THE EUROPEAN
PHYSICAL JOURNAL PLUS

Regular Article

A simplified estimate of the effective reproduction number R_t using its relation with the doubling time and application to Italian COVID-19 data

Gianluca Bonifazi[1,2]**, Luca Lista**[3,4,a] **, Dario Menasce**[5]**, Mauro Mezzetto**[6]**,**
Daniele Pedrini[5]**, Roberto Spighi**[2]**, Antonio Zoccoli**[2,7]

[1] Università Politecnica delle Marche, Ancona, Italy
[2] INFN Sezione di Bologna, Bologna, Italy
[3] Università degli Studi di Napoli Federico II, Naples, Italy
[4] INFN Sezione di Napoli, Naples, Italy
[5] INFN Sezione di Milano Bicocca, Milano, Italy
[6] INFN Sezione di Padova, Padova, Italy
[7] Alma Mater Studiorum Università di Bologna, Bologna, Italy

Abstract A simplified method to compute R_t, the effective reproduction number, is presented. The method relates the value of R_t to the estimation of the doubling time performed with a local exponential fit. The condition $R_t = 1$ corresponds to a growth rate equal to zero or equivalently an infinite doubling time. Different assumptions on the probability distribution of the generation time are considered. A simple analytical solution is presented in case the generation time follows a gamma distribution.

1 Introduction

The effective reproduction number R_t is one of the main parameters that controls the evolution of an infection. It recently gained importance during the COVID-19 pandemic outbreak and is used as one of the indicators to determine restrictive measures such as regional or national lock-downs.

Different algorithms for its computation are available [1–5], some of which are very CPU intensive.

Implementations are also available as software packages [6] for a number of algorithms, and results are presented on websites [7–9] with regular updates.

CPU-effective algorithms offer the advantage that estimates can be derived in real time as soon as new data are published. Often, results of simplified algorithms don't differ too much from the results of more accurate methods, in particular due to the limited quality of input data.

[a] e-mail: luca.lista@infn.it (corresponding author)

Springer

The following proposes a simplified approach to the estimate of R_t based on the determination of the doubling time, or equivalently the growth rate. The growth rate can be simply obtained as the slope parameter from a linear interpolation, in a certain interval of time, of the logarithms of the daily number of infected persons.

2 The effective reproduction number, R_t

We assume I_t is the number of infected persons at the time t, measured as number of days from a conventional beginning of the epidemic, defined as $t = 0$.

Each contagious person can infect other people during his infection period. We assume that a person that got infected at a day d will infect, on average, a certain number of other persons that become infectious at the day $t > d$ with a discrete probability distribution w_s, with $s = t - d$. The newly infected people, on turn, may infect more people with the same mechanism. $s = t - d$ is defined as the generation time in literature and corresponds to the time interval between infector-infected pair.

The probability distribution w_s is normalized to unity:

$$\sum_{s=1}^{\infty} w_s = 1 \, . \tag{1}$$

In practice, after a sufficiently large amount of time, *i.e.*: for a sufficiently large value of s, w_s becomes negligible. An estimate of w_s from Italian infection data, unfortunately from a limited number of cases, is published in [10] where w_s is approximated with a gamma distribution.

At a time t, the expected number of infected persons, $\mathbb{E}[I_t]$ can be determined from I_d, $d = 0, \ldots, t - 1$, according to [3], as:

$$\mathbb{E}[I_t] = R_t \sum_{d=0}^{t-1} I_d w_{t-d} \, , \tag{2}$$

or, equivalently, defining $s = t - d$, as:

$$\mathbb{E}[I_t] = R_t \sum_{s=1}^{t} I_{t-s} w_s \, . \tag{3}$$

The simplest assumption on w_s is a constant generation time g, which is equivalent $w_s = \delta_{gs}$ where δ_{gs} is a Kronecker delta, *i.e.*: $w_g = 1$ and $w_s = 0$ for $s \neq g$. In this case, Eq. 3 becomes:

$$\mathbb{E}[I_t] = R_t I_{t-g} \, . \tag{4}$$

For COVID-19, the average generation time, defined as the mean value of a gamma distribution fitted to the Italian data, is $g = 6.7 \pm 1.9$ days [10]. The Robert Koch Institute (RKI) takes instead for Germany the value $g = 4$ that gives a very simple estimate $\hat{R}_t$ of R_t [4] [1]:

$$\hat{R}_t = \frac{I_t}{I_{t-g}} \, , \tag{5}$$

---

[1] We tested this algorithm over the R_t values computed by RKI for the German cases as reported in [11], and we found an excellent agreement with the data published by the RKI.
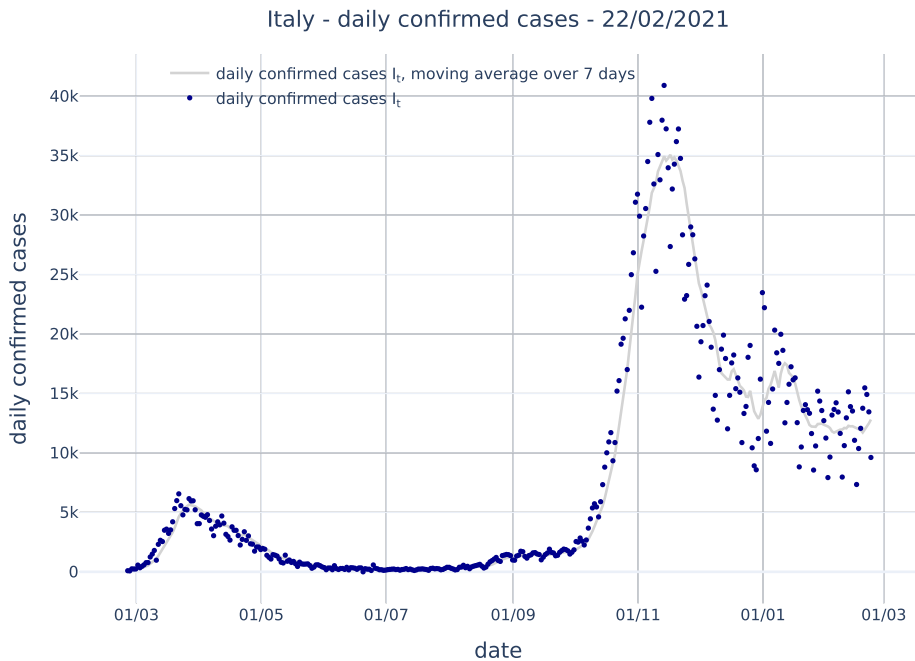
Fig. 1 Number of daily confirmed cases, I_t, for Italy according to public COVID-19 Italian data from the Italian Dipartimento di Protezione Civile, dark blue dots. The moving average over 7 days is also shown as light gray line

or the smoother ratio of the moving averages over g days:

$$\hat{R}_t = \frac{\sum_{d=t-g+1}^{t} I_d}{\sum_{d=t-g+1}^{t} I_{d-g}} \, . \tag{6}$$

Usually, the moving average over few days does not sufficiently smooth the distribution of the number of daily infected cases I_t. In particular, the lower number of swab tests taken during the weekend causes a "ripple" structure that requires a further smoothing to be applied to the input data before evaluating Eq. 6.

Figure 1 shows the number of daily confirmed cases, I_t for Italy according to public COVID-19 Italian data from the Italian Dipartimento di Protezione Civile. The large dispersion of data is clearly visible, in particular around the more stable moving average over 7 days, also shown in the figure.

3 Relation between R_t and doubling time

Another indicator of the growth of the epidemic is the doubling time τ_2 defined as the time required to double the number of infected persons, assuming an exponential growth.

Given n consecutive counts of infected people, $I_{t-n+1}, \ldots, I_t$, the following function model can interpolate the n counts:

$$I_t = A \, e^{\lambda t} \, , \tag{7}$$

or, equivalently:

$$I_t = A\, 2^{t/\tau_2} . \tag{8}$$

The growth rate λ is related to the doubling time τ_2 by:

$$\tau_2 = \frac{\log 2}{\lambda} . \tag{9}$$

Estimates of A and λ, or equivalently τ_2, can be determined with a numerical fit procedure. In particular, the exponential fit can be conveniently implemented as a linear regression on $\log I_t$.

Assuming $R_t = R$ constant during the considered time interval, the evolution model in Eq. 4 represents an exponential growth. In a time period formed by a number of days n which is an integer multiple of g: $n = Ng$, we have:

$$\mathbb{E}[I_t] = R^N I_h , \tag{10}$$

where $h = t - n + 1$, or:

$$\mathbb{E}[I_t] = R^{n/g} I_h . \tag{11}$$

Changing the base from R to e gives:

$$\mathbb{E}[I_t] = e^{(n \log R)/g} I_h . \tag{12}$$

Comparing with Eq. 7, considering that $t = h + n + 1$, and $A\, e^{\lambda t} = A\, e^{\lambda n} e^{\lambda(h+1)}$, we have:

$$\lambda = \frac{\log R}{g} , \quad A = I_h e^{-\lambda(h+1)} , \tag{13}$$

hence the estimate $\hat{R}$ of R is:

$$\hat{R} = e^{g\hat{\lambda}} = e^{(g \log 2)/\hat{\tau}_2} , \tag{14}$$

where $\hat{\lambda}$ and $\hat{\tau}_2$ are the estimates of λ and τ_2, respectively.

4 Simplified algorithm

We studied the progression of the COVID-19 pandemic in Italy, considering the data published on daily basis by Italian Dipartimento di Protezione Civile [13]. For each day t, we perform an exponential fit to the n last days' counts, $I_{t-n+1}, \ldots, I_t$. We determine an estimate $\hat{\tau}_2$ of the doubling time τ_2, or an equivalent estimate $\hat{\lambda}$ of the growth rate λ, from a fit of the model in Eq. 7 or Eq. 8. Then, assuming a reasonable estimate g of the average generation time, we estimate R_t according to Eq. 14 as:

$$\hat{R}_t = e^{g\hat{\lambda}} = e^{(g \log 2)/\hat{\tau}_2} = 2^{g/\hat{\tau}_2} . \tag{15}$$

There are some advantages of Eq. 15 compared to the simplified model from Eq. 6:

- Equation 15 can also be applied in case g, the average generation time, is not an integer, while Eq. 6 must approximate g to the nearest integer.
- The exponential fit better follows an exponential growth in the considered time interval, as it is the case when R_t is a constant, with respect to a moving average.

At the cost of a modest increase in the computing time, yet maintaining very good speed, we consider the method proposed here to be more flexible and reliable compared to the method adopted in [4]. Moreover, the data smoothing can be tuned by including a sufficient number of points in the fit. In this way, no preliminary smoothing of the data is needed before the application of the algorithm.

