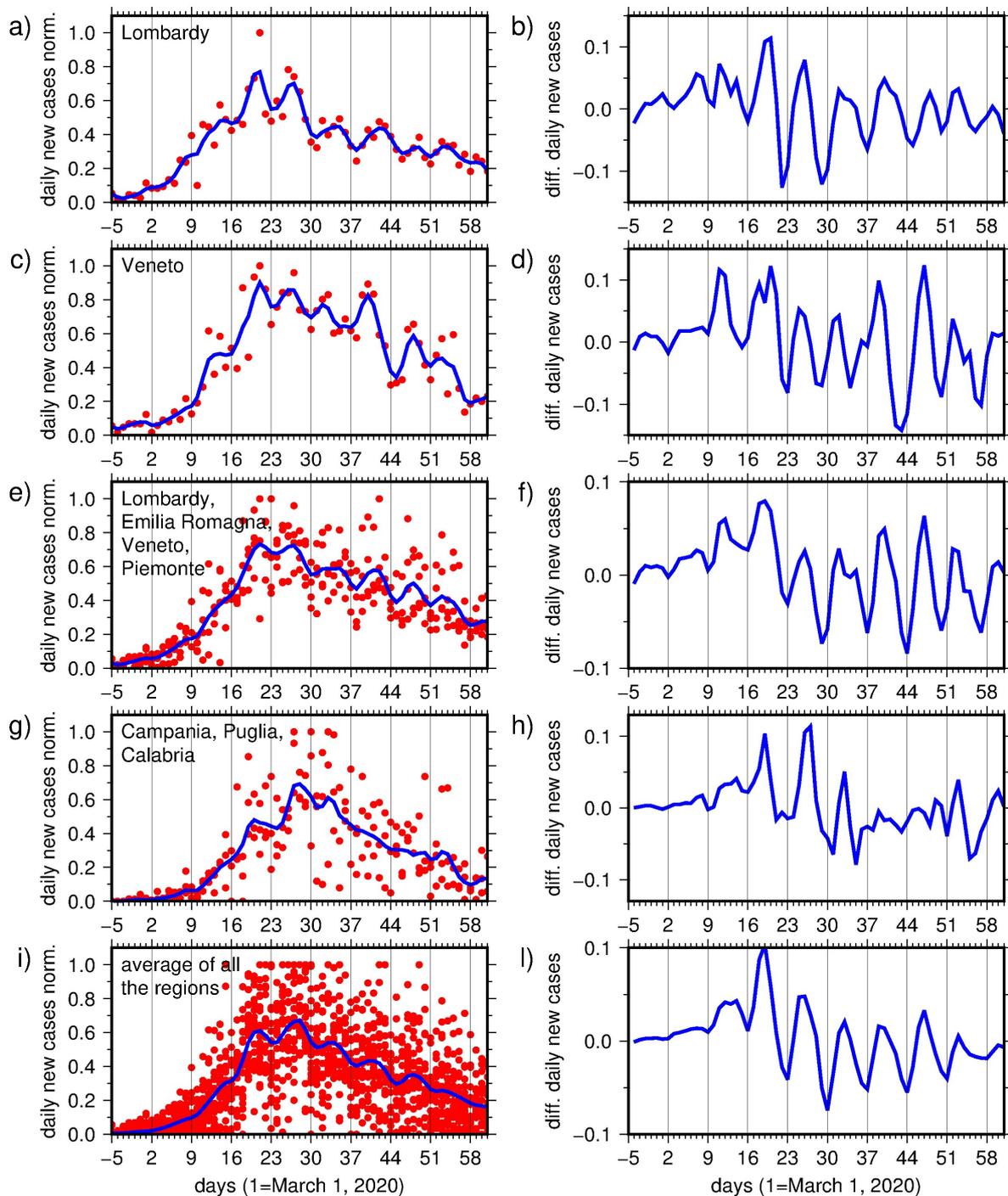


Figure 3. Time series considered in this study for some Italian regions and groups of them. Panels on the left: daily new Covid-19 cases (normalized values, red points) and the corresponding smoothed curve estimated by kernel regression (X_{NS} in equation 2). Panels on the right: time series of the first differences of the smoothed curve on the left (dX_{NS} in equation 3).

