Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique

Luis Alvarez[a,1], **Miguel Colom**[b], **Jean-David Morel**[c], **and Jean-Michel Morel**[b]

[a]CTIM. Departamento de Informática y Sistemas, Universidad de Las Palmas de Gran Canaria. Spain; [b]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-94235, Cachan, France.; [c], Laboratoire de Physiologie Intégrative et Systémique Ecole Polytechnique Fédérale de Lausanne, AI 1144 Station 15 CH-1015 Lausanne Switzerland

The COVID-19 pandemic has undergone frequent and rapid changes in its local and global infection rates, driven by governmental measures, or the emergence of new viral variants. The reproduction number R_t indicates the average number of cases generated by an infected person at time t and is a key indicator of the spread of an epidemic. A timely estimation of R_t is a crucial tool to enable governmental organizations to adapt quickly to these changes and assess the consequences of their policies. The EpiEstim method is the most widely accepted method for estimating R_t. But it estimates R_t with a significant temporal delay. Here, we propose a new method, *EpiInvert*, that shows good agreement with EpiEstim, but that provides estimates of R_t several days in advance. We show that R_t can be estimated by inverting the renewal equation linking R_t with the observed incidence curve of new cases, i_t. Our signal processing approach to this problem yields both R_t and a restored i_t corrected for the "weekend effect" by applying a deconvolution + denoising procedure. The implementations of the EpiInvert and EpiEstim methods are fully open-source and can be run in real-time on every country in the world, and every US state through a web interface at www.ipol.im/epiinvert.

COVID-19 | Renewal equation | Reproduction number | Integral equations

The reproduction number R_t is a key epidemiological parameter evaluating transmission potential of a disease over time. It is defined as the average number of new infections caused by a single infected individual at time t in a partially susceptible population (1). R_t can be computed from the daily observation of the incidence curve i_t, but requires empirical knowledge of the probability distribution Φ_s of the delay between two infections (2, 3).

There are two different models for the incidence curve and its corresponding infection delay Φ. In a theoretical model, i_t would represent the real daily number of new infections, and Φ_s is sometimes called *generation time* (4, 5) and represents the probability distribution of the time between infection of a primary case and infections in secondary cases. In practice, neither parameter is easily observable because the infected are rarely detected before the appearance of symptoms and tests will be negative until the virus has multiplied over several days. What is routinely recorded by health organizations is the number of *new detected, incident cases*. When dealing with this real incidence curve, Φ_s is called *serial interval* (4, 5). The serial interval is defined as the delay between the onset of symptoms in a primary case and the onset of symptoms in secondary cases (5).

R_t is linked to i_t and Φ through the *renewal equation*, first formulated for birth-death processes in a 1907 note of Alfred Lotka (6). We adopt the Nishiura et al. formulation (7, 8),

$$i_t = \sum_{s=f_0}^{f} R_{t-s} i_{t-s} \Phi_s \quad \text{for} \quad t = 0, .., t_c, \qquad [1]$$

where t_c represents the last time at which i_t was available, f_0 and f are the maximal and minimal observed times between a primary and a secondary case. The underlying epidemiological assumption of this model is that the time-varying factor R_t causes a constant proportional change in an individual's infectiousness, over the course of their entire infectious period, based on the day on which they were infected. In this case we refer to R_t as the case reproductive number. According to Cori et al. (5), "It is the average number of secondary cases that a case infected at time step t will eventually infect (9)."

It is important to note that secondary infections are sometimes detected before primary ones, and therefore the minimum delay f_0 is generally negative (see Fig. 2). Equation [1] does not yield an explicit expression for R_t. Yet, an easy solution can be found for the version of the renewal equation proposed in Fraser (9) (equation (9)), and Cori et al in (5),

$$i_t = R_t \sum_{s=f_0}^{f} i_{t-s} \Phi_s. \qquad [2]$$

Significance Statement

Based on a signal processing approach we propose a method to compute the reproduction number R_t, the transmission potential of an epidemic over time. R_t is estimated by minimizing a functional that enforces: (i) the ability to produce an incidence curve i_t corrected of the weekly periodic bias produced by the "weekend effect", obtained from R_t through a renewal equation ; (ii) the regularity of R_t. A good agreement is found between our R_t estimate and the one provided by the currently accepted method, EpiEstim, except our method predicts R_t several days closer to present. We provide the mathematical arguments for this shift. Both methods, applied every day on each country, can be compared at www.ipol.im/epiinvert.

L. Alvarez and J-M. Morel designed and performed research and experiments and wrote the paper. L. Alvarez implemented the method. M. Colom built the online interface and collected and processed data. J.D. Morel rewrote parts and designed the statistical analysis and presentation of the results.

