The short infectious period of COVID-19 indicates the efficiency of fast switching periodic protocols

Eugenio Lippiello (eugenio.lippiello@unicampania.it)
University of Campania https://orcid.org/0000-0003-1444-1889
Lucilla de Arcangelis
University of Campania
Giuseppe Petrillo
University of Campania "Luigi Vanvitelli"

Article

Keywords: Quantification of Transmissibility, Public Health Responses, Generation Times, Numerical Simulations, Fast Periodic Switching, Alternating Quarantine

Posted Date: March 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-359482/v1

The short infectious period of COVID-19 indicates the efficiency of fast switching periodic protocols

Lippiello E. [1], de Arcangelis L. [2], Petrillo G. [1]

[1] *Department of Mathematics and Physic, University of Campania "L. Vanvitelli", 81100 Caserta, Italy.*

[2] *Department of Engeenering, University of Campania "Luigi Vanvitelli", 81100, Caserta, Italy*

Correspondence to E.L. (email: eugenio.lippiello@unicampania.it).

Abstract The Covid-19 disease pandemic is showing the importance of an accurate quantification of transmissibility in order to design and tune public health responses. Transmissibility is usually quantified in terms of the reproduction number R_t, the average number of secondary cases caused by an infected individual. Here we show the central role also played by $w(z)$, the distribution of generation times z, namely the time between successive infections in a transmission chain. We obtain an accurate estimate of $w(z)$ by means of a novel method which allows us to simultaneously obtain its evaluation together with the measure of R_t, over the course of an epidemic, and the number of exogenous infected cases. We use one year of data from COVID-19 officially reported cases in the 21 Italian regions, since the first confirmed case on February 2020. We find that $w(z)$ is a distribution very peaked around its average value $\overline{z} \simeq 6$ days with a standard deviation σ smaller than one day. This estimate of σ is much smaller than previous ones obtained by means of contact tracing from the distribution of serial intervals. We perform extended numer-

ical simulations to demonstrate that, because of the small value of σ, post-lockdown mitigation policies such as fast periodic switching and/or alternating quarantine can be extremely efficient.

Introduction

An accurate estimate of transmission parameters is fundamental in monitoring the spreading of a disease during a pandemic [1–9]. The coronavirus disease 2019 (COVID-19) pandemic has shown the relevance of an accurate evaluation of the time-dependent reproduction number R_t to monitor the effect of containment measurements imposed by local governments. R_t values are usually estimated from mathematical models and, commonly, values $R_t > 1$ are interpreted as an indicator that the infection will be able to start spreading in a population, which does not occur if $R_t < 1$. Its estimate depends on the model used and on the other quantities present in the model, as the probability distribution $w(z)$ of the generation time z, i.e. the time difference between the dates of infection of successive cases in a transmission chain. The distribution $w(z)$ is usually obtained by means of a probabilistic reconstruction of transmission trees. However this is a very difficult task for two different reasons: Firstly, the information about who infected whom is usually not available, particularly when the percentage of asymptomatic infections is large. Secondly, the times of infection are rarely known. To overcome this latter difficulty $w(z)$ is usually approximated by the distribution of the serial intervals, which is the difference in dates of symptom onset between a pair of a primary and its secondary case. Taking into account the heterogeneity of the timing of symptom onsets, as well as, all the drawbacks related to the uncertainty in the identification of

the correct infector-infected pair, the fitted distribution of serial intervals is expected to be broader than the "true" $w(z)$ [10,11]. More precisely we expect that the analysis of serial intervals provides a reasonable estimate of the average value $\overline{z}$ of $w(z)$ but overestimates its standard deviation σ. Data collected for COVID-19 in different geographic areas indicate a mean value of serial intervals ranging from 4 to 7.5 days, with a standard deviation ranging between 3 days up to 3 weeks inside each area [12–14]. We remark that a correct estimate of σ is important to correctly tune the optimal time for quarantine which should be at least larger than $\overline{z} + \sigma$ and also to plan optimized lock-down strategies. A first attempt to extrapolate, directly from the infection data, the parameters controlling the functional form of $w(z)$ can be found in Wallinga & Teunis (WT) [15]. In their seminal paper WT propose a method based on the maximization of the likelihood L formulated in terms of $w(z)$ and of the infection network, in which the nodes represent cases and the directed edges between the nodes represent transmission of infection between cases. White & Pagano (WP) [16] have subsequently proposed a simpler formulation for L, expressed only in terms of the daily series of incidence rate $I(t)$, i.e. the rate of people infected at the calendar time t. Their study, however, is restricted to the evaluation of $w(z)$ in the case of a constant reproduction number $R_t = R_0$.

Here we present a generalization of the WP approach which allows us to extract, simultaneously, from $I(t)$ the temporal evolution of $R_t(t)$ together with the more appropriate functional form of $w(z)$ consistent with the epidemic curve. More precisely, we are able to explore the whole range of parameters controlling $w(z)$ and to identify the optimal ones. Our procedure also provides as an

output the daily number of exogenous infected cases $\mu(t)$, which can be implemented in numerical simulations to test the efficiency of the procedure. We discuss in details the results of our approach using data for the COVID-19 disease in Italian regions, as well as the influence of periodicity in the collection of epidemic data. Finally, we enlighten the relevance of the information contained in $w(z)$ in order to design alternative strategies for virus containment in absence of vaccines.

Basic Definitions

The basis of epidemic models can be usually identified in the renewal equation [1,3]

$$I(t) = \int_{-\infty}^{t} ds\beta(t,s)I(s) + \mu(t) \tag{1}$$

where $\beta(t,s)$ is the transmissibility, defined in such a way that $\beta(t,s)dt$ is the average number of people someone, infected at the calendar time s, infects during the subsequent time interval $[t, t+dt)$, where dt is a small time. Under stationary conditions $\beta(t,s)$ is time translationally invariant and it is usually written as $\beta(t,s) = R_0 w(t-s)$, where R_0 is the reproduction number, defined as the average number of secondary cases per primary case. Conversely, when the contact rate changes in time, $\beta(t,s)$ explicitly depends on both times and one is compelled to introduce an effective reproduction number $R_t(t)$, which is commonly written as $R_t(t) = \int_{-\infty}^{t} ds\beta(t,s)$. Under the assumption that the generation time distribution $w(z)$ remains stationary over time and therefore it depends only on the time difference $z = t-s$, one can write $\beta(t,s) = R_t(t)w(t-s)$ [3,17]. Therefore $R_t(t)$ specifically describes the changes in the contact rate and presents the advantage to be directly obtained from Eq.(1). At the same time it does not have a clear physical meaning and,

rigorously speaking, the value $R_t(t) = 1$ cannot be used as a threshold to discriminate between a spreading or a decaying epidemic. An alternative definition is provided by the cohort reproduction number [3,17] $R_s(s) = \int_s^\infty dt\beta(t,s)$ which is the average number of secondary infections among a cohort experiencing an infection at time s. By definition, $R_s(s) = \int_s^\infty dt R_t(t)w(t-s)$ and therefore under stationary conditions $R_t(t) = R_s(t) = R_0$ whereas, when $R_t(t)$ changes in time, $R_s(t)$ can be viewed as the convolution of the functions $R_t(t)$ and $w(t-s)$. Therefore, in terms of Fourier transforms $\mathcal{F}(R_s(t),\omega) = \mathcal{F}(R_t(t),\omega)\,\mathcal{F}(w(t),-\omega)$, where $\mathcal{F}(x(t),\omega)$ is the Fourier transform of the function $x(t)$. In particular, when the distribution $w(z)$ is very peaked around its average value $\overline{z}$ one has $R_s(t-\overline{z}) \simeq R_t(t)$, which implies that a change in $R_t(t)$ at the time t becomes already visible in $R_s(s)$ at a previous time $s \simeq t - \overline{z}$.

The algorithms commonly adopted to estimate R_t or R_s [15,18] usually assume that $w(z)$ is already known and provided by the reconstruction of the transmission tree. Since $w(z)$ substantially depends only on biological parameters, as for instance the pathogen growth followed by immune suppression, it is expected to assume very small values, $w(z) \simeq 0$, for small time differences z and to decay exponentially, or even faster, after a characteristic time. For this reason $w(z)$ is usually approximated by a gamma, a Weibull or a normal distribution [3,10,12]. Next we present a novel method to extract $w(z)$ directly from data.