In the following sections, we will introduce extensions of Eq. 15 that allow a more precise determination of $\tilde{R}_t$ than with the simplified assumption that $w_s = \delta_{gs}$, i.e.: s is constant and equal to g.

5 Uncertainty estimate

Given Eq. 15, the uncertainty on $\hat{R}_t$ is determined by the uncertainties on $\hat{\lambda}$ (or τ_2) and the uncertainty on g. Namely, if $\sigma_{\hat{\lambda}}$ and σ_g are the uncertainties on $\hat{\lambda}$ and g, respectively, within a Gaussian error approximation, the variance of R_t is given by:

$$\text{Var}[\hat{R}_t] = \left(\frac{\partial \hat{R}_t}{\partial \hat{\lambda}}\sigma_{\hat{\lambda}}\right)^2 + \left(\frac{\partial \hat{R}_t}{\partial g}\sigma_g\right)^2 = (ge^{\hat{\lambda}g}\sigma_{\hat{\lambda}})^2 + (\hat{\lambda}e^{\hat{\lambda}g}\sigma_g) = (e^{\hat{\lambda}g})^2(g^2\sigma_{\hat{\lambda}}^2 + \hat{\lambda}^2\sigma_g^2). \tag{16}$$

The error on $\hat{R}_t$ is:

$$\sigma_{\hat{R}_t} = \sqrt{\text{Var}[\hat{R}_t]} = \hat{R}_t\sqrt{g^2\sigma_{\hat{\lambda}}^2 + \hat{\lambda}^2\sigma_g^2}. \tag{17}$$

The uncertainty on $\hat{\lambda}$ derives from the exponential fit procedure, while the uncertainty on g depends on how well the probability distribution of the generation time w_s is known. From [10], the estimate of w_s and its average g for COVID-19 in Italy is known from a limited number of cases.

In particular, when $\hat{\lambda} = 0$ (infinite doubling time), which corresponds to $\hat{R}_t = 1$, σ_g doesn't contribute to the $\hat{R}_t$ uncertainty. This means that an imperfect assumption on g does not affect the condition $\hat{R}_t = 1$ which is important to determine the turning point of infection, from growing to receding, or vice versa.

The uncertainty computed in Eq. 17 does not take into account the systematic uncertainty due to the assumed approximation that the generation time s is constant, and equal to g. Moreover, the assumption of Gaussian uncertainties may not hold for an asymmetric distribution.

6 Effect of finite width in the w_s distribution

The deviation of w_s from the hypothesis of a constant generation time $s = g$ may be approximately estimated in the continuum approximation. Equation 3 for a continuous time variable t may be rewritten as:

$$i(t) = \rho(t)\int_0^t i(t-s)\,w(s)\,\mathrm{d}s\,, \tag{18}$$

where $\rho(t)$ and $i(t)$ are the continuum equivalent of R and I_t, respectively.

The normalization condition is:

$$\int_0^\infty w(s)\,ds = 1\,. \tag{19}$$

If s is a constant equal to g, we have $w(s) = \delta(s - g)$, where δ is a Dirac's delta function. Hence:

$$i(t) = \rho(t)\,i(t - g)\,. \tag{20}$$

Assuming an exponential growth $i(t) = A\,e^{\lambda t}$, one has:

$$A\,e^{\lambda t} = \rho(t)\,A\,e^{\lambda(t-g)} = \rho(t)\,A\,e^{\lambda t}e^{-\lambda g}\,, \tag{21}$$

which gives the continuous version of Eq. 15, where $\rho(t) = \rho$ is a constant:

$$\rho = e^{\lambda g}\,. \tag{22}$$

Assuming, instead, that $w(s)$ deviates from the Dirac's delta assumption and has average value g and standard deviation σ, we may write Eq. 18 applying a series expansion of $i(t - s)$ around $s = g$:

$$i(t) = \rho(t) \int_0^t \left[i(t - g)\,w(s) - i'(t - g)\,(s - g)\,w(s) + \frac{1}{2}i''(t - g)\,(s - g)^2\,w(s) + \cdots \right] ds\,. \tag{23}$$

We assume that $w(s) \simeq 0$ for $s > t$, so that the integration can be extended from 0 to ∞ instead of 0 to t.

After the integration, in the first term the normalization condition of $w(d)$ can be applied. The second term vanishes, and in the third term the definition of standard deviation σ of $w(s)$ can be applied. Equation 23 becomes:

$$i(t) \simeq \rho(t) \left[i(t - g) + \frac{\sigma^2}{2}i''(t - g) \right]\,. \tag{24}$$

If we assume again $i(t) = A\,e^{\lambda t}$, hence $i''(t) = A\,\lambda^2\,e^{\lambda t}$, Eq. 24, becomes:

$$A\,e^{\lambda t} = \rho(t) \left[A\,e^{\lambda t}e^{-\lambda g} + A\,\frac{\sigma^2}{2}\lambda^2 e^{\lambda t}e^{-\lambda g} \right]\,. \tag{25}$$

The term $A\,e^{\lambda t}$ simplifies. If $\lambda^2\sigma^2 \ll 1$, we may write, approximately:

$$1 = \rho(t)\,e^{-\lambda g}\left(1 + \frac{\sigma^2}{2}\lambda^2\right) \simeq \rho\,e^{-\lambda g}e^{\lambda^2\sigma^2/2}\,, \tag{26}$$

hence:

$$\rho = e^{\lambda g - \lambda^2\sigma^2/2}\,. \tag{27}$$

Equation 27 has already been reported in [12]. This result implies the width of the distribution w_s has the effect to replace g in Eq. 22 with an "effective" generation time g^{eff} that is somewhat smaller than the true average value and depends on λ according to:

$$g^{\text{eff}} = g - \lambda\frac{\sigma^2}{2}\,. \tag{28}$$

In order to take into account more details of the distribution, more terms may be added to Eq. 23. Those would add a dependency of ρ on the higher moments: skewness, kurtosis and possibly more, if required by the desired accuracy. Those cases are not considered in the present work.