The authors declare no competing interests

[1] Luis Alvarez. E-mail: lalvarez@ulpgc.es

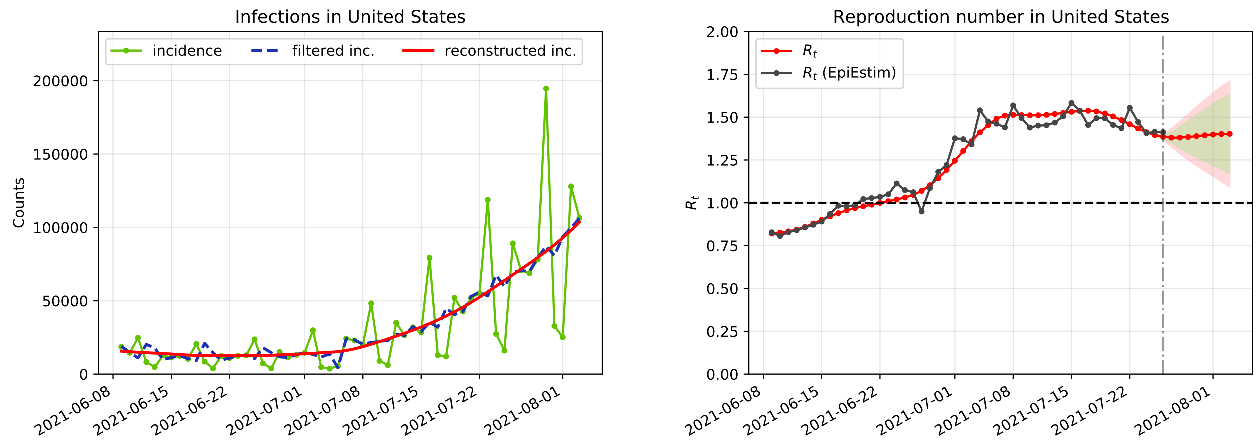

Fig. 1. Illustration of the EpiInvert method on the USA incidence curve of new cases. On the left: in green, the raw oscillating curve of incident cases up to August 3, 2021. In blue, the incidence curve after correction of the "week-end bias". In red, the incidence curve simulated from R_t after the inversion of the renewal equation. On the right: in black, R_t, the reproduction number estimated by the current EpiEstim method, adopted by most health experts (10), shifted back eight days. Estimating its value every day guides the health policy of each country. Having R_t larger than 1, as it is the case for the USA on August 3, 2021 means that the pandemic is expanding. In red, the estimation of R_t by the EpiInvert method. This estimate, obtained by compensating the week-end bias and inverting the integral equation, has a temporal shift of about eight days with respect to EpiEstim. The shadowed areas give the 90% and 95% confidence intervals for the R_t estimation.

By this equation, R_t is derived at time t from the past incidence values i_{t-s} by a simple division, provided that $f_0 \geq 0$:

$$R_t = \frac{i_t}{\sum_{s=f_0}^{f} i_{t-s}\Phi_s}. \qquad [3]$$

The underlying epidemiological assumption of this model is that the time-varying factor of R_t causes a change in the infectiousness only on the day on which transmission occurs*. In this case we refer to R_t as the instantaneous reproduction number. This R_t estimate, implemented by the EpiEstim software, is highly recommended in a very recent review (11) signed by representatives from ten different epidemiological labs from several continents.

EpiEstim is the standard method to compute a real-time estimation of the reproduction number, and of widespread use. In its stochastic formulation, the first member i_t of Equation [2] is assumed to be a Poisson variable, and the second member of this equation is interpreted as the expectation of this Poisson variable. This leads to a maximum likelihood estimation strategy to compute R_t (see (5, 12–15)). A detailed description of EpiEstim methods can be found in the supporting information.

Comparing Equations [2] and [1] shows that when applied with the same serial interval and case incidence curve, the second equation is derived from the first by assuming R_t constant on the serial interval support $[t-f, t-f_0]$. Replacing R_{t-s} by R_t in Equation [1] indeed yields Equation [2]. A more accurate interpretation of the quotient on the right of Equation [3] would be

$$R_{t-\mu} = \frac{i_t}{\sum_{f_0}^{f} i_{t-s}\Phi_s}, \qquad [4]$$

where μ is a central value of the probability distribution of the serial interval Φ that could be, for instance, the median or the mean. In the Ma et al. (16) estimate of the serial interval for Covid-19, we have $\mu \simeq 5.5$ for the median and $\mu \simeq 6.7$ for the mean. This supports the hypothesis that EpiEstim estimates R_t with an average delay of more than 5 days.

In practice, the way the sliding average of the incidence is calculated causes another delay. Indeed, as illustrated in Figure 1 the raw data of the incidence curve i_t can oscillate strongly with a seven-day period. This oscillation has little to do with the Poisson noise used in most aforementioned publications. Government statistics are affected by changes of testing and polling policies and by week-end reporting delays. These recording delays and subsequent rash corrections result in impulse noise, and a strong weekly periodic bias observable on the incidence curve (in green) on the left of figure 1.

To reliably estimate the reproduction number, a regularity constraint on R_t is needed. Cori et al., initiators of the EpiEstim method (5), use as regularity constraint the assumption that R_t is locally constant in a time window of size τ ending at time t (usually $\tau = 7$ days). This results in smoothing the incidence curve with a sliding mean over 7 days. This assumption has two limitations: it causes a significant resolution loss, and an additional $\frac{\tau}{2} = 3.5$ backward shift in the estimation of R_t, given that R_t is assumed constant in $[t-\tau, t]$.

In summary, the computation of R_t by equations Eq. (1) and Eq. (2) raises three challenges:

1. The renewal equation Eq. (1) involves future values of i_t, those for $t+1, \cdots, t-f_0$.

2. Its second form Eq. (2) used by the standard method estimates R_t with a backward shift of about 5 days.

3. Smoothing of the week-end effect causes a 3.5 days shift backward.

These cumulative backward shifts may cause a time delay of up to 8.5 days. We shall give an experimental confirmation of such delays by two independent methods: using a simulator with synthetic ground truths, and a thorough study of the incidence curves of 55 countries. The practical meaning of this

*Cori et al. (5): "We assume that, once infected, individuals have an infectivity profile given by a probability distribution w_s, dependent on time since infection of the case, s, but independent of calendar time, t. (...) R_t is the average number of secondary cases that each infected individual would infect if the conditions remained as they were at time t."

study is that the value of R_t computed by EpiEstim at time t might refer approximately to R_{t-8}[†].

Here, we address these three issues by proposing a method to invert the renewal equations Eq. (1) and Eq. (2). The inversion method developed for Eq. (1) is illustrated in Figure 1 (right), where the EpiEstim result using the renewal equation Eq. (2) (in black) is superposed with the estimate (in red) of R_t by EpiInvert using Eq. (1). After registering both, the black EpiEstim curve stops eight days before EpiInvert, the red curve. More generally we found, using the incidence curve of 55 countries, that the median of the temporal shift between the EpiEstim and EpiInvert R_t estimates using the form Eq. (1) of the renewal equation is about 8.24 days, and that the median of the RMSE approximation error between both estimates is just about 0.036.

This result is slightly surprising, given that the interpretation of R_t in both equations is different, and that the serial interval used in both equations also is different. In Eq. (2) the serial interval is indeed truncated to preserve the temporal causality of this equation. This excellent 0.036 fit nevertheless suggests that the EpiInvert method, applied to the renewal equation Eq. (1), is compatible with the EpiEstim method, but brings an information closer to present. This fact will be investigated experimentally in Sections 3 and 4.