Log-likelihood evaluation

The only empirical parameter we consider is the daily incidence which is a discrete se-

ries $\{I(m)\}_{m=1,..N}$, where $I(m)$ represents the daily number of infected, recorded on the m-th day, and N is the number of considered days. As a consequence we also consider a discretized transmissibility $\beta(m,j)$, representing the average number of infections induced during the m-th day by people infected on the j-th day. Introducing the daily cohort number $R_s(m)$ such as $\beta(m,j) = R_s(j)w(m-j)$, where $w(m) = \int_m^{m+1} w(z)dz$ is the discretized generation time distribution, the renewal Eq.(1) gives the average value $E[I(m)]$ of the daily incidence on day m

$$E[I(m)] = \sum_{j=0}^{m-1} R_s(j)w(m-j)I(j) + \mu(m) \tag{2}$$

where $\mu(m) = \int_m^{m+1} \mu(t)dt$ is the daily number of imported cases during the m-th day. A similar equation can be written in terms of the effective reproduction number $R_t(m)$, however we prefer to consider $R_s(m)$, more suitable for numerical implementation. Indeed, if the discrete series $\{R_s(m)\}_{m=1,..,N}$ and $\{\mu(m)\}_{m=1,..,N}$, as well as the functional form of $w(m)$ are assigned, one can easily simulate the epidemic curve $\{I(m)\}_{m=1,..,N}$ according to a generation tree algorithm (see Methods).

In the generation process the number of individuals infected on the m-th day is assumed to be Poisson distributed, $P_P[I(m)] = \frac{E[I(m)]^{I(m)}e^{-E[I(m)]}}{I(m)!}$, with the expected value $E[I(m)]$ obtained from Eq.(2). The likelihood of the time series $\{I(m)\}_{m=1,...,N}$, for assigned sequences $\{R_s(m)\}_{m=1,..,N}$, $\{\mu(m)\}_{m=1,..,N}$ and for a given functional form of $w(m)$ is given by $L(\{I\},\{R_s\},\{\mu\},\{w\}) = \prod_{m=1}^{N} P_p[I_m]$. The best series $\{R_s\},\{\mu\},\{w\}$ compatible with the recorded data $\{I\}$ are the ones that maximize the likelihood. We perform this maximization process assuming that the functional form of $w(m)$ is assigned and depends on few tuning parameters. In particular, we consider

the two cases of $w(z)$ either a gamma or a Weibull distribution (see Methods). In both cases $w(z)$ is fully characterized by its average value $\overline{z}$ and the standard deviation σ, and the search for their optimal values is among the purposes of our approach. This leads to an expression for $L\left(\{I\},\{R_s\},\{\mu\},\overline{z},\sigma\right)$ which is equivalent to the one obtained by WP [16], except for the fact that we keep explicitly into account the temporal dependence of $R_s(m)$. Furthermore, we introduce a smoothness constraint on $R_s(m)$, by penalizing its second difference [19], in order to impose that $R_s(m)$ is a smooth function which does not change abruptly between two subsequent days. The final step in our approach is to consider the logarithm of the likelihood $LL = \log\left(L\right)$ which allows us to split L into the sum of different terms which can be more easily evaluated, thus providing a more efficient maximization procedure. The final expression for LL is given by

$$LL\left(\{I_m\},\{R_s\},\{\mu\},\overline{z},\sigma,V\right) = \sum_{m=1}^{N} I_m \log\left(E(I_m)\right) - \sum_{m=1}^{N} E(I_m) - \frac{1}{2V}\sum_{m=2}^{N-1}\left(R_s(m-1)+R_s(m+1)-2R_s(m)\right)^2 \tag{3}$$

where V is the parameter that controls the degree of smoothness of $R_s(m)$. In the Methods we present an optimized procedure based on the Markov-chain-Monte-Carlo method to find the best series $\{R_s\}$ and $\{\mu\}$ corresponding to the maximum of the log-likelihood LL [20,21]. Our algorithm is sufficiently fast that we can obtain the optimal LL for the full range of relevant values of $\overline{z}$ and σ.

It is important to clarify the difference between the standard deviation σ of $w(z)$ and the standard deviation of the probability distribution $\psi(\Delta t_{rec})$ of Δt_{rec}. Here Δt_{rec} is the temporal

distance between the day of the infection and the day when this infection is identified and recorded in the data set. Usually $\psi(\Delta t_{rec})$ is broadly distributed with a standard deviation of some days. However, as shown in the Methods, $\psi(\Delta t_{rec})$ does not affect the estimated value of σ obtained via the log-Likelihood maximization procedure.

Results: The Reproduction number R_s and the optimal $w(z)$

For each Italian region we consider the series of the daily number of new infected $\{I\}$ and the daily number of new tested people $\{n^T\}$. In the main text we present results for the region Lombardy where the first outbreak of Covid-19 has been documented in Europe and which is characterized by a widespread diffusion of the disease since March 2020. The same analysis has been performed for all the other Italian regions and in the Supplementary Information (SI) we present results for other five more populated regions (Lazio, Campania, Sicily, Veneto, Emilia-Romagna) .

In Fig.1 we plot $\{I\}$ and indicate with vertical lines the starting time of government's restrictions, which combines lockdown restrictions and closures with measures such as testing policy and contact tracing, etc. Fig.1 clearly shows the presence of two waves in the disease spreading with the maximum of infection reached on March 24 2020, during the first wave, and on November 10 2020, during the second one. Here we present results assuming that $w(z)$ is a gamma distribution, $w(z) = (\tau^{-a}/\Gamma(a) z^{a-1}) \exp(-z/\tau)$, which depends on two parameters, $a \geq 1$ and $\tau > 0$, and where $\Gamma(a)$ is the gamma function. The mean value of $w(z)$ is given by $\overline{z} = a\tau$ and its variance

$\sigma = a\tau^2$. In the SI (Fig.Supp.3) we show that a Weibull distributed $w(z)$ produces similar results.

We use the data plotted in Fig.1 to extract the information about $\{R_s\}$, $\{\mu\}$ and $w(z)$ according to our maximization procedure of $LL\left(\{I\}, \{R_s\}, \{\mu\}, \overline{z}, \sigma, V\right)$, exploring in detail a wide range of $\overline{z}$ and σ values. In Fig.2 we plot the temporal variation of $R_s(m)$ and $\mu(m)$, which have been obtained for $\overline{z} = 6.2$ and $\sigma = 0.9$ corresponding to a maximum of LL (Fig.3) and therefore representing an optimal description of the recorded sequence $\{I\}$. In order to verify the efficiency of our procedure we implement these optimal series $\{R_s\}$, together with the optimal choices of $\overline{z}$ and σ, in the generation tree algorithm (Methods). In the algorithm we also implement the optimal series $\{\mu\}$ extracted from the LL maximization and therefore there is only one free parameter $I(0)$, representing the initial value of infected people on the 0-th day. In Fig.1 we show that the numerical sequence $\{I\}$ simulated via the generation tree algorithm very well overlaps with the experimental one. As a further support we also compare our findings for $R_s(m)$ with the one provided by the Wallinga-Teunis (WT) algorithm (see Eq.(5) in Methods) [15]. We observe (Fig.2) that the two approaches provide results in good agreement. The temporal evolution of $R_s(m)$ observed in Fig.2 is consistent with its expected dependence on the contact rate. We indeed find a clear decrease of $R_s(m)$ after the application of strong confinement measures (after red lines) with a weak tendency to an increase after the removal of these measures. It is interesting to observe a clear peak of $R_s(m)$ in the middle of August which reflects the intense social activity typical of Italian summer vacation at the turn of the Assumption (August 15th). In the lower panel of Fig.2 we plot the weekly average value of $\mu(m)$ which fluctuates around an average value $\mu \sim 15$, with no par-

ticularly interesting trend. The estimate of $\mu(m)$ is however a necessary information to implement in the generation tree algorithm (see Methods) to simulate the synthetic sequence $\{I\}$. This allows a comparison with the recorded one in order to verify the accuracy of the inversion procedure.

An interesting result is provided by Fig.3a where we plot $LL\left(\{I\},\{R_s\},\{\mu\},\overline{z},\sigma,V\right)$ as function of $\overline{z}$. The equivalent of Fig.3a for a Weibull distributed $w(z)$ is presented in Fig.Supp.3a in SI. We find that LL, at fixed τ, presents a non-monotonic behavior with two maxima which are more evident for small τ values. The position $\overline{z}$ of the absolute maximum appears to be quite independent of τ. We wish to stress that the presence of local maxima makes very complicated the identification of the optimal a and τ by means of standard methods for LL maximization [22]. Automatic procedures can be indeed trapped in a local maximum. Conversely, our fast LL evaluation allows us to explore a wide range of a and τ values, except for $\tau < 0.03$ which are not numerically accessible because of the divergence of $\Gamma(a)$ when $a \gg 1$. Considering data at fixed $\overline{z}$ we notice that LL tends to decrease for increasing τ, i.e. increasing σ. The main result is that the optimal value (maximum value of LL) is obtained for $\overline{z} = 6.2$ days and $\tau = 0.03$, corresponding to $\sigma = 0.2$ days. A previous estimate $\overline{z} = 6.6$ days was provided for Lombardy from the distribution of serial intervals, extracted from the analysis of contact tracing data [13,14]. At the same time, the weighted average value from the 12 studies of serial intervals from Asiatic regions analyzed in [12] gives $\overline{z} = 5.5$ days. We find that the different estimates of $\overline{z}$ are in reasonable agreement whereas important differences are found in the σ values. For instance serial intervals in Lombardy give $\sigma = 23.8$ days [13,14] which is much larger than our estimate $\sigma = 0.2$ days. We remark that a similar

estimate of σ is obtained for all 21 Italian Regions (see SI).