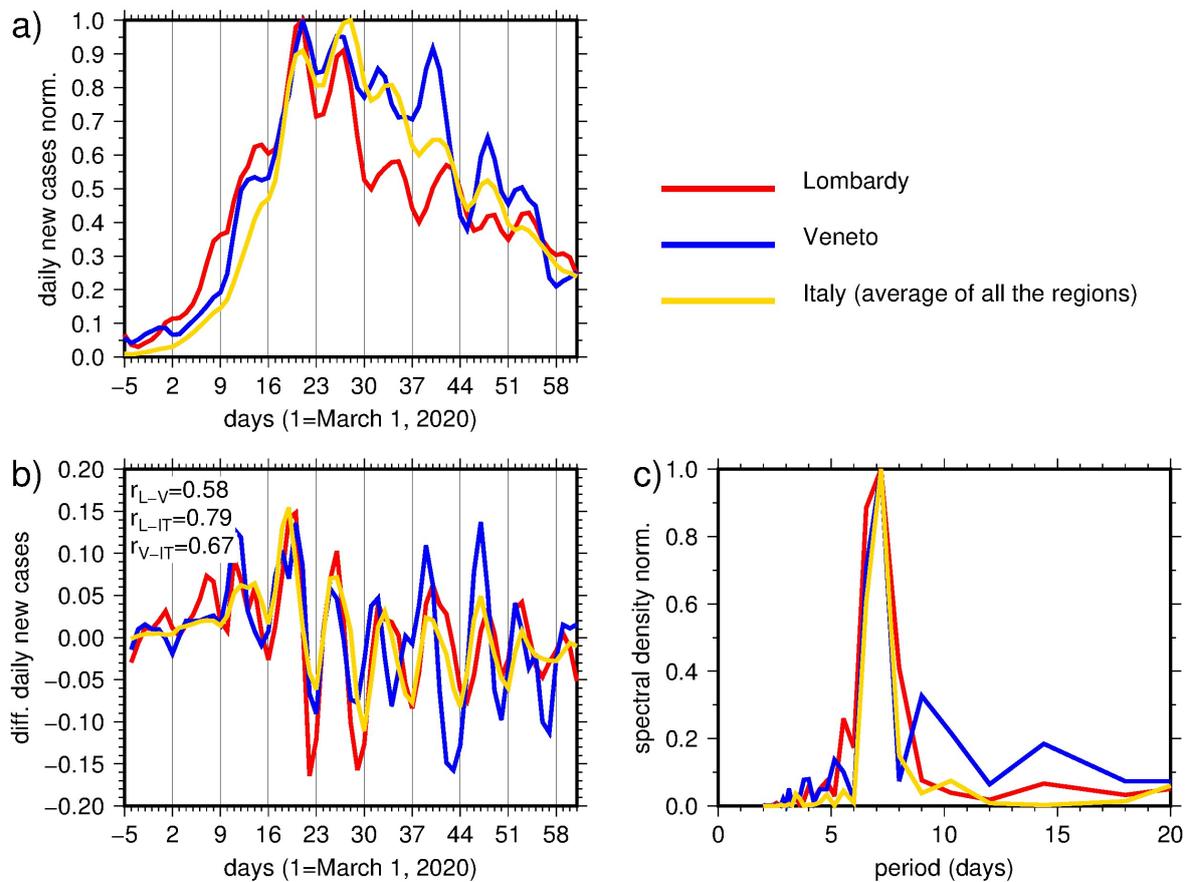


Figure 4. Time evolution of the Covid-19 epidemic in the Italian regions Lombardy and Veneto compared with the average behavior of all the regions. a) smoothed normalized time series of daily new cases (equation 2); b) time series of the first differences of the smoothed curves (equation 3); c) power spectra density of the time series in (b). In (b) the Pearson correlation coefficient for couples of time series is also reported. The vertical grids marks the Monday of each week.