The general integral equation [1] is a functional equation in R_t. Integral equations have been previously used to estimate R_t: in (17), the authors estimate R_t as the direct deconvolution of a simplified integral equation where i_t is expressed in terms of R_t and i_t in the past, without using the serial interval. Such inverse problems involving noise and a reproducing kernel can be resolved through the Tikhonov-Arsenin (18) variational approach involving a regularization term. This method is widely used to solve integral equations and convolutional equations (19). The solution of the equation is estimated by an energy minimization. The regularity of the solution is obtained by penalizing high values of the derivative of the solution. Our variational formulation includes the correction of the weekly periodic bias, or "weekend effect". The standard way to deal with a weekly periodic bias is to smooth the incidence curve by a seven days sliding mean. This implicitly assumes that the periodic bias is additive. The present study supports the idea that this bias is better dealt with as multiplicative. In the variational framework, the periodic bias is therefore corrected by estimating multiplicative periodic correction factors. This is illustrated on the left graphic of Fig. 1 where the green oscillatory curve is transformed into the blue filtered curve by the same energy minimization process that also computes R_t (on the right in red) and reconstructs the incidence curve up to present by evaluating the renewal equation using the computed R_t and the filtered incidence curve (on the left, in red).

In this work we use two versions of the renewal equation formulation to compute R_t. It is, however, possible to formulate statistical models for R_t that do not take into account the serial interval and the renewal equation. For instance, in (20), the author proposes to use the model:

$$log(R_t) = log(R_{t-1}) + \sigma Z_t - \alpha i_{t-1}, \qquad [5]$$

where Z_t is an independent and identically distributed sequence of standard normal random variables, σ is the dispersion of the random walk and α is the coefficient of drift. The model was fit to the provided incidence data by applying Bayesian inference on the parameter and state space with assumed prior distributions.

1. Available serial interval functions for SARS-CoV-2

As we saw, the *serial interval* in epidemiology refers to the time between successive observed cases in a chain of transmission. Du et al. in (21) define it as "the time duration between a primary case (infector) developing symptoms and secondary case (infectee) developing symptoms."

Du et al. in (21) obtained the distribution of the serial interval by a careful inquiry on 468 pairs of patients where one was the probable cause of the infection of the other. The serial distribution Φ obtained in (21) has a significant number of cases on negative days, meaning that the infectee had developed symptoms up to $f_0 = 10$ days before the infector. In addition to this first serial interval, we test a serial interval obtained by Nishiura et al. in (22) using 28 cases, which is approximated by a log-normal distribution, and a serial interval obtained by Ma et al. in (16) using 689 cases. As proposed by the authors this serial interval has been approximated by a shifted log-normal to take into account the cases in the negative days. In Fig. 2 we show the profile of the three serial intervals. There is good agreement of the serial intervals obtained by Du et al. (21) and Ma et al. (16)[‡]. Note that $f_0 = -4$ for the Ma et al. serial interval, $f_0 = 0$ for Nishiura et al. and $f_0 = -10$ for Du et al. The discrete support of Φ is therefore contained in the interval $[f_0, f]$.

We are assuming that the serial interval profile does not change across the time. As stated in (23) (equation [10]), it is nevertheless possible to use a more general form of the renewal equation (1),

$$i_t = \sum_{s=f_0}^{f} R_{t-s} i_{t-s} \Phi_{t-s,s} \quad \text{for} \quad t = 0, .., t_c, \qquad [6]$$

where $\Phi_{t-s,s}$ is the forward serial interval which takes into account that the onset of symptoms and transmission potential can jointly depend on the life history of a disease. The forward serial interval measures the time forward from symptom onset of an infector, obtained from a cohort of infectors that developed symptoms at the same time $t - s$. This more general form of the renewal equation is used in (23) to properly link the initial epidemic growth to the reproduction number $R0$. The variational approach proposed in the present work can be easily extended to compute R_t from i_t and $\Phi_{p,s}$, provided an estimation of $\Phi_{p,s}$ is available.

Transmissibility can also depend on coronavirus lineage. For instance in (24), the authors show that the SARS-CoV-2 variant B.1.1.7 has a 43 to 90% higher reproduction number than preexisting variants. It cannot be ruled out that these new variants have a different serial interval than preexisting ones.

[†] The lack of confidence in the computation of R_t is illustrated by the following fact: the official value of R_t is updated weekly and not daily by the official French online app Anticovid. This actually introduces an additional average 3.5 delay in the publication of this index!

[‡] In the online interface (www.ipol.im/epiinvert) the users can, optionally, upload their own distribution for the serial interval.

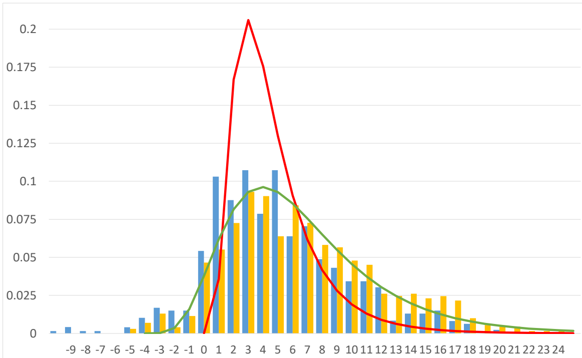

Fig. 2. Serial intervals used in our experiments: the discrete one proposed by Du et al. in (21) (solid bars in blue), the serial interval proposed by Ma et al. (16) (solid bars in orange) and its shifted log-normal approximation (in green), finally a log-normal approximation of the serial interval proposed by Nishiura et al. in (22) (in red).