How to manage periodicity in data collection

We next explore if the very small value of σ obtained from the LL maximization could be affected by the weekly periodicity in the testing procedure, since a smaller number of tests are performed during the weekend with respect to the working weekdays. This periodicity is clearly enlightened by the Discretized Fourier Transform (DFT) of $\{n^T\}$, plotted in Fig.4a, which clearly presents a peak at a frequency value equal to one, where time is measured in week units. The behavior of the DFT of $\{n^T\}$ at small frequencies can be conversely related to the periodicity caused by the two waves. Indeed, an increasing (decreasing) number of infected induces a larger (smaller) number of tests, leading to a correlation between the two signals $\{n^T\}$ and $\{I\}$. The same peak at small frequencies is indeed also found in the DFT of $\{I\}$ where a second smaller peak at $f = 1week^{-1}$ is still observed (Fig.4a). This second peak can be attributed to the weekly periodicity in the number of tests.

In the following we develop a simple argument to disentangle the daily number of infected $\{I\}$ from the daily test number $\{n^T\}$. More precisely we assume that the daily number of identified infected, during the m-th day, can be viewed as the sum of two contributions $I(m) = I^{(\phi)}(m) + I^{ran}(m)$. Here $I^{(\phi)}(m)$ is the "disentangled" incidence daily rate which can be interpreted as the daily number of infected people that should have been identified, independently of the number of tests performed on the m-th day, with the only condition $n^T(m) \geq I^{(\phi)}(m)$.

This number, for instance, includes infected with symptoms but also all people which have been in strict contact with them. The term $I^{ran}(m)$, conversely, represents the number of infected identified on the m-th day substantially by chance, according to a random search within a population N_P after performing $n^T(m)$ tests. Indicating with $I^{TOT}(m)$ the total number of new infected individuals during the m-th day, one has $I^{ran}(m) = n^T(m)I^{TOT}(m)/N_P$. This leads to $I(m) = I^{(\phi)}(m) + I^{TOT}(m)n^T(m)/(N_P\phi_1)$, where the coefficient $\phi_1 < 1$ has been introduced to take into account that the search is not fully random but it is usually focused on a subset $N_P\phi_1$ of the total population, which has been more exposed to the infection. Since it is reasonable to expect that $I^{(\phi)}(m)$ represents a fixed fraction $\phi_2 < 1$ of the total number of infected, $I^{(\phi)}(m) = \phi_2 I^{TOT}(m)$, we finally obtain

$$I^{(\phi)}(m) = \frac{I(m)}{1 + \frac{n^T(m)}{\phi N_P}} \tag{4}$$

where $\phi = \phi_1\phi_2$ is a parameter which can be fixed by imposing that $\{I^{(\phi)}\}$ is not causally related to $\{n^T\}$. More precisely we observe that (inset of Fig.4a) the Pearson's correlation coefficient between the temporal series $\{I^{(\phi)}\}$ and $\{n^T\}$, $\rho\left(\{I^{(\phi)}\}, \{n^T\}\right)$, is a monotonic increasing function of ϕ. We therefore identify the optimal threshold ϕ^* for decorrelation by imposing $\rho\left(\{I^{(\phi)}\}, \{n^T\}\right) = \rho\left(\{I^*\}, \{n^T\}\right)$. Here $\{I^*\}$ is the temporal series obtained by randomly reshuffling the original series $\{I\}$ in such a way that $I^*(m) = I(m^*)$, where $m^* = m + k^*$ and k^* is an integer random number uniformly distributed in the interval $[-3 : 3]$. By construction $\{I^*\}$ cannot present the weekly periodicity of $\{I\}$, as confirmed by its DFT (Fig.4a). The condition $\rho\left(\{I^{(\phi)}\}, \{n^T\}\right) = \rho\left(\{I^*\}, \{n^T\}\right)$. therefore allows us to obtain the series $\{I^{(\phi)}\}$, after setting

$\phi = \phi^*$ in Eq.(4), whose daily variation is uncorrelated to the weekly periodicity of $\{n^T\}$. In Fig.3b we plot $LL^{(\phi)} = LL\left(\{I^{(\phi)}\}, \{R_s\}, \{\mu\}, \overline{z}, \sigma, V\right)$ as function of $\overline{z}$, for different values of τ and $\phi = \phi^*$. This figure still shows the presence of a maximum at $\overline{z} = 6.1$ and interestingly, at fixed $\overline{z}$, $LL^{(\phi)}$ non monotonically depends on τ, with the largest value reached for $\tau = 0.15$. This leads to $\sigma = \tau\overline{z} = 0.9$ days which represents a more reasonable estimate than the smaller value suggested by Fig.3a. This is an indication that $\{I^{(\phi)}\}$ is not affected by the periodicity in data collection and provides a more accurate estimate of the parameters a and τ controlling $w(z)$. Similar results are obtained for the majority of Italian regions (see SI) even if, in few cases, the value of ϕ^* is so small to hide the information contained in $\{I^{(\phi)}\}$ leading to a less significant $LL^{(\phi)}$ (see Fig.Supp.6d, as an example).

Fast switching protocols

The importance of an accurate estimate of $R_s(m)$ to monitor public health responses is nowadays widely recognized. Here we stress also the relevance of a correct estimate of the parameters controlling $w(z)$. More precisely, we discuss the influence of σ in fast periodic switching protocols [23] where periods of stringent lockdown alternate with periods when only weak social constraints are imposed. The same consideration applies to the protocol of alternating quarantine [24], where the population is subdivided in two non interacting subsets, each one subject to a fast periodic switching protocol in phase opposition: While one subset is in full lockdown, the other subset has quite regular activity. These protocols can have the advantage to suppress the virus outbreak, while

at the same time they allow for continued economic activities. More precisely, we take as reference value $R_s(m) = 3$ measured in Lombardy at the beginning of October 2020, in a temporal period where substantially all the activities, including schools, were open. We therefore assume that during an interval of duration T_{NLD} the reproduction number $R_t(m)$ assumes the constant value $R_t(m) = 3$, whereas in the subsequent period of duration T_{LD} a rigid lockdown decreases $R_t(m)$ to very low values. It is quite intuitive that the fast periodic switching protocols can be efficient only if $T_{LD} > \overline{z} + \sigma$, so that an infected individual will have a low probability to be still infectious when s/he comes back to normal activity. To test this point we simulate the evolution of $I(m)$ over a temporal interval of one year, assuming a periodic $R_t(m)$ which alternates, each $T_{LD} = T_{NLD} = 2\overline{z}$, between the values of $R_t(m) = 3$ and $R_t(m) = 0.25$, leading to a periodic $R_s(m)$ (inset of Fig.5). We always assume a gamma distributed $w(z)$ with $\overline{z} = 6.1$ days, obtained from our analysis, and we consider two values of σ, $\sigma = 0.9$ obtained by our analysis and $\sigma = 18$ days estimated from serial intervals. Results, plotted in Fig.5, clearly show that for a sufficiently large σ, $I(m)$ presents fluctuations around an average exponential growth which is steeper for larger σ. Obviously, in this case the fast switching protocol does not work. Conversely, for $\tau = 0.15$ ($\sigma = 0.9$) the exponential growth is replaced by a much slower power law increase and $I(m)$ fluctuates around an average value which changes very slowly in time. We can finally conclude that only in the case of small σ fast switching strategies give the possibility to keep the disease spreading always under control leaving simultaneously enough space to normal activities.

Discussion and Conclusions

We have presented a detailed analysis of the log-likelihood $LL\left(\{I\},\{R_s\},\{\mu_m\},\overline{z},\sigma,V\right)$ for the daily incidence rate $\{I\}$ of COVID-19 in different italian regions, implementing the optimal time series $\{R_s\}$ and $\{\mu\}$ for a wide range of the parameters $\overline{z}$ and σ which control the shape of the generation time distribution $w(z)$. To our knowledge no similar analysis has been performed also for other viruses. The main conclusion is that the maximum of LL for COVID-19 in Italy is observed for values of the standard deviation σ of $w(z)$ much smaller than previous estimate. More precisely our results indicate that an individual remains infectious in a small temporal window between $[5:7]$ days, from the starting time of his/her infection, and s/he is not infectious outside this interval. This information can be useful to better characterize the biological processes promoting the transmission of the disease and in planning optimized strategies for mitigating the spreading of the virus. For instance, our study indicates that it is sufficient a quarantine period no longer than 8 days. Moreover, we have shown that it could be possible to keep the number of daily infected quite constant by means of a fast periodic switching protocol which alternates periods of 12 days of rigid lock-down with unconstrained periods. In this way it is possible to avoid the exponential growth of the virus spreading while keeping socio-economic continuity at about the 50% of maximum capacity.