7 "Exact" solution

If we assume, as in the previous section, that $i(t)$ is an exponential function, or at least that it can be approximated to an exponential function within a time interval that is at least as wide as the time range in which $w(s)$ is not negligible, $\rho(t)$ can be computed "exactly," and is constant within that interval.

If we assume $i(t) = A\,e^{\lambda t}$, Eq. 18 becomes:

$$A\,e^{\lambda t} = \rho(t) \int_0^t A\,e^{\lambda(t-s)} w(s)\,\mathrm{d}s = A\,e^{\lambda t} \rho(t) \int_0^t e^{-\lambda s} w(s)\,\mathrm{d}s . \tag{29}$$

Simplifying the term $A\,e^{\lambda t}$, as in the previous cases, $\rho(t)$ can be computed as:

$$\rho(t) = \left[\int_0^t e^{-\lambda s} w(s)\,\mathrm{d}s \right]^{-1} . \tag{30}$$

If $w(s)$ is negligible for values of $s > t$, we can extend the integration from 0 to ∞, and $\rho(t) = \rho$ does not depend on t:

$$\rho = \left[\int_0^\infty e^{-\lambda s} w(s)\,\mathrm{d}s \right]^{-1} . \tag{31}$$

This result is also reported in [12].

Note that if $\lambda = 0$, Eq. 31 becomes:

$$\rho = \left[\int_0^\infty w(s)\mathrm{d}s \right]^{-1} . \tag{32}$$

The normalization of $w(s)$ implies $\rho = 1$, regardless of the details of the probability distribution $w(s)$.

8 The case of a gamma distribution

In [10], $w(s)$ is approximated to a gamma distribution:

$$w(s) = \frac{s^{\kappa-1} e^{-s/\theta}}{\theta^\kappa\,\Gamma(\kappa)} , \tag{33}$$

where κ and θ, the shape and scale parameters, are determined with a fit to the Italian data. Equation 31 becomes:

$$\rho = \theta^\kappa\,\Gamma(\kappa) \left[\int_0^\infty s^{\kappa-1} e^{-s(\lambda+1/\theta)}\,\mathrm{d}s \right]^{-1} , \tag{34}$$

where the integration can be performed analytically:

$$\rho = \theta^\kappa\,\Gamma(\kappa) \left[-\frac{\Gamma(\kappa,(\lambda+1/\theta)s)}{(\lambda+1/\theta)^\kappa} \Bigg|_{s=0}^{s=\infty} \right]^{-1} . \tag{35}$$

With some simplification of the Γ functions, the result is:

$$\rho = (1 + \lambda\theta)^{\kappa} . \tag{36}$$

The above equation is valid for $-1/\theta < \lambda < \infty$. Again, $\lambda = 0$ corresponds to $\rho = 1$ for any values of κ and θ, as demonstrated in general in the previous section.

9 R_t and τ_2 as indicators of the epidemic evolution

R_t is often used as indicator of the epidemic evolution. As we have seen, there is a very close relation between the Effective Reproduction Number and doubling time. The estimate of the doubling time τ_2 can be determine directly from the number of infected people, while R_t also requires an estimate of the average generation time g, which propagates an extra uncertainty with respect to the estimate of τ_2.

The main feature of R_t is the passage through the threshold value of one: $R_t > 1$ indicates a growing phase, while $R_t < 1$ indicates a receding phase of the epidemic. Those conditions are equivalent to $\tau_2 > 0$ and $\tau_2 < 0$, respectively, as evident form Eq. 15. In the case $\hat{\lambda} = 0$, $\hat{R}_t$ is not affected by the uncertainty on the estimate of g.

For this reason, we consider τ_2, or equivalently λ, a better indicator of the situation of the epidemic compared to R_t, which may be of interest for other epidemiology purposes.

10 Results

Figure 2 shows R_t, evaluated with the presented algorithm assuming a constant generation time, using the public Italian COVID-19 data released by the Italian Dipartimento di Protezione Civile [13]. Different values of the average generation time g have been assumed, from 3 to 7 days.

The magnitude of the dependence of R_t on g gives also a clue about the uncertainty on R_t due to imperfect knowledge of g, which mainly affects the regions where R_t is significantly different from 1.

Figure 3 shows instead the evaluation performed with the three models discussed above:

1. Eq. 15, assuming a constant generation time of $g = 6.7$ days;
2. Eq. 28, assuming a mean value of 6.7 days and a standard deviation of 4.88 days;
3. Eq. 36, assuming a gamma distribution having parameters $\kappa = 1.87$ and $\theta = 3.57$ days, respectively, as determined in [10].

Note that the mean of the gamma distribution is equal to the product $\kappa\theta$.