2. Computing R_t by a variational method

As explained in the previous section, we aim at solving two versions of the renewal equation

$$i_t = F(R, i, \Phi, t) \quad \text{for} \quad t = 0, .., t_c, \qquad [7]$$

where

$$F = F_1 \equiv R_t \sum_{s=f_0}^{f} i_{t-s}\Phi_s; \qquad F = F_2 \equiv \sum_{s=f_0}^{f} i_{t-s}R_{t-s}\Phi_s. \quad [8]$$

F_2 corresponds to the case reproductive number formulation (equation [1]) and F_1 to the instantaneous reproduction number formulation (equation [2]). Both formulations of the renewal equation are valid, and we can apply our methodology to both. As we shall see, this leads to anticipate by several days the estimate of R_t. Equation [2] is also used in the classic Wallinga Teunis method (4), as shown in the supporting information. This last method is widely used to compute R_t retrospectively.

Correcting the week-end effect We must first formulate a compensation for the weekend effect, which in most countries is stationary, strong, and the main cause of discrepancy between i_t and its expected value $F(i, R, \Phi, t)$. To remove the weekend effect we estimate periodic multiplicative factors defined by a vector $\mathbf{q} = (q_0, q_1, q_2, q_3, q_4, q_5, q_6)$.

The variational framework we propose to estimate R_t is therefore given by the minimization of the energy

$$E(\{R_t\}; \mathbf{q}) = \sum_{t=0}^{t_c} \left(\frac{q_{t\%7} i_t - F(\{q_{t\%7} i_t\}, R, \Phi, t)}{median_{(t-\tau,t]}(i)} \right)^2 + \qquad [9]$$
$$w \sum_{t=1}^{t_c} (R_t - R_{t-1})^2$$

where $t\%7$ denotes the remainder of the Euclidean division of t by 7, $t = 0$ represents the beginning of the epidemic spread and t_c the date of the last available incidence value.

The weekend effect has varied over the course of the pandemic. Hence, for the estimate of $\mathbf{q}$ it is better to use a time interval $[t_c - T + 1, T]$ where T is fixed in the experiments to $T = 56$ (8 weeks). This two months time interval is long enough to avoid overfitting and small enough to ensure that the testing policy has not changed too much. The optimization of R_t is instead performed through the whole time interval $[0, t_c]$. The corrected value $\hat{i}_t = q_{t\%7} i_t$ amounts to a deterministic attenuation of the weekend effect on i_t. An obvious objection is that this correction might not be mean-preserving. To preserve the number of accumulated cases in the period of estimation, we therefore add the constraint

$$\sum_{t=t_c-T+1}^{t_c} i_t = \sum_{t=t_c-T+1}^{t_c} \hat{i}_t = \sum_{t=t_c-T+1}^{t_c} q_{t\%7} i_t, \qquad [10]$$

to the minimization problem [9].

In that way, the multiplication by the factor $q_{t\%7}$ produces a redistribution of the cases i_t during the period of estimation, but it does not change the global amount of cases. In Equation [9], $median_{(t-\tau,t]}(i)$ is the median of i_t in the interval $(t-\tau, t]$ used to normalize the energy with respect to the size of i_t. In the experiments we use $\tau = 21$. The first term of E is a data fidelity term which forces the renewal equation [7] to be satisfied as much as possible. The second term is a classic Tikhonov-Arsenin regularizer of R_t. As in the case of EpiEstim, this method provides a real-time estimate of R_t up to the date, t_c, of the last available incidence value. Yet, in contrast with EpiEstim, this method takes advantage for $t < t_c$ of the knowledge of the incidence curve $i_{\bar{t}}$ for $\bar{t} \in [t, t_c]$. This improves the posterior accuracy of the R_t estimate.

The regularization weight. The regularization weight $w \geq 0$ is a dimensionless constant weight fixing the balance between the data adjustment term and the regularization term.

Boundary conditions of the variational model. Since $t = 0$ is the beginning of the epidemic spread where the virus runs free, one is led to use an estimate of $R_0 = R0$ based on the basic reproduction number $R0$. (In the supporting information we present a basic estimation of $R0$ from the initial exponential growth rate of the epidemic obtained as in (25)), therefore, to solve Equation [9], we add the boundary condition $R_0 = R0$. The proposed inversion model provides an estimation of R_t up to the date, t_c, of the last available incidence value. Yet if $f_0 < 0$, the functional [9] involves a few future values of R_t and i_t for $t_c \leq t \leq t_c - f_0$. These values are unknown at present time t_c. We use a basic linear regression using the last seven values of i_t to extrapolate the values of i_t beyond t_c. We prove in the supporting information, that the boundary conditions and the choice of the extrapolation procedure have a minor influence in the estimation of R_t in the last days when minimizing [9].

All of the experiments described here can be reproduced with the online interface available at www.ipol.im/epiinvert. This online interface allows one to assess the performance of the method applied to the total world population and to any country and any state in the USA, with the last available data. We detail our daily sources in the supporting information.

An empirical confidence interval for R_t. In absence of a statistical model on the distribution of R_t, no theoretical *a priori* confidence interval for this estimate can be given. Nevertheless, a realistic confidence interval is obtained by the following procedure: let us denote by $R_t^{\tilde{t}}$ the EpiInvert estimate at time t using the incidence curve up to the date $\tilde{t} \geq t$. Therefore

$R_t^{t_c}$ represents the final EpiInvert estimate of R_t using the incidence data up to the last available date t_c. As shown below using the real and simulated data, for $F \equiv F_2$, R_t^{t+k} stabilizes for $k \geq 8$. We can therefore consider R_t^{t+8} as an approximation of the reproduction number ground truth. We want to provide an empirical confidence interval $I_t = [R_t^{t_c} - r(t), R_t^{t_c} + r(t)]$ such that 95% of times $R_t^{t+8} \in I_t$ (for $t = t_c, t_c - 1, ..., t_c - 7$). To define $r(t)$ we use, on the one hand, a measure of the variation of R_t in the last few days given by

$$\sigma(t) = \sqrt{\frac{\sum_{n=1}^{N} (R_t^{t_c} - R_t^{t_c - n})^2}{N}}, \qquad [11]$$

where $R_t^{t_c-n}$ in $(t_c - n, t_c]$ are obtained by linear extrapolation. In our experiments we use $N = 3$. On the other hand, we use, supported by results obtained below for real and simulated data, that the error in the estimation of R_t grows linearly when t approaches t_c (the last time at which i_t was available). Combining $\sigma(t)$ with a linear function with respect to $(t_c - t)$ we obtain the following expression for $r(t)$:

$$r(t) = \sigma(t) + (B - C(t_c - t))_+ \qquad [12]$$

where B and C are parameters of the estimation and $(x)_+ \equiv max(0, x)$. The advantage of this empirical approach is that the estimation of the confidence interval is adapted to the variation of R_t in the last few days. Using 16500 experiments on real data corresponding to R_t estimations on 300 different values for the last used day, t_c, in 55 countries, we obtain, in the case of $F \equiv F_2$, that using $B = 0.24$ and $C = 0.03$, 95% of times $R_t^{t+8} \in I_t$ (for $t = t_c, t_c - 1, ..., t_c - 7$). In the same way for the empirical 90% confidence interval we obtain $B = 0.16$ and $C = 0.022$. If we consider now $F \equiv F_1$, then it is observed that R_t^{t+k} stabilizes for $k = 3$. Using it as ground truth, the obtained empirical confidence intervals for R_t^{t+3} are given by $B = 0.04$ and $C = 0.016$ (in the case of 95%) and $B = 0.02$ and $C = 0.009$ (in the case of 90%) These empiric intervals are displayed for each t in the online algorithm www.ipol.im/epiinvert and have the aspect of fattened curves above and below R_t.

Efficiency measure of the weekly bias correction. We estimate the correction of the weekly periodic bias by the efficiency measure

$$\mathcal{I} = \sqrt{\frac{\sum_{t=t_c-T+1}^{t_c} \left(\hat{i}_t - F(\hat{i}, R, \Phi, t)\right)^2}{\sum_{t=t_c-T+1}^{t_c} \left(i_t - F(i, R1, \Phi, t)\right)^2}}. \qquad [13]$$

$\mathcal{I}$ represents the reduction factor of the RMSE between the incidence curve and its estimate using the renewal equation after correcting the week-end bias. $\hat{i}_t = i_t q_{t\%7}$ and R are the optimal values for the energy [9] and $R1$ denotes the R estimate without correction of the weekly bias. The value of $\mathcal{I}$ can be used to assess whether it is worth applying the correction of the weekly periodic bias to a given country in a given time interval.

Estimation of the temporal shift between EpiEstim and EpiInvert. In what follows, we will denote by R_t^{Epi} the EpiEstim estimation of the reproduction number by Cori et al. in (5), detailed in the supporting information. As we have argued above, we expect a significant temporal shift between the EpiInvert estimate of R_t and R_t^{Epi}, of the order of 9 days. This expectation is strongly confirmed by the experimental results, and can be checked by applying the proposed method to any country using the online interface available at www.ipol.im/epiinvert. In summary, the time shift between both methods should be a half-week (3.5 days) for $F \equiv F_1$ and by Equation [4] of about $\mu + 3.5 \simeq 9$ for $F \equiv F_2$. This will be verified experimentally by computing the shift $\tilde{t}$ between R_t^{Epi} and R_t yielding the best RMSE between both estimates:

$$\tilde{t} = \arg\min_{t \in [0,12]} \mathcal{S}(t) \equiv \sqrt{\frac{\sum_{k=t_c-T+1}^{t_c} (R_{k-t} - R_k^{Epi})^2}{T}} \qquad [14]$$

where $T = 56$ (8 weeks) and where we evaluate R_{k-t} for non-integer values of $k - t$ by linear interpolation.

Summary of the algorithm parameters and options.

- choice of the serial interval : the default options are the serial intervals obtained by Ma et al. (we use the shifted log-normal approximation), Nishiura et al. and Du et al.. The users can also upload their own serial interval;
- choice of the renewal equation used, $F \equiv F_1$ or $F \equiv F_2$;
- Correction of the weekly periodic bias (option by default)

The regularization weight w is always fixed to 5, the value we obtain below, experimentally, by comparing with EpiEstim.

Summary of the output displayed at www.ipol.im/epiinvert. First we draw two charts. In the first one we draw R_t and R_t^{Epi} shifted back $\tilde{t}$ days where $\tilde{t}$ is defined in [14]. R_t is surrounded by a shaded area that represents the above defined empirical confidence intervals. In the second chart, we draw the initial incidence curve i_t in green, the incidence curve after the correction of the weekly periodic bias $\hat{i}_t = i_t q_{t\%7}$ in blue, and the evaluation of the renewal equation given by $t \to F(\hat{i}, R, \Phi, t)$ in red. For each experiment we also compute :

1. R_{t_c} : last available value of the EpiInvert R_t estimate.
2. $R_{t_c}^{Epi}$: last available value of the EpiEstim estimate R_t^{Epi}.
3. $\tilde{t}$: optimal shift (in days) between R and R^{Epi} defined in [14].
4. $\mathcal{S}(\tilde{t})$: RMSE between R and R^{Epi} shifted back $\tilde{t}$ days (defined in [14]).
5. $\mathcal{V}(i)$: variability of the original incidence curve, i_t, given by :

$$\mathcal{V}(i) \equiv \frac{\|i'\|_{L^1[t_c-T, t_c]}}{\|i\|_{L^1[t_c-T, t_c]}} \approx \frac{\sum_{t=t_c-T+1}^{t_c} |i_t - i_{t-1}|}{\sum_{t=t_c-T+1}^{t_c} i_t} \qquad [15]$$

6. $\mathcal{V}(\hat{i})$: variability of the filtered incidence $\hat{i}_t$ after the correction of the weekly periodic bias.
7. $\mathcal{I}$: reduction factor of the RMSE error between the incidence curve and its estimate using the renewal equation after the correction of the weekly periodic bias (defined in [13]).
8. $\mathbf{q} = (q_0, .., q_6)$: the correction coefficients of the weekly periodic bias (q_6 corresponds to the t_c, the last time at which i_t was available).