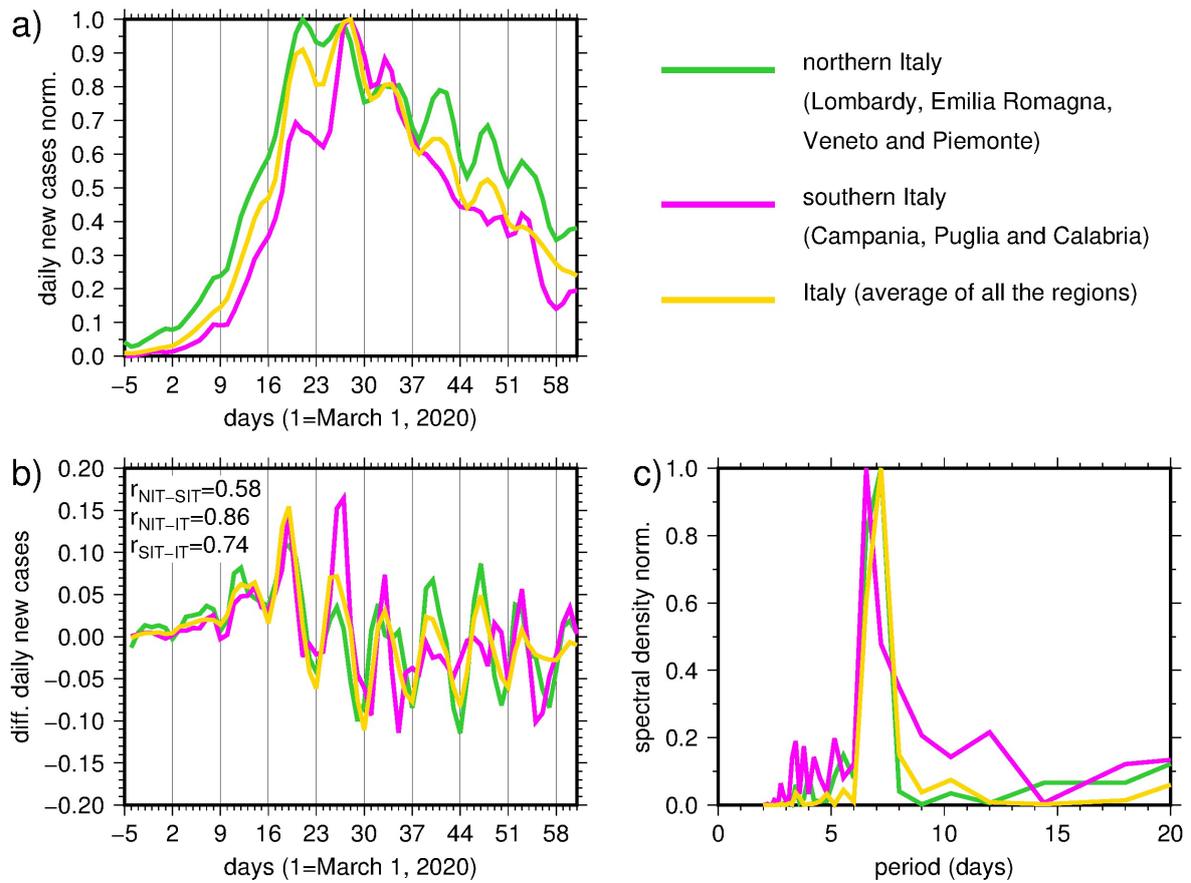


Figure 5. Time evolution of the Covid-19 epidemic for groups of regions in northern and southern Italy compared with the average behavior of all the regions. The meaning of the graphs is the same as in Fig.4.