Methods

The generation tree algorithm

We present a standard procedure for a self-exciting Branching Process where the time series

$I(m)$ is simulated according to a generation tree algorithm [25]. The first step is setting the number $I(0)$ of infected persons on the day 0 and use the daily number of imported infections $\mu(m)$ to obtain $n_0 = I(0) + \sum_{m=1}^{N} \mu(m)$, the number of infected people. Using the terminology of branching processes, this is the zero-th order generation and we index with $k_0 \in [1, n_0]$ each infector, defined as mother element, which generates a number $n(k_0)$ of off-springs, i.e. the newly infected elements. The number $n(k_0)$ of off-springs depends on $R_s(t(k_0))$ evaluated at the occurrence time $t(k_0)$ of their mother, according to a Poisson distribution $P_P(n(k_0))$ with the expected value $E(n(k_0)) = R_s(t(k_0))$. This is the first order generation containing $n_1 = \sum_{k_0=1}^{n_0} n(k_0)$ elements, each one infected at a time $t(k_1) = t(k_0) + m$, where m is a random time extracted according to the probability generating function $w(m)$. Only off-springs infected during the observational time window $t(k_1) \in [1, N]$ are considered. At the subsequent step $(j+1)$ the previous step is repeated considering as mother elements the n_j off-springs of the previous $(j-1)$ generation. In this way one obtains n_{j+1} new off-springs and the process is iterated up to the final generation j_f, such that $n_{j_f+1} = 0$. The numerical code is available for open access at https://github.com/Statistical-Mechanics-Group-Caserta/covid-maximum-loglikelihood-estimation

The log-likelihood maximization procedure

The algorithm assumes an initial trial value of $\{R_s^0\}$ and $\{\mu^0\}$. At each Monte-Carlo (MC) trial we randomly select a day $m' \in [1, N]$ and extract $\delta R = q_r r R_s(m')$ with r uniformly distributed between $[-1/2, 1/2]$ and $q_r \ll 1$. We evaluate $E(I_m)$ replacing in Eq.(2) $R_s(m') =$

$R_s(m')+\delta R$. The new value of $E(I_m)$ is used in Eq.(3) for the evaluation of the trial log-likelihood LL'. If $LL' > LL$ the new value of $R_s(m')$ is more consistent with data and its value is therefore updated, otherwise it is discarded. A similar procedure is applied to the series $\{\mu\}$ with the trial value $\mu(m') = \mu_(m') + q_\mu r$,with $q_\mu \ll 1$ accepted only if it leads to a larger value of $LL' > LL$ from Eq.(3). We complete a Monte-Carlo step when N trials have been performed. We notice that the new value $R_s(m')$ only affects the terms with $m > m'$ in the first sum in the rhs of Eq.(3) and therefore the number of operation in a MC step is of order $N \times N/2$ making the computation very fast with 5000 MC steps involving about 10 seconds of standard CPU time, when $N = 360$ days. Since $\{\mu\}$ only weakly affects LL, the evaluation is optimized by including the trials on $\{\mu\}$ only each 20 MC steps. We have verified that simulations evolve towards an asymptotic value, for a large number of MC steps, which is independent of the specific initial choice of $\{R_s^0\}$ and $\{\mu^0\}$. Indeed, only the initial value $R_s^0(0)$ is relevant since it remains constant during the simulation and therefore affects the other values of $R_s(m)$, because of the smoothness constraint (Eq.3). For this reason we extract this value by means of the Wallinga-Teunis (WT) method [15] which, by the definition of the cohort reproduction number $R_s(m) = \sum_{j=m}^{N} \beta(j,m)$, leads to

$$R_s(j) = \sum_{m=0}^{N} \frac{I(j+m)w(j)}{\sum_{l=0}^{N} I(j+l-m)w(l)}, \tag{5}$$

which is obtained implementing the definition $\beta(m,j) = R_t(m)w(m-j)$ and using $R_t(m) = E(I(m))/\sum_{j=1}^{N} w(m-j)I(j)$, given by Eq.2 after setting $\mu(m) = 0$. In particular, we use the initial value of $R_s^0(m)$ from Eq.(5) for $m \leq 3$, whereas for $m \geq 4$ we assume that $R_s^0(m)$ linearly decreases down to $R_s^0(m) = 0.1$ and then remains constant at larger m. We also fix

$\mu^0(m) = 0$ for all $m \in [1, N]$. Results do not depend on this initial choice. We have obtained a rough estimate of the minimum value of the parameter $V = 3.5/\sqrt{\sum_{m=1}^{N} I(m)}$ in Eq.(3) which makes fluctuations of $R_s(m)$ sufficiently smooth and we keep this value always fixed. We have verified that after 5000 MC steps the simulation reaches its asymptotic value, namely $\{R_s\}$ and $\{\mu\}$ do not vary appreciably for additional MC steps.The numerical code is available for open access at https://github.com/Statistical-Mechanics-Group-Caserta/covid-maximum-loglikelihood-estimation

The independence between $w(z)$ and $\psi(\Delta t_{rec})$

We indicate with $I^{(true)}(m)$ the number of infections which have been truly occurred the m-th day. This infection, however, are identified and reported only subsequent days. More precisely, according to the definition of $\psi(\Delta t_{rec})$, the $I^{(true)}(m)$ infections leads to a incidence daily rate of recorded infections

$$I(m) = \sum_{n=-\infty}^{\infty} I^{(true)}(n)\psi(m-n) \tag{6}$$

where we extended the sum over n from $(-\infty, \infty)$ assuming that $I^{(true)}(m) = 0$ if m is smaller than the first detection day of and also $\psi(j) = 0$ if $j \leq 0$. The renewal Eq.(2) is expected to hold for $I^{(true)}(m)$

$$E[I^{(true)}(m)] = \sum_{j=-\infty}^{m-1} R_s(j)w(m-j)I^{(true)}(j) + \mu(m) \tag{7}$$

but we will show that it also holds for $I(m)$. Inserting Eq.(7) in Eq.(6) we obtain

$$E[I(m)] = \sum_{n=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Big(R_s(j)w(n-j)I^{(true)}(j) + \mu(n)\Big)\psi(m-n). \tag{8}$$

We have extended the sum over j to ∞ setting $w(j) = 0$ if $j \leq 0$ and this allows us to exchange the two sum, leading to

$$E[I(m)] = \sum_{k=-\infty}^{\infty} w(k) \sum_{j=-\infty}^{\infty} R_s(j) I^{(true)}(j) \psi(m-k-j) + \mu^{(rec)}(m), \tag{9}$$

with $\mu^{rec}(m) = \sum_{n=-\infty}^{\infty} \mu(n)\psi(m-n)$. We next assume that the time evolution of $R_s(j)$ is sufficiently slower such that it can be considered roughly constant during the time scale where $\psi(m-j-k)$ goes to zero

$$\begin{aligned} &\sum_{j=-\infty}^{\infty} R_s(j) I^{(true)}(j) \psi(m-k-j) \simeq \\ &R_s(m-k) \sum_{j=-\infty}^{\infty} I^{(true)}(j) \psi(m-k-j) = R_s(m-k) I(m-k), \end{aligned} \tag{10}$$

where in the last step we used Eq.(6). We finally obtain

$$E[I(m)] \simeq \sum_{k=-\infty}^{\infty} R_s(k) w(m-k) I(k) + \mu^{rec}(m) \tag{11}$$

which is the renewal equation for $I(m)$ which no longer depends on $\psi(\Delta t_{rec})$. In particular the generation time distribution $w(z)$ obtained from Eq.(11), for the daily recorded incidence rate $I(m)$, is the same controlling the evolution of $I^{(true)}(m)$ (Eq.(7)).

We consider data provided by Protezione Civile for the 21 Italian regions and collected in https://github.com/DavideMagno/ItalianCovidData. We consider the time series from February 24 2020 $(m = 1)$ up to February 24 2021 $(m = N)$ for global $N = 366$ days.

Data availability

1. Kermack, W. O., McKendrick, A. G. & Walker, G. T. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115**, 700–721 (1927). URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118. https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0118.

2. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Science Publications, Oxford, 2002).

3. Grassly, N. C. & Fraser, C. Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* **6**, 477 (2008).

4. Roberts, M. G. & Heesterbeek, J. A. P. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *J. Math. Biol.* **55**, 803 (2007).

5. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences* **274**, 599–604 (2007). URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2006.3754. https://royalsocietypublishing.org/doi/pdf/10.1098/rspb.2006.3754.

6. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences*

101, 6146–6151 (2004). URL https://www.pnas.org/content/101/16/6146. https://www.pnas.org/content/101/16/6146.full.pdf.

7. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015). URL https://link.aps.org/doi/10.1103/RevModPhys.87.925.

8. Liu, Q.-H. *et al.* Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences* **115**, 12680–12685 (2018). URL https://www.pnas.org/content/115/50/12680. https://www.pnas.org/content/115/50/12680.full.pdf.

9. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of covid-19. *Proceedings of the National Academy of Sciences* **117**, 16732–16738 (2020). URL https://www.pnas.org/content/117/29/16732. https://www.pnas.org/content/117/29/16732.full.pdf.

10. Svensson, A. A note on generation times in epidemic models. *Mathematical Biosciences* **208**, 300–311 (2007). URL https://www.sciencedirect.com/science/article/pii/S0025556406002094.

11. te Beest, D., Wallinga, J., Donker, T. & van Boven, M. Estimating the generation interval of influenza a (h1n1) in a range of social settings. *Epidemiology* **24**, 244–250 (2013).

12. Rai, B., Shukla, A. & Dwivedi, L. K. Estimates of serial interval for covid-19: A systematic review and meta-analysis. *Clinical Epidemiology and Global Health* **9**, 157–161 (2021). URL https://www.sciencedirect.com/science/article/pii/S2213398420301895.

13. Cintia, P. & et al. The relationship between human mobility and viral transmissibility during the covid-19 epidemics in italy. *arXiv preprint arXiv:2006.03141* **.**, . (2020). URL https://arxiv.org/abs/2006.03141.

14. Marziano, V. *et al.* Retrospective analysis of the italian exit strategy from covid-19 lockdown. *Proceedings of the National Academy of Sciences* **118** (2021). URL https://www.pnas.org/content/118/4/e2019617118. https://www.pnas.org/content/118/4/e2019617118.full.pdf.

15. Wallinga, J. & Teunis, P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. *American Journal of Epidemiology* **160**, 509–516 (2004). URL https://doi.org/10.1093/aje/kwh255. https://academic.oup.com/aje/article-pdf/160/6/509/179728/kwh255.pdf.

16. Forsberg White, L. & Pagano, M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Statistics in Medicine* **27**, 2999–3016 (2008). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.3136. https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.3136.

17. Fraser, C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS ONE* **2**, 1–12 (2007). URL https://doi.org/10.1371/journal.pone.0000758.

18. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology* **178**, 1505–1512 (2013). URL https://doi.org/10.1093/aje/kwt133. https://academic.oup.com/aje/article-pdf/178/9/1505/17341195/kwt133.pdf.

19. Hens, N. *et al.* Estimating the effective reproduction number for pandemic influenza from notification data made publicly available in real time: A multi-country analysis for influenza a/h1n1v 2009. *Vaccine* **29**, 896–904 (2011). URL https://www.sciencedirect.com/science/article/pii/S0264410X10006754.

20. Bottiglieri, M., Lippiello, E., Godano, C. & de Arcangelis, L. Comparison of branching models for seismicity and likelihood maximization through simulated annealing. *Journal of Geophysical Research: Solid Earth* **116**, n/a–n/a (2011). URL http://dx.doi.org/10.1029/2009JB007060. B02303.

21. Lippiello, E., Giacco, F., Arcangelis, L. d., Marzocchi, W. & Godano, C. Parameter estimation in the ETAS model: Approximations and novel methods. *Bulletin of the Seismological Society of America* **104**, 985–994 (2014). URL http://www.bssaonline.org/content/104/2/985.abstract. http://www.bssaonline.org/content/104/2/985.full.pdf+html.

22. Myung, I. J. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology* **47**, 90–100 (2003). URL https://www.sciencedirect.com/science/article/pii/S0022249602000287.

23. Bin, M. *et al.* Post-lockdown abatement of covid-19 by fast periodic switching. *PLOS Computational Biology* **17**, 1–34 (2021). URL https://doi.org/10.1371/journal.pcbi.1008604.

24. Meidan, D. *et al.* Alternating quarantine for sustainable epidemic mitigation. *Nat. Commun.* **12**, 220 (2021).

25. de Arcangelis, L., Godano, C., Grasso, J. R. & Lippiello, E. Statistical physics approach to earthquake occurrence and forecasting. *Physics Reports* **628**, 1 – 91 (2016). URL //www.sciencedirect.com/science/article/pii/S0370157316300011.

Author contributions E.L. conceived the study and developed the algorithm. E.L., L. de A. and G. P. contributed extensively to data analysis and to write the manuscript.

Competing Interests The Authors declare no competing interests.

Materials & Correspondence Correspondence and material requests can be addressed to E.L. (eugenio.lippiello@unicampania.it).

Acknowledgements E. L. and G. P. acknowledge support from project PRIN201798CZLJ. L. de A. acknowledges support from project PRIN2017WZFTZP. E.L. e L. de A. acknowledge support from VALERE

project $E - PASSION$ of the University of Campania “L. Vanvitelli”.

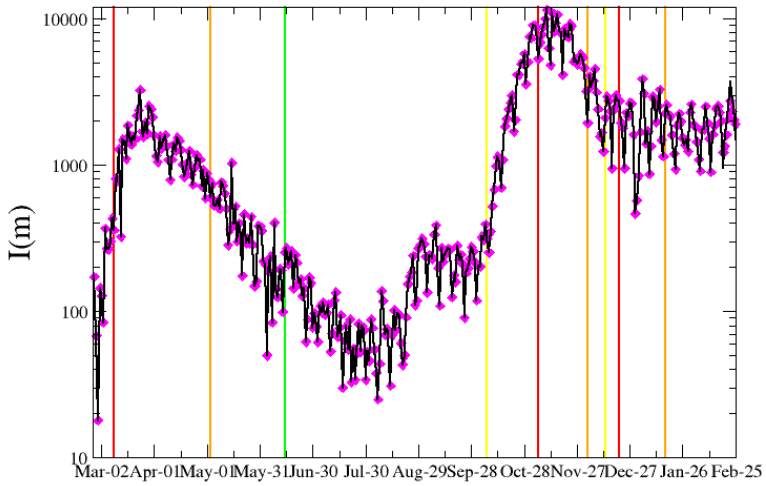

Figure 1: Black continuous lines represent the daily incidence of COVID-19 for the Lombardy from 02/24/2020 up to 02/24/2021. Color vertical lines indicate the starting time of different containment measures with a color code ranging from red, orange, yellow up to green as a rough indicator of the severity of these restrictions, decreasing from red to green, i.e strong restrictions are imposed in the temporal period after a red line whereas weak ones after a green one. Magenta diamonds represent the result of numerical simulations implementing the best estimate for $\{R_s\}$ and $\{\mu\}$, for $\tau = 0.15$, $a = 6.2\tau$, provided by the LL maximization procedure. The overlap with experimental data is achieved setting $I(0) = 250$.

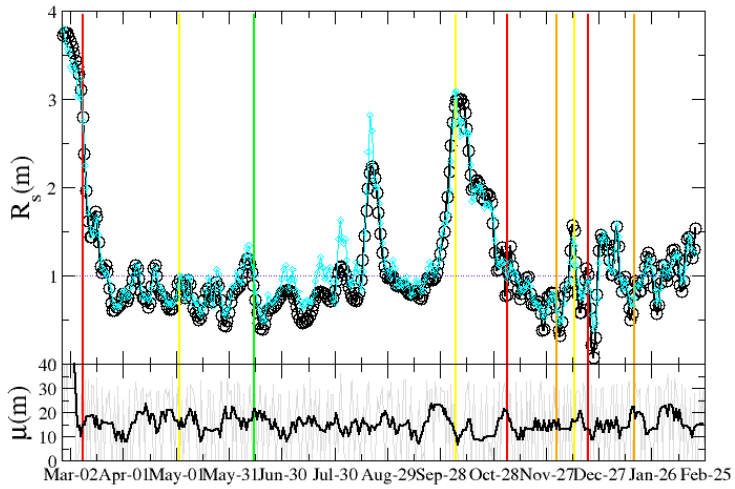

Figure 2: (Upper panel) The cohort reproduction number $R_s(m)$ of COVID-19 for the Lombardy from 02/24/2020 up to 02/18/2020. Color vertical lines indicate the starting time of different containment measures (see caption of Fig.1). Black circles represent $R_s(m)$ obtained by means of the log-likelihood maximization procedure whereas cyan diamonds are used for $R_s(m)$ estimated from the WT method (Eq.(5) in Methods). (Lower panel) The daily number of imported cases $\mu(m)$ estimated by the log-likelihood maximization procedure is plotted in thin grey whereas solid line is used for its weekly average.