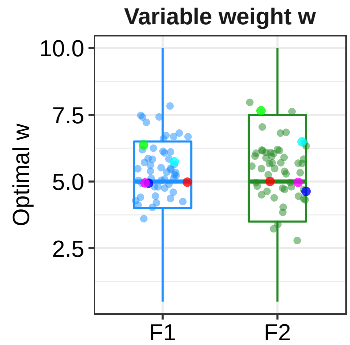

Fig. 3. Distribution of w for F_1 and F_2 when the regularization weight w and the delay $\tilde{t}$ are optimized independently for each country to minimize the average error $\mathcal{S}(\tilde{t})$ between the EpiEstim and the EpiInvert methods on a time lapse of 56 days. France in blue, Japan in green, Peru in cyan, South Africa in magenta, USA in red.

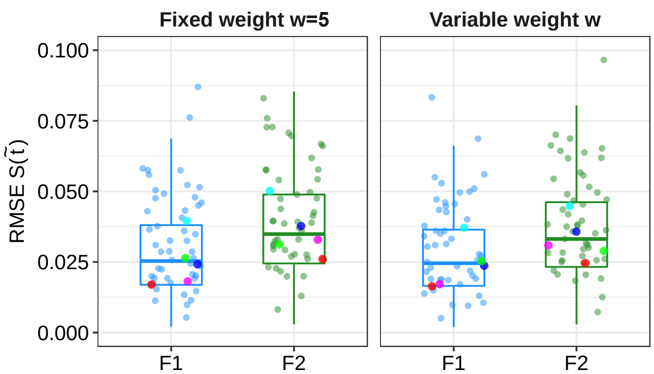

Fig. 4. Average error $\mathcal{S}(\tilde{t})$ between the EpiEstim and the EpiInvert estimates of R_t for each country. On the left w is fixed and on the right it is the optimal weight per country. France in blue, Japan in green, Peru in cyan, South Africa in magenta, USA in red.

3. Results on incidence curves from 55 countries

To estimate a reference value for the regularization parameter w we used the incidence data up to July 17, 2021 for the 55 countries showing the larger number of cases. For each country, we performed 30 experiments. Starting with the incidence data up to July 17, in each experiment we removed the last 10 days from the incidence data used in the previous experiment. In that way we got a large variety of real epidemic scenarios. We optimized the RMSE $\mathcal{S}(\tilde{t})$ between R_t and R_t^{Epi} shifted back $\tilde{t}$ days (defined in [14]). This optimization was performed with respect to w and $\tilde{t}$. The goal was to fix w, the only parameter of the method, so that the result of EpiInvert is as close as possible to EpiEstim in the days where both methods predict R_t. The second goal of this optimization was to estimate the effective time shift $\tilde{t}$ between both methods.

In Fig. 3 we show the box plot of the distribution of w for F_1 and F_2 when w was optimized independently for each experiment to minimize the average error over 56 days between the EpiEstim and the EpiInvert methods. The median of the distribution of w is 5 for F_1 and F_2 which indicated that a common value of w could be fixed for all countries. Here and in all figures to follow, each dot represents the average of all experimental results associated to a country.

In Fig 4, we show, for the versions F_1 and F_2 of the renewal

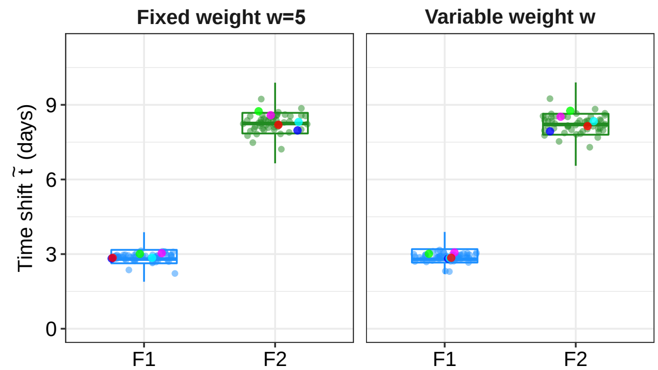

Fig. 5. Optimal time shift $\tilde{t}$ obtained by minimizing the mean error $\tilde{S}(t)$ over 56 days between the EpiEstim and the EpiInvert estimates of R_t for each country. The time shift is, as predicted by our theoretical analysis, close to 3 days for F_1 and slightly above 8 days for F_2. On the left w is fixed and on the right it is the optimal weight per country. France in blue, Japan in green, Peru in cyan, South Africa in magenta, USA in red.

equation, the average error $\mathcal{S}(\tilde{t})$ over 56 consecutive days of the error between the EpiEstim and the EpiInvert estimates of R_t for each country. The median of the overall average error is 0.025 for F_1 and 0.034 for F_2.

As shown in Fig. 4, the agreement between R_t and R_t^{Epi} shifted back by the optimal delay $\tilde{t}$ is overwhelming. As is apparent by comparing the box plots on the left and right, the increase of the error $\mathcal{S}(\tilde{t})$ was insignificant when fixing $w = 5$ for all countries ("fixed weight") instead of optimizing jointly on w and $\tilde{t}$ for all countries ("variable weight"). In all experiments, we therefore fixed the value of w to 5 for all countries. Once fixed, we optimized again $\mathcal{S}(\tilde{t})$ with respect to $\tilde{t}$.

In the box plot of Fig. 5 we show, for the versions F_1 and F_2 of the renewal equation, the optimal time shift $\tilde{t}$ obtained by minimizing the mean error $\tilde{S}(t)$ over 56 days between the EpiEstim and the EpiInvert estimates of R_t for each country. As is apparent by comparing the box plots on the left and right, there is almost no change on $\tilde{t}$ when fixing w for all countries ("fixed weight") instead of optimizing jointly on w and $\tilde{t}$ for all countries. We obtain respectively $\tilde{t} = 2.88 \pm 0.47$ for variable w and $\tilde{t} = 2.87 \pm 0.49$ for F_1 with fixed w, and similarly for F_2: $\tilde{t} = 8.24 \pm 0.82$ and $\tilde{t} = 8.27 \pm 0.80$. These results are in good agreement with the discussion about the EpiEstim method we have presented above, which led to predict a time delay of 3.5 days for $F \equiv F_1$ and more than 8 days for $F \equiv F_2$. The difference between the predicted time delay and the observed one therefore is about 0.5 days. This is easily explained by the regularization term in Equation [9], which forces R_t to resemble R_{t-1}. In summary, these experiments show that EpiEstim predicts at time t a value R_t which corresponds to day $t - 8.5$ or $t - 3.5$, and that EpiInvert predicts at time t a value R_t which corresponds to day $t - 0.5$.