All three methods give similar values for R_t close to 1, but exhibit some discrepancy at more extreme values. Compared to the "exact" solution that assumes a gamma distribution (Eq. 36), assuming a fixed generation time (Eq. 15) gives a result that is about 9% larger at the highest value and about 4% larger at the lowest value. Including the contribution of the standard deviation term (Eq. 28) gives a reduction of about 12% at the larges value and 3% at the lowest value. Using (Eq. 15) with a lower "effective" g may improve the agreement with the "exact" solution at higher values at the cost of a poorer agreement at lower values. This is effectively done in the implementation of the RKI algorithm.

Figure 4 shows the application of different algorithms to the official Italian COVID-19 data published by the Italian Dipartimento di Protezione Civile [13]. The algorithm presented in this paper is noted as CovidStat and assumes a gamma distribution with the parameters
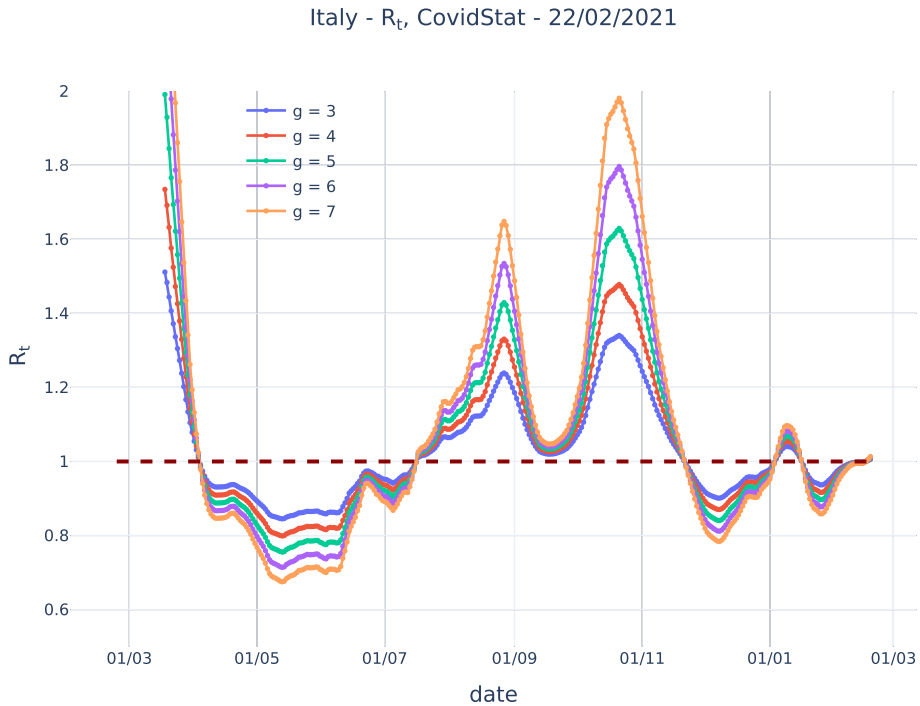
Fig. 2 R_t evaluated on the public COVID-19 Italian data released by the Italian Dipartimento di Protezione Civile with the presented algorithm assuming different constant values of the generation time g from 3 to 7 days

reported above. It is compared with algorithms by Wallinga and Teunis [1], Bettencourt and Ribeiro [2], Cori et al. [3], and RKI [4]. Algorithms by Wallinga and Teunis and Cori et al. use the details of the probability distribution w_s and are here implemented assuming the same w_s as our algorithm. Bettencourt and Ribeiro uses a fixed time, that we have set to 7 days.

The method proposed here has been implemented with an exponential fit to the last 14 days. The RKI algorithm has been applied with generation time $g = 5$, since the original implementation with $g = 4$ showed significant discrepancy with respect to the other algorithms, consistently with what can be noted in Fig 3. A smoothing of the infection data with a Savitzky-Golay filter [14] using a time window of 15 days and a third-order polynomial was also applied to the infection data before applying the RKI algorithm.

The comparison of the proposed method with other algorithms shows a good agreement, considering the possible source of uncertainties and the intrinsic "ripple" structure of the data that may depend on the applied smoothing. In particular, agreement of our method is very good with the Wallinga-Teunis and with the Cori et al. algorithms. The agreement with the Bettencourt-Ribeiro is also good, considering that it includes a "ripple" structure due to the data fluctuations. The agreement with the RKI method is also reasonable after the assumed constant generation time is "tuned," with a residual disagreement for the cases where $R_t < 1$. This feature is consistent with what can be observed comparing the "exact" solution computed for the gamma distribution to the one computed assuming a fixed generation time "tuned" to the more convenient value $g = 5.5$, as shown in Fig 5.
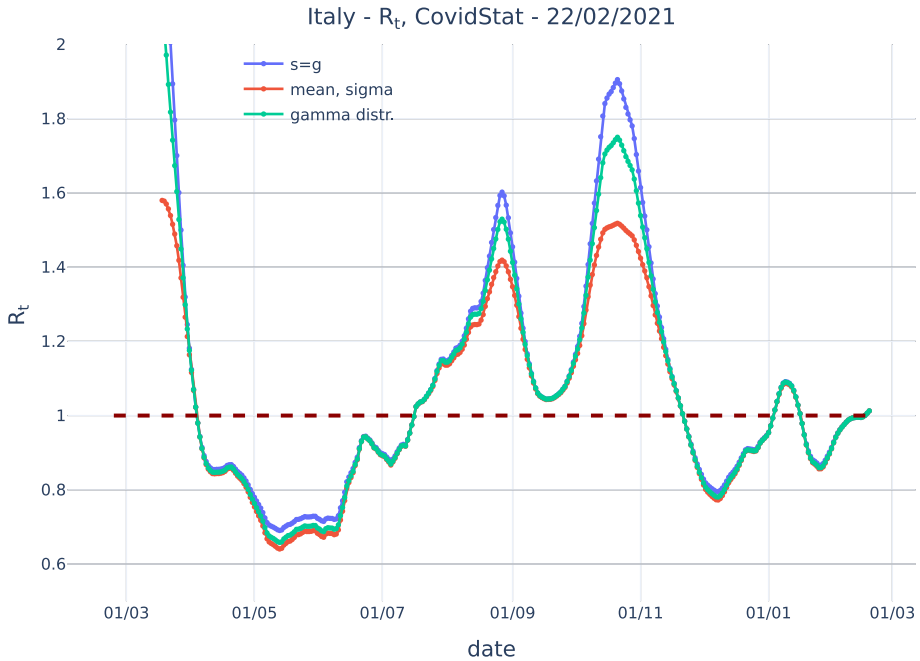
Fig. 3 R_t evaluated on the public COVID-19 Italian data released by the Italian Dipartimento di Protezione Civile with assuming a constant generation time, assuming a mean value and a standard deviation contribution, and assuming a gamma distribution. The assumed parameters are taken from [10]