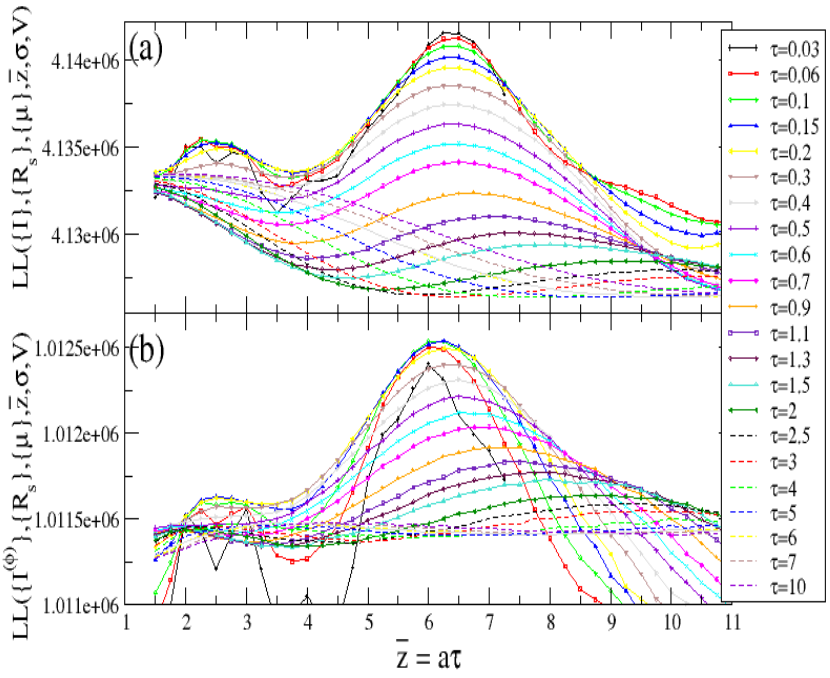

Figure 3: (Panel a) The log-likelihood $LL\left(\{I\},\{R_s\},\{\mu\},\overline{z},\sigma,V\right)$ obtained from the daily incidence of COVID-19 in Lombardy, is plotted as a function of $\overline{z}=a\tau$. Different curves correspond to different values of τ, which implies a different $\sigma=\tau\overline{z}$. (Panel b) As in the upper panel for $LL\left(\{I^{(\phi)}\},\{R_s\},\{\mu\},\overline{z},\sigma,V\right)$ with $I^{(\phi)}(m)$ given in Eq.(4) for $\phi=\phi^*=5E-4$.

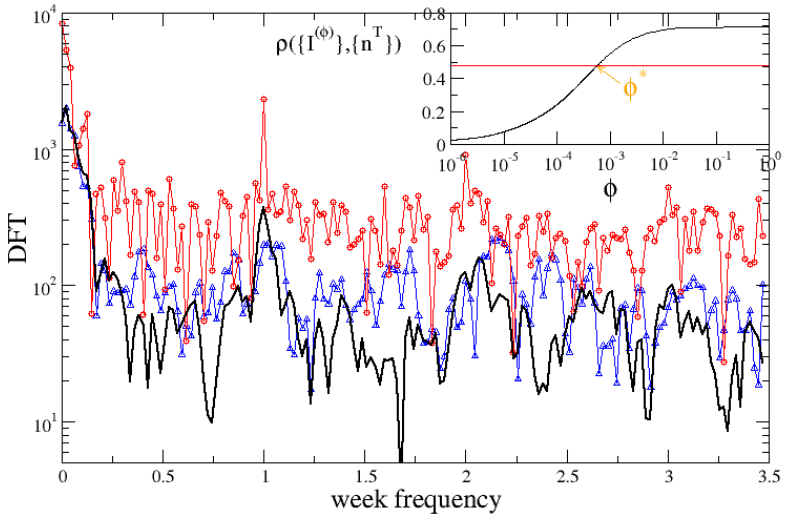

Figure 4: The DFT of the series $\{I\}$ (black lines), $\{n^T\}$ (red circles) and $\{I^*\}$ (blue triangles) as function of the frequency. Time is measured in week units such that the peak in 1 corresponds to a weekly periodicity. (Inset) The Pearson's correlation coefficient $\rho\left(\{I^{(\phi)}\},\{n_T\}\right)$ as function of ϕ. The horizontal red line represent the value $\rho\left(\{I^*\},\{n^T\}\right)$.

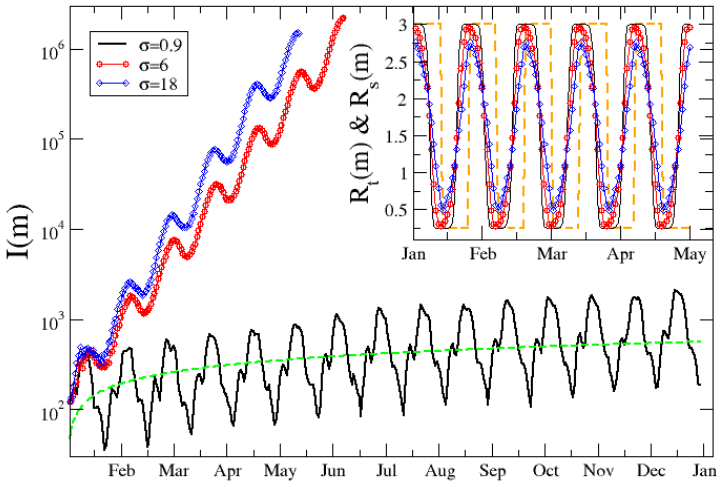

Figure 5: The daily incidence $I(m)$ expected for a periodic fast switching protocol which alternates period without constrains to rigid lockdown periods, each one lasting 12 days. We assume $I(0) = 1000$ at the starting time. Different colors and symbols are used for different values of σ in the probability distribution of generation time $w(z)$, which is gamma distributed. The green dashed line is a power law fit $m^{1/2}$ for the temporal evolution of the average value of $I(m)$ when $\sigma = 0.9$. (Inset) We plot with dashed orange line the evolution of $R_t(m)$ implemented in the numerical model. The other curves represent the evolution of $R_s(m)$ for different values of τ, with the same color code and symbols of the main panel.