We now explore the internal coherence of the EpiInvert predictions. Let us denote by $R_t^{\tilde{t}}$ the EpiInvert estimate at time t using the incidence curve up to the date $\tilde{t} \geq t$. Since the estimate of EpiInvert at each day evolves with the knowledge of the incidence in later days, when $\tilde{t}$ increases, the estimation $R_t^{\tilde{t}}$ becomes more accurate and, as shown later using simulated data, we can consider that $R_t^{\tilde{t}}$ stabilizes and approaches the final estimation when $\tilde{t} = t + 3$ for $F \equiv F_1$ and $\tilde{t} = t + 8$

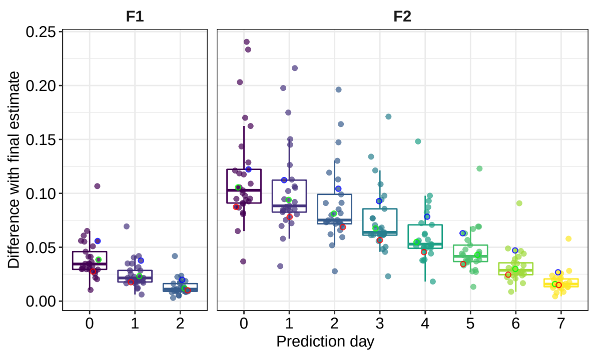

Fig. 6. Internal relative error between the EpiInvert estimations depending on the prediction day k. Each dot represents the average value on 300 experiments performed on one country for different values of t. On the left, for $F \equiv F_1$, we compare for $k = 0, 1, 2$, the relative errors $|R_t^{t+k} - R_t^{t+3}|$. On the right, for $F \equiv F_2$, we compare, in the same way, $|R_t^{t+k} - R_t^{t+8}|$ for $k = 0, .., 7$. For $F \equiv F_2$, we see that $|R_t^{t+k} - R_t^{t+8}|$ goes down almost linearly with respect to k. France in blue, Japan in green, Peru in black, South Africa in magenta, USA in red. The robustness of the prediction is positively affected by incidence numbers.

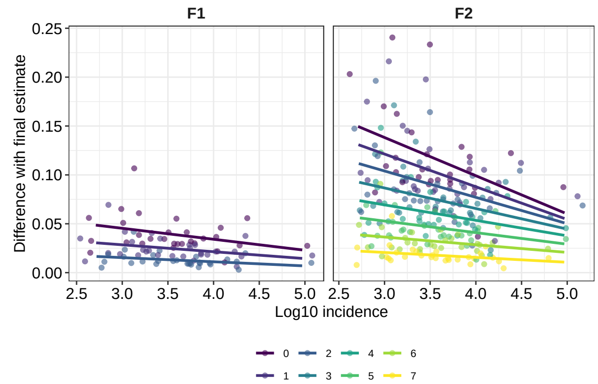

Fig. 7. Linear regression of the internal relative error between the EpiInvert estimation as a function of the mean incidence. The regression lines are clearly decreasing, which means that a higher incidence favors a better estimate of R_t.

for $F \equiv F_2$. Fig. 6 gives a box plot of the distributions of the internal relative error between the EpiInvert estimations depending on the prediction day k.

Fig. 7 shows, for each prediction day $k = 0, 1, ..,$ the linear regression of the internal relative error between the EpiInvert estimations, viewed as a function of the mean incidence of the country. These regression lines are clearly decreasing, which means that a higher incidence favors a better estimate of R_t. Last but not least, we evaluate the reduction obtained on the "week-end effect". Fig. 8 shows a regression plot of the RMSE reduction factor $\mathcal{I}$ (see [13]) obtained by applying correcting coefficients to reduce the "week-end effect". This reduction decreases from about 0.7 to 0.4, the plots being ordered in increasing order of average incidence. This indicates that higher incidences lead to a more regular 7 days periodicity of the week-end effect. In https://ctim.ulpgc.es/covid19/BoxPlots/ Fig. 6, 7 and 8 are presented in interactive mode with tooltip detailed statistics on each country.

4. Validation on epidemic simulations

To evaluate the accuracy of the proposed technique, we used simulated data where the ground truth for R_t (that we denote by R_t^{GT}) is similar to the one proposed in Gostic et al. (11). This ground truth simulates the impact of a strict lockdown at the beginning of the epidemic spread. Initially, $R_t^{GT} = R_0 > 1$, then, a strict lockdown is implemented at time $t = 0$ and R_t^{GT} becomes $R_i < 1$. After t' days (the lockdown duration) the social-distancing measures start relaxing to keep the R_t value stabilized around 1. The parameters to define R_t^{GT} are R_0, R_i, t' and s, which determines the slope of the transitions between R_0 and R_i and between R_i and 1. the larger s, the sharper the transition. For a technical description of the definition of R_t^{GT}, see the supporting information. The ground truth of the incidence curve, that we denote by i_t^{GT}, is computed from the renewal equation using R_t^{GT} as reproduction number. Since the ground truth of the incidence curve is defined up to the multiplication by a constant factor, the simulator allows users

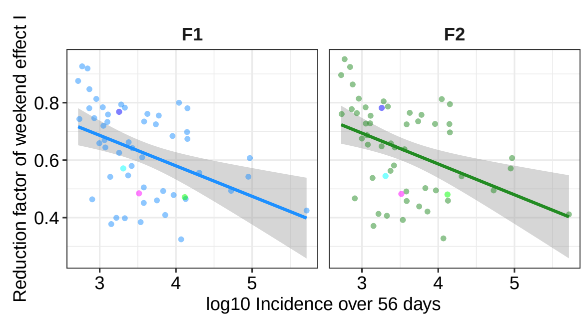

Fig. 8. Reduction factor $\mathcal{I}$ (see [13]) obtained by applying correcting coefficients to reduce the "weed end effect". This reduction decreases from about 0.7 to less than 0.4. The plots are ordered in increasing order of average incidence.

to tune the additional parameter i_{max}, which represents the maximum value of the incidence curve in the whole period. We simulated the observed incidence curve i_t assuming that $i_t = \mathcal{P}(i_t^{GT} q'_{t\%7})$ follows a Poisson distribution of mean $i_t^{GT} q'_{t\%7}$ where $\mathbf{q}' = (q'_0, .., q'_6)$ is the vector of the weekly bias correction factors.