Figure 6 shows the estimated growth rate λ and the corresponding R_t for Italy data. Estimates are done with an exponential fit over the last 14 days. For R_t the contribution to uncertainty due to the propagation of the statistical uncertainty on λ is, in most of the range, much smaller than the total uncertainty that also takes in to account the uncertainty on the parameters that model $w(s)$, according to the estimate from [10]. This contribution to the total uncertainty is particularly large as the values of R_t depart from one. For $R_t = 1$, as noted before, the uncertainty contribution form the parameters that model $w(s)$ is null. The magnitude of the total uncertainty is comparable with what is obtained from the algorithm by Cori et al. that tales into account the uncertainty on $w(s)$.

11 Performances

We compared the CPU time required to run the five algorithms considered in this paper. The benchmarks ran on a dedicated cluster with 32 cores/64 threads on two AMD EPYC 7301 processors and 64GB RAM. The algorithm ran on a single thread avoiding any multithread implementation. The results are reported in Table 1.

The algorithm proposed in this paper outperforms all other algorithms, in particular when the number of cases is large, as for Itay and Lombardia. The comparison with the RKI algorithm is not very meaningful. RKI estimates R_t as the ratio of the number of infected persons last $g = 5$ days divided by the the number of infected persons in the previous g days, which takes a very small CPU time. Nonetheless, our implementation is largely dominated by the overhead introduced by the python module `pandas` [15] compared to `numpy` [16],
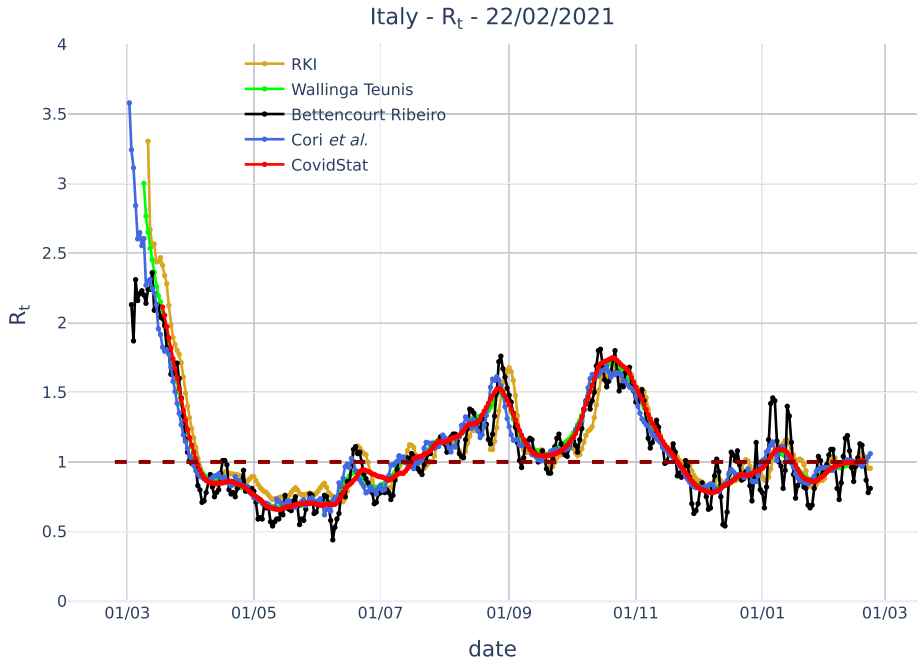
Fig. 4 Comparison of R_t computed using different algorithms with public COVID-19 Italian data from the Italian Dipartimento di Protezione Civile. The algorithm presented in this paper is noted as CovidStat and assumes a gamma distribution with known parameters. It is compared with algorithms by RKI, Wallinga and Teunis, Bettencourt and Ribeiro, and Cori et al.

which is faster, and is the one we use for the CovidStat algorithm. The choice was only dictated by convenience, and we didn't consider any porting of our implementation of the RKI algorithm to `numpy`, that would outperform the CovidStat algorithm, because the gain would be negligible anyway.