Figures

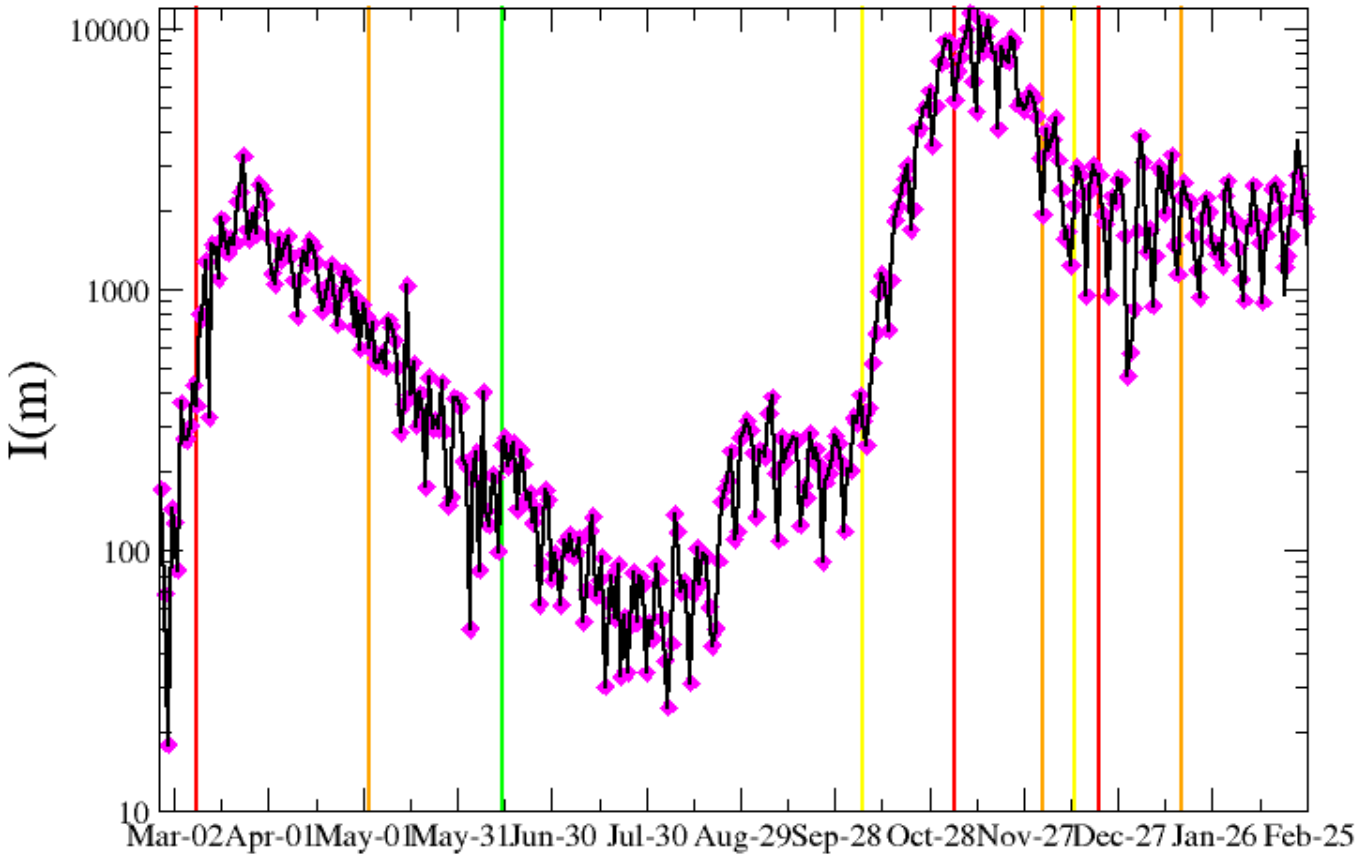

Figure 1

Black continuous lines represent the daily incidence of COVID-19 for the Lombardy from 02/24/2020 up to 02/24/2021. Color vertical lines indicate the starting time of different containment measures with a color code ranging from red, orange, yellow up to green as a rough indicator of the severity of these restrictions, decreasing from red to green, i.e strong restrictions are imposed in the temporal period after a red line whereas weak ones after a green one. Magenta diamonds represent the result of numerical simulations implementing the best estimate for {Rs} and {μ}, for $\tau = 0.15$, $a = 6.2\tau$, provided by the LL maximization procedure. The overlap with experimental data is achieved setting $I(0) = 250$.

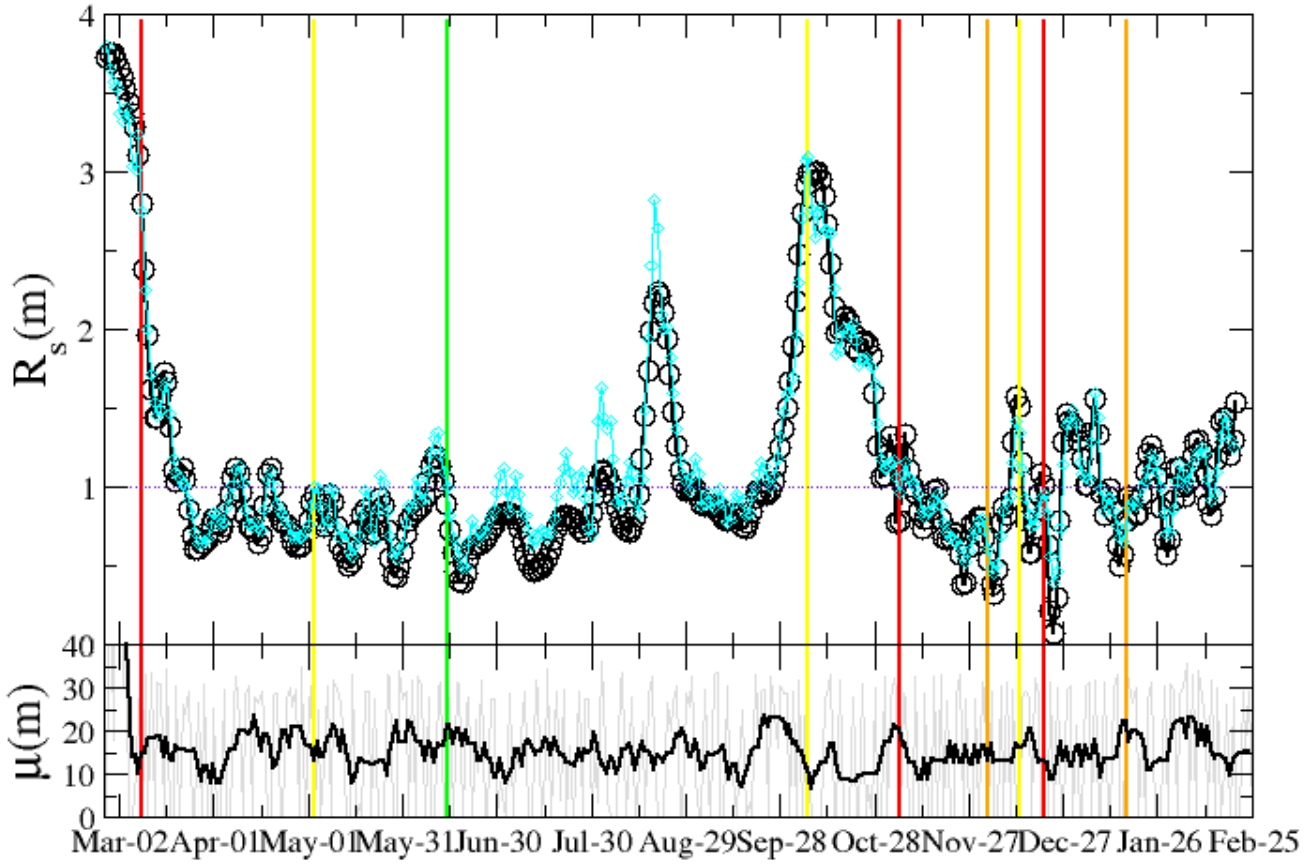

Figure 2

(Upper panel) The cohort reproduction number Rs(m) of COVID-19 for the Lombardy from 02/24/2020 up to 02/18/2020. Color vertical lines indicate the starting time of different containment measures (see caption of Fig.1). Black circles represent Rs(m) obtained by means of the log-likelihood maximization procedure whereas cyan diamonds are used for Rs(m) estimated from the WT method (Eq.(5) in Methods). (Lower panel) The daily number of imported cases μ(m) estimated by the log-likelihood maximization procedure is plotted in thin grey whereas solid line is used for its weekly average.

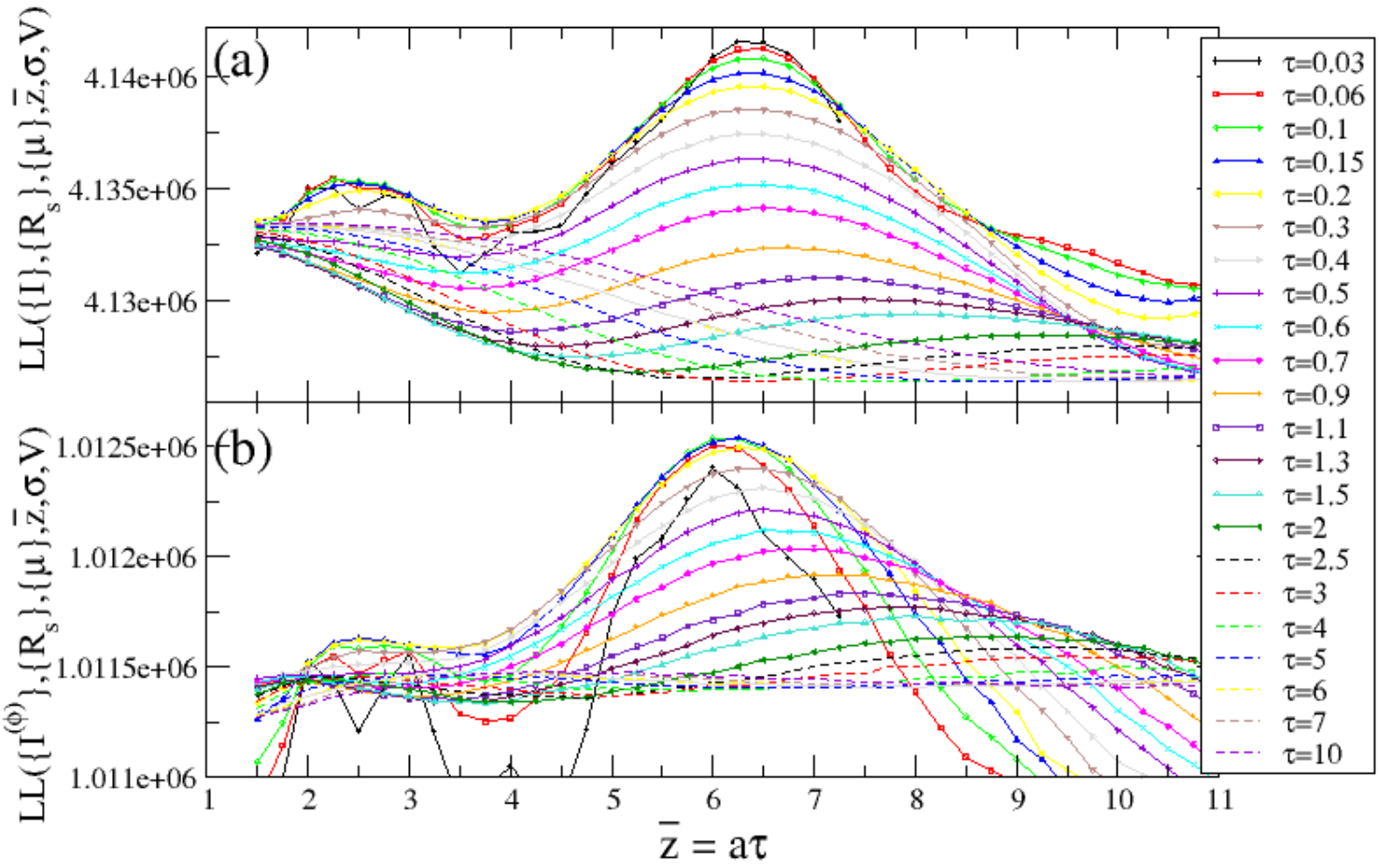

Figure 3

(Panel a) The log-likelihood... (please see manuscript .pdf for full caption)