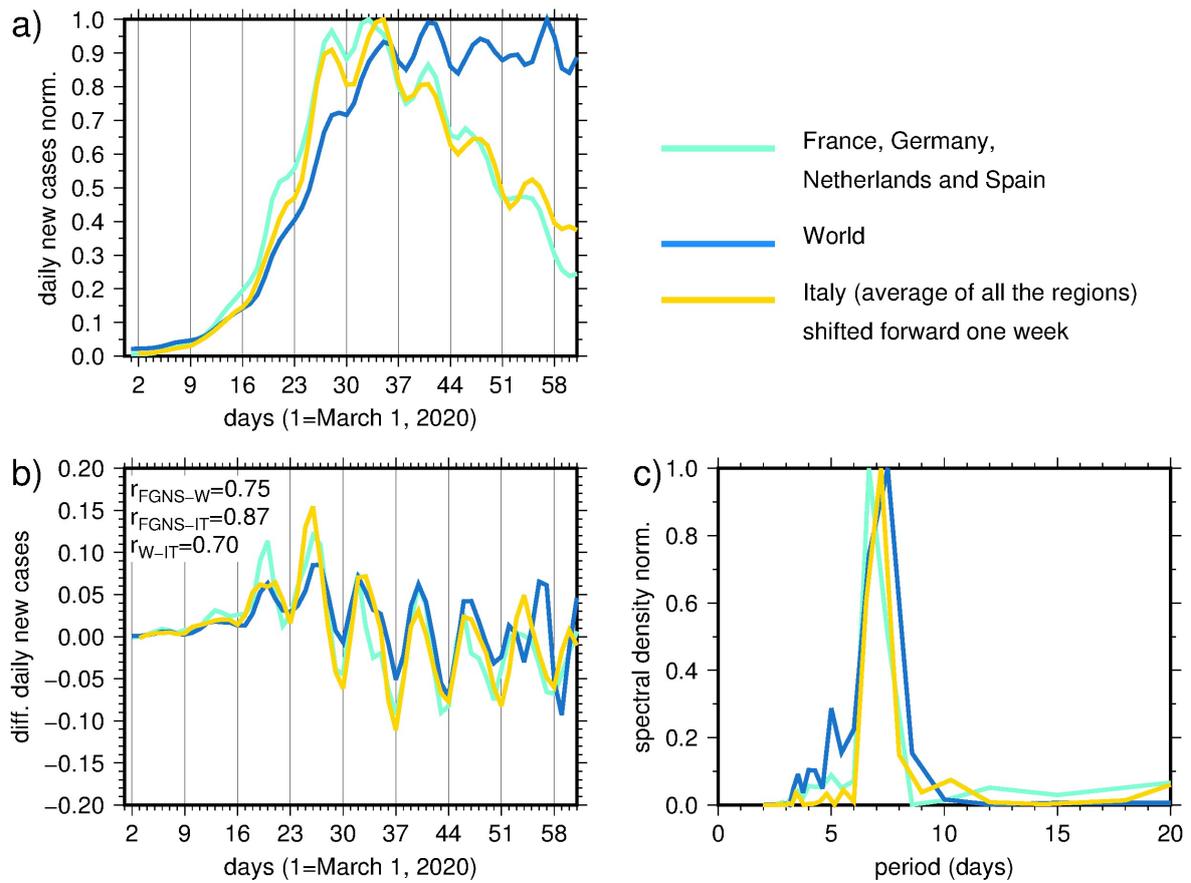


Figure 6. Comparison between the time evolution of the Covid-19 epidemic in Italy, a group of European countries and worldwide. The meaning of the graphs is the same as in Fig.4. The time series of Italy is shifted forward one week.