We report in the CovidStat website [9] R_t estimates for Italy, for North, Center and South separately, for the 20 Italian regions, and for the autonomous provinces of Bolzano and Trento. On the aforementioned dedicated 64-thread cluster, each geographic area running on a separate thread, the computation takes about 30 minutes for all five algorithms, including all the data management overhead.

In addition, we compute R_t for the 107 provinces and for about 30 countries. For those, we only compute the CovidStat R_t estimate in order to reduce the required computation time. This evaluation takes a negligible CPU compared with the other methods of computation of R_t.

Updates are published on our website daily and are produced automatically, with no human intervention, as soon as the data from the Dipartimento della Protezione Civile are available.

12 Conclusion

A simplified method to determine an estimate of R_t based on a local exponential fit is presented. The method can be applied assuming a fixed generation time, including the contribution of the standard deviation of the generation time distribution, or assuming a func-
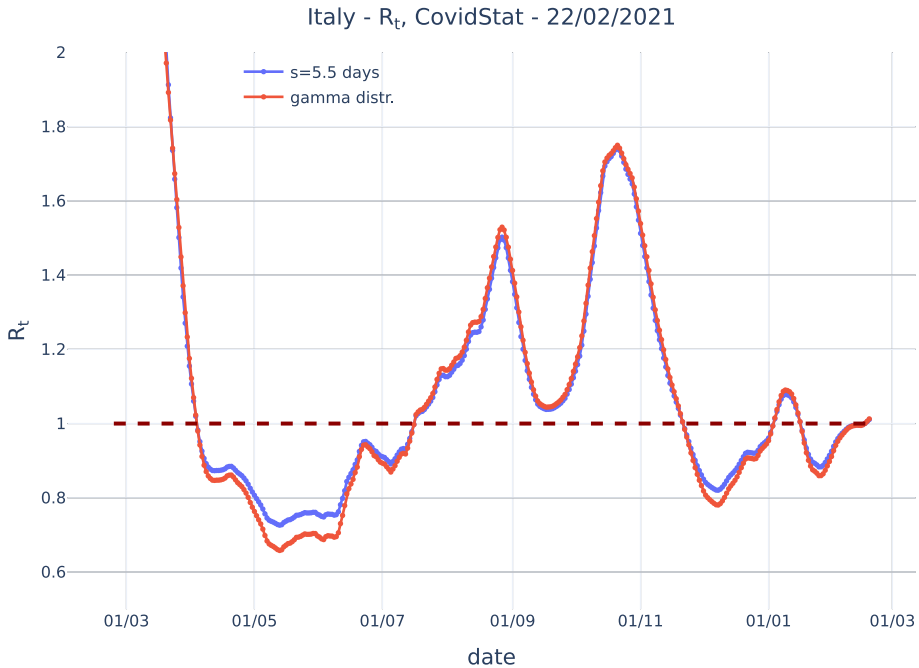
Fig. 5 Comparison of R_t computed assuming a gamma distribution and assuming a constant generation time "tuned" to $g = 5.5$ in order to reduce the disagreement for $R_t > 1$. A residual disagreement for $R_t < 1$ is visible

Table 1 CPU time in seconds required to run the five R_t algorithms for Italy and five Italian regions with decreasing number of inhabitants and the number of infected persons. Emilia-Romagna has lower number of inhabitants, but significantly more infected persons compared to Lazio and Campania. The specs of the cluster used for the benchmark are reported in the text

Geographic area	Inhabitants (mln., 2019)	CovidStat	Wallinga Teunis	Bettencourt Ribeiro	Cori et al.	RKI
Italy	60.36	0.138	510.1	80.3	11.5	0.587
Lombardia	10.06	0.120	80.8	78.3	11.5	0.585
Lazio	5.88	0.105	30.0	76.1	11.2	0.586
Campania	5.80	0.103	31.9	33.9	11.2	0.584
Emilia-Romagna	4.46	0.101	33.5	75.2	11.4	0.585
Basilicata	0.56	0.098	20.1	28.9	11.3	0.584

tional form for the probability distribution of the generation time. If a gamma distribution is assumed, a simple analytic solution is reported. The method offers some advantages compared to the simplified method adopted by the Robert Koch Institut, yet preserving good computing performances that makes it suitable for a real-time evaluation.

Results of the method applied to the public Italian COVID-19 data have been presented. The proposed method shows a good agreement with other, more complex, algorithms available in literature and implemented in public software packages.