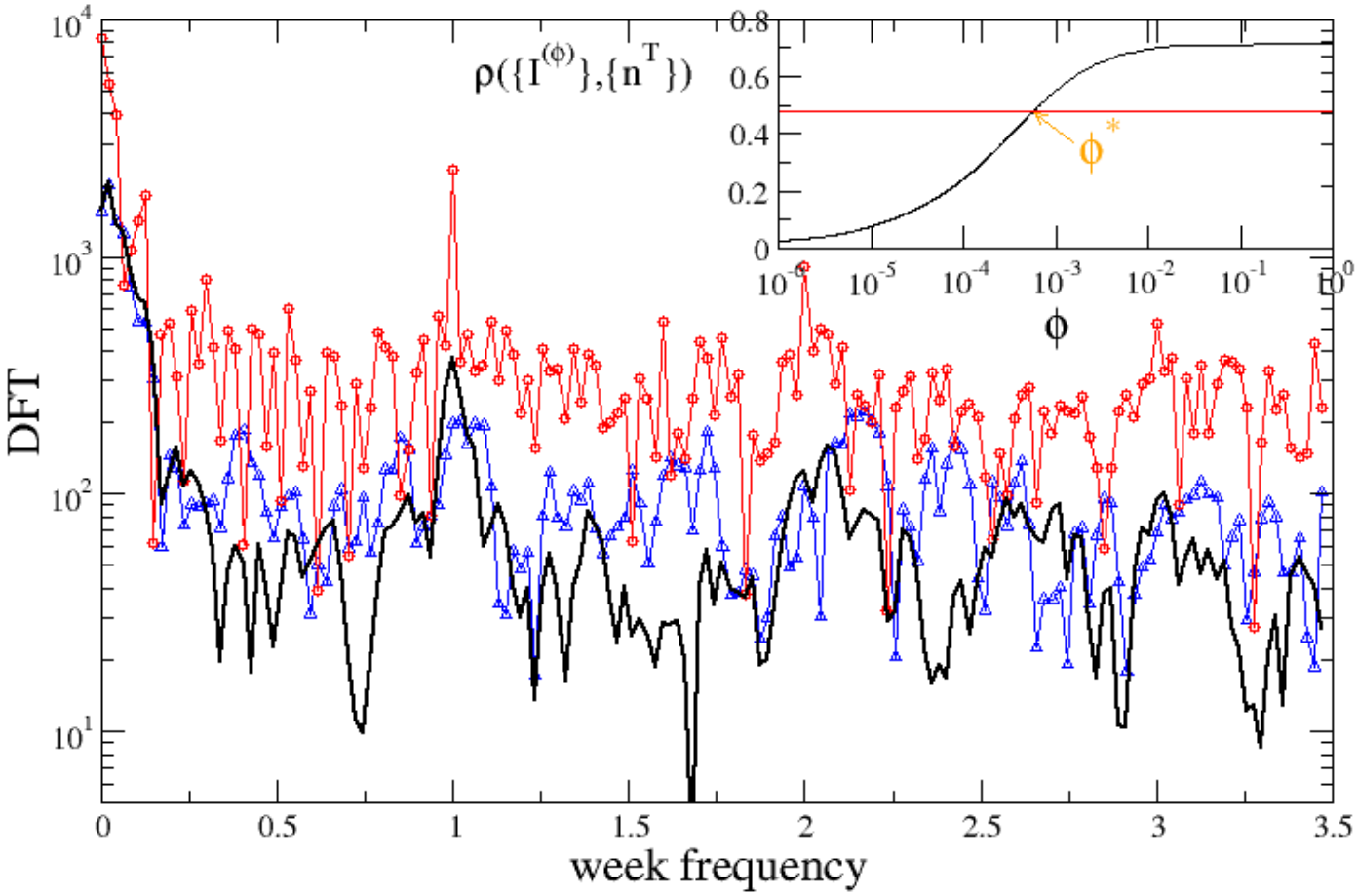

Figure 4

The DFT of the series... (please see manuscript .pdf for full caption)

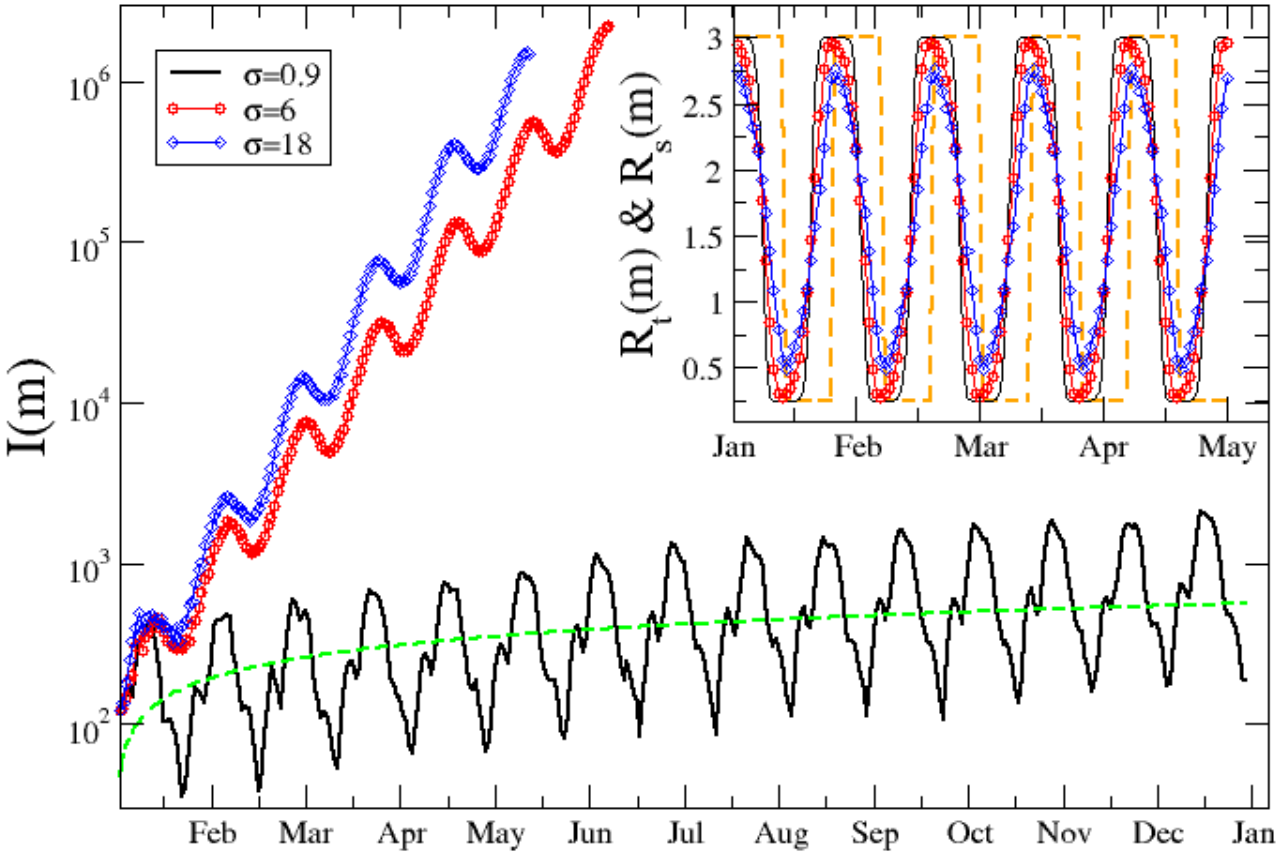

Figure 5

The daily incidence I(m) expected for a periodic fast switching protocol which alternates period without constrains to rigid lockdown periods, each one lasting 12 days. We assume I(0) = 1000 at the starting time. Different colors and symbols are used for different values of σ in the probability distribution of generation time w(z), which is gamma distributed. The green dashed line is a power law fit m1/2 for the temporal evolution of the average value of I(m) when σ = 0.9.(Inset) We plot with dashed orange line the evolution of Rt(m) implemented in the numerical model. The other curves represent the evolution of Rs(m) for different values of τ, with the same color code and symbols of the main panel.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- rtgtSI.pdf