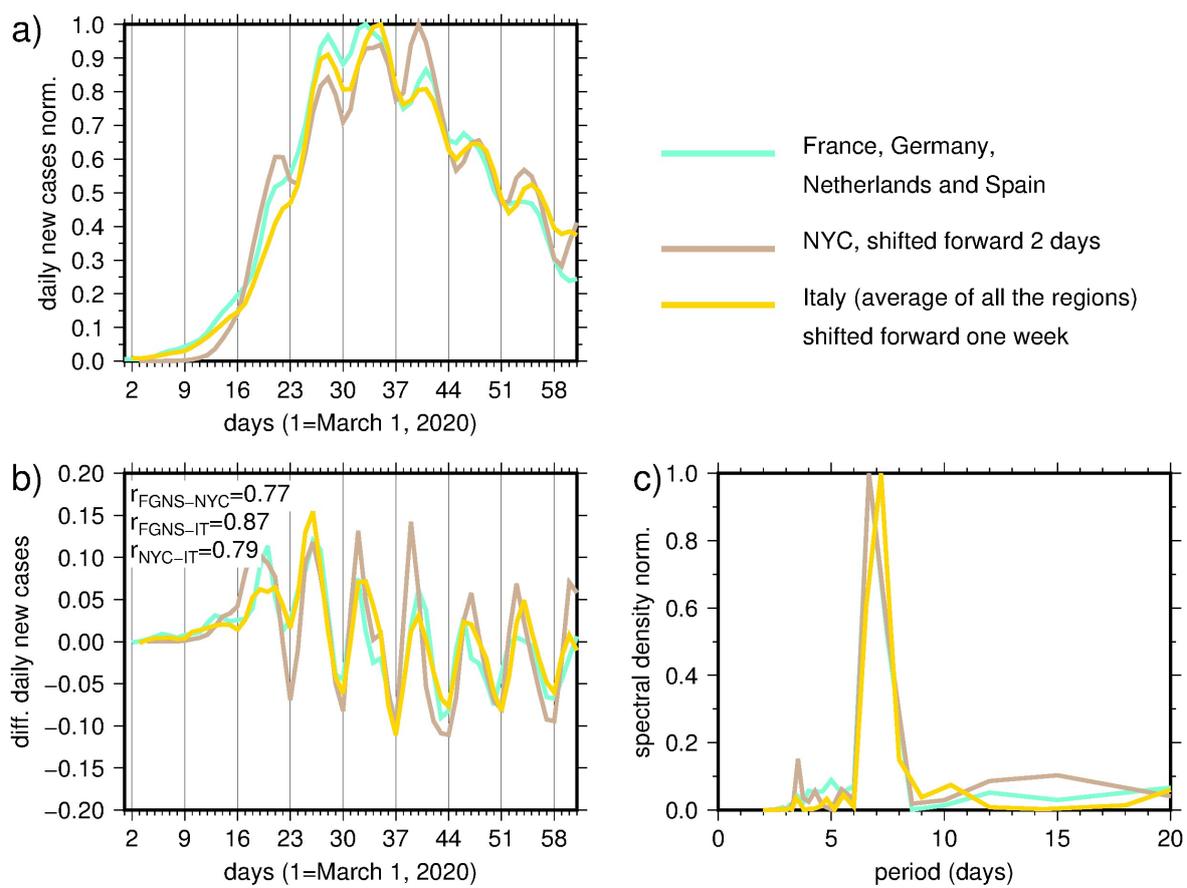


Figure 7. Comparison between the time evolution of the Covid-19 epidemic in Italy, a group of European countries and New York City. The meaning of the graphs is the same as in Fig.4. The time series of Italy and NYC are shifted forward 7 and 2 days, respectively.

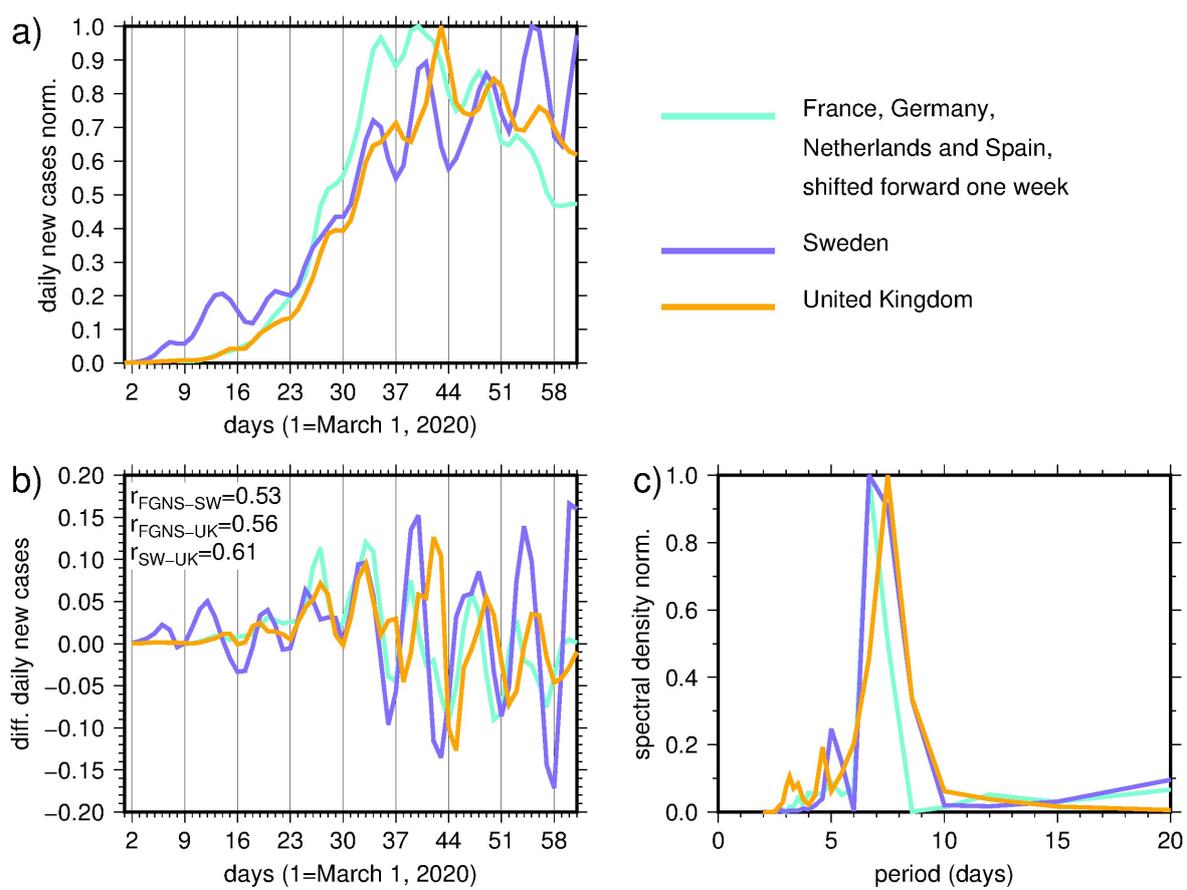


Figure 8. Comparison between the time evolution of the Covid-19 epidemic in a group of European countries, in the United Kingdom and in Sweden. The meaning of the graphs is the same as in Fig.4. The time series of the group of European countries is shifted forward 7 days.