We note a close relation between R_t and the doubling time of the number of infections τ_2, or equivalently the growth rate λ. In particular, the condition $R_t > 1$ is equivalent to
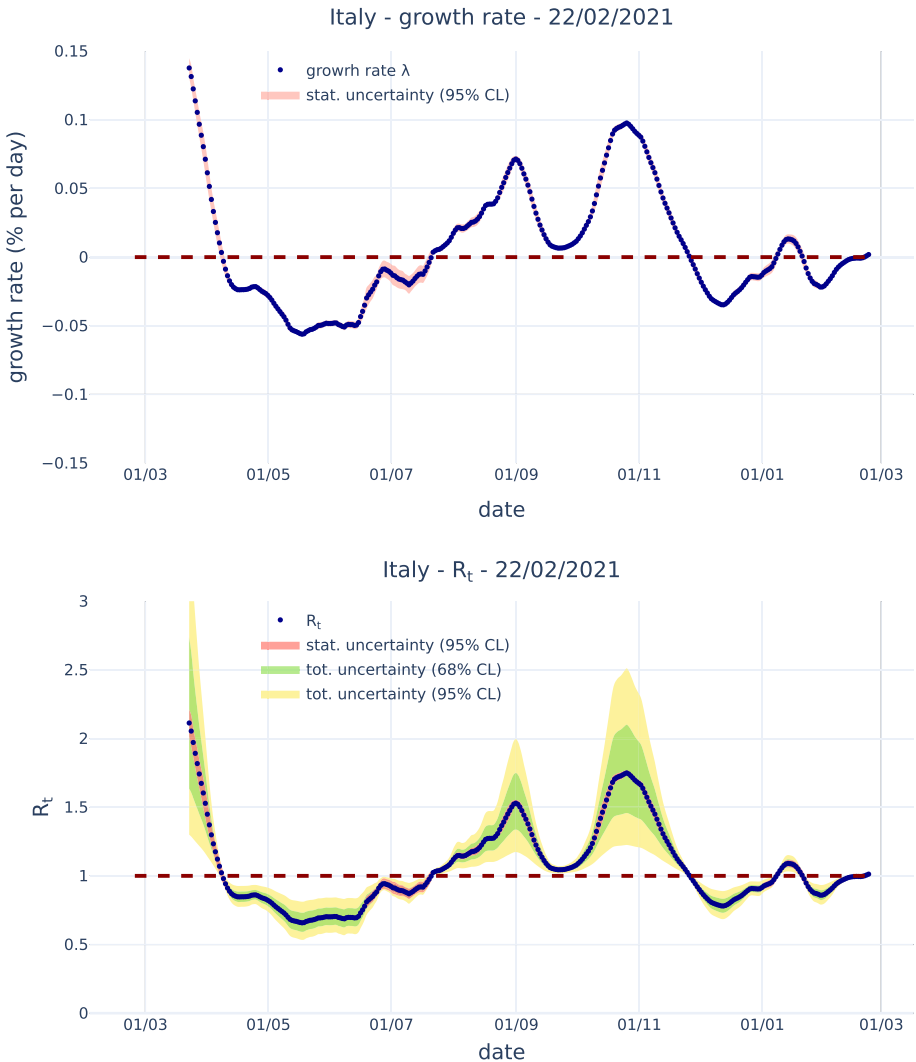
Fig. 6 Growth rate λ (top) and R_t (bottom). For the growth rate λ, the statistical uncertainty band at 95% Confidence Level is shown. For R_t the contribution to uncertainty due to the propagation of the statistical uncertainty on λ at 95% confidence level is shown together with the total uncertainty at the 68% and 95% confidence level, that also takes in to account the uncertainty on the parameters that model $w(s)$. All data refer to Italy according to public COVID-19 Italian data from the Dipartimento di Protezione Civile

$\tau_2 > 0$ or $\lambda > 0$. Since the determination of R_t is affected by additional uncertainty sources compared to τ_2, we consider τ_2 or λ to be a more sound and simpler indicator of the condition of growing or receding epidemic compared to R_t, while R_t may have more importance in other contexts of epidemiological interest.

We publish in real time daily estimates of R_t as computed by our algorithm and by all the other algorithms quoted in this article for the cases in Italy and all the Italian regions under [9]. Daily values for the major world countries are also reported.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: Available at `covid19.infn.it`.]

References

1. J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am. J. Epidemiol. **160**(6), 509–516 (2004). https://doi.org/10.1093/aje/kwh255
2. L.M.A. Bettencourt, R.M. Ribeiro, Real time bayesian estimation of the epidemic potential of emerging infectious diseases. PLoS ONE **3**(5), e2185 (2008). https://doi.org/10.1371/journal.pone.0002185
3. A. Cori, N.M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics. Am. J. Epidemiol. **178**(9), 1505–1512 (2013). https://doi.org/10.1093/aje/kwt133
4. Robert Koch Institut, *Erläuterung der Schätzung der zeitlich variierenden Reproduktionszahl* (2020), https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/R-Wert-Erlaeuterung.pdf
5. K. Systrom, *The Metric We Need to Manage COVID-19. R_t: the effective reproduction number*, (2020), http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/
6. EpiEstim: *Estimate Time Varying Reproduction Numbers from Epidemic Curves*, https://cran.r-project.org/web/packages/EpiEstim/index.html
7. K. Systrom, R_t, *Effective Reproduction Number*, https://rt.live/
8. G. Bonifazi, *R_t COVID-19 Italia, Numero effettivo di riproduzione del virus*, https://rt-italy.live/
9. CovidStat INFN, https://covid19.infn.it/
10. D. Cereda et al., *The early phase of the COVID-19 outbreak in Lombardy, Italy*, arXiv:2003.09320 (2020)
11. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/Nowcasting_Zahlen.xlsx?__blob=publicationFile
12. J. Wallinga, M. Lipsitch, How generation intervals shape the relationship between growth rates and reproductive numbers. Proc Biol Sci. **274**(1609), 599–604 (2007). https://doi.org/10.1098/rspb.2006.3754
13. Dipartimento della Protezione Civile, *Dati COVID-19 Italia*, https://github.com/pcm-dpc/COVID-19
14. A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 1964 **36**(8), 1627–1639 (1964). https://doi.org/10.1021/ac60214a047
15. Pandas, https://pandas.pydata.org/
16. NumPy, https://numpy.org/