To simulate the weekly bias the simulator proposes 19 options of real bias correction factors, pre-estimated using the incidence curve of 19 countries. Note that, in agreement with the Poisson model, the weekly bias is applied first on the deterministic incidence curve. It is followed by the Poisson simulation, which takes this biased deterministic value as parameter. The simulator finally uses EpiInvert and EpiEstim to compute R_t from the biased Poisson process realization i_t. An online implementation of this simulator is available at www.ipol.im/episim.

A statistical analysis of the results was performed on 4800 simulations, obtained by varying regularly the parameters $R_0 \in [1.5, 2]$, $R_i \in [0.5, 0.8]$, $s \in [0.1, 2.]$, $i_{max} \in [1000, 30000]$, $t' = 28$, pre-estimated weekly bias from 19 countries, and the extra option of not applying weekly bias. The regularization parameter was fixed to $w = 5$, which is the optimal value obtained with real data. In the case of $F \equiv F_1$, we compared the ground truth R_t^{GT} with R_t^{Epi} (the EpiEstim estimation), R_t^t (the EpiInvert estimation using i_t up to the time t), R_t^{t+3} (the EpiInvert estimation at time t using i_t up to the time $t+3$), and the final estimate $R_t^{t_c}$. In the simulator, $t_c > t+8$ is the last day used in the simulation, which depends on the lockdown duration. In the case of $F \equiv F_2$, we compared the ground truth with R_t^{Epi}, R_t^t, R_t^{t+4}, R_t^{t+8} and $R_t^{t_c}$.

Fig. 10 shows a thorough comparison on a lockdown scenario of the results of the R_t estimation methods. These simulations confirm the theoretically anticipated time delays between the various considered estimates of R_t. Contrarily to EpiEstim, EpiInvert updates the estimated values of R_t when days pass by. This estimate of R_t obtained k days later, denoted by R_t^{t+k}, stabilizes near the (blue) ground truth R_t^{GT} for $k = 8$ (B, red) with the F2 model, and for $k = 3$ for the F1 model (A, red). Indeed it uses, for each t, the incidence values up to 8 days (resp. 3 days) later. Nevertheless, for the F1 model, the timely estimate R_t^t (A, orange) is very close to the ground truth R_t^{GT} (A, blue) and much closer to it than the EpiEstim estimate (A, black). For the F2 model, the timely estimate R_t^t (B, orange) is not that close to R_t^{GT}. Indeed, the estimation uses the values of i_t up to time t, so there is only partial information to compute the reproduction number, which still depends on the future values of i_t. Yet, the R_t^{t+4} estimate (B, magenta), which is delayed by only 4 days, is considerably closer to the ground truth (in blue), and R_t^{t+8} is still closer. Observe that EpiEstim (in black) provides by far the worst estimation of R_t^{GT}.

In Table 1, we show the distributions of the optimal time delay between R_t^{GT} and its various estimates by minimizing the RMSE between both curves. For the F1 model, the EpiEstim estimate shifted back by 2.65 days has an error of 0.053. R_t^t, computed on the same day by EpinInvert, has an error of 0.044 with a delay of just 0.84 days. In short, EpiInvert gets a better estimate 2 days in advance with respect to EpiEstim. A similar conclusion arises for the F_2 model. EpiEstim, when shifted back by 8.42 days, has an error of 0.108. Waiting for 8 days and shifting back by 0.44 days the result of R_t^{t+8} yields an inferior error, 0.075. But a result almost as good (0.078) is obtained by taking the result of R_t^t and shifting it back by 5.41 days. There is no particular gain in waiting longer for better estimates of EpinInvert : the estimate does not improve with time and is stack at 0.075. In summary, this result (based on simulations) leads to the following recommendations:
a) The EpiEstim estimate at time t, R_t must be shifted back by 8.42 days;
b) The EpinInvert synchronous estimate R_t^t made at time t must be shifted back by 5.41 days. It is more precise than the EpiEstim estimate (an 0.078 error against 0.108) and it is obtained three days earlier (a 5.41 days delay against 8.42);
c) Nevertheless, as we have seen in Fig. 6, the EpiInvert ex post estimate $k \to R_t^{t+k}$ stabilizes after 5 days to a value which is very close to the ground truth, without the need for shifting back its value.

In the above estimates, replacing R_t^{t+8} by $R_t^{t_c}$ does not change this conclusion. The difference between these estimates is negligible. Indeed, R_t^{t+k} no longer varies for $k \geq 8$.

5. CONCLUSION

The reproduction number R_t can be estimated by solving a renewal equation linking R_t, the case incidence curve i_t, and the serial interval Φ_s. We considered two formulations of the renewal equation. The first one ($F \equiv F_1$) estimates the instantaneous reproduction number. The second one ($F \equiv F_2$) estimates the case reproductive number. Resolving these equations is challenging, because the daily incidence data i_t recorded by health administrations is noisy and shows a strong quasi-periodic behavior. In order to get an estimate of R_t we introduced a classic regularity constraint on R_t and we corrected the weekly periodic bias observed in the incidence curve i_t by a simple variational formulation. Our proposed variational model, EpiInvert, also computes empirical confidence intervals. In contrast to former methods, EpiInvert can use serial intervals with distributions containing negative days (as it is the case for COVID-19). Thus, it avoids an artificial truncation of the serial interval, and it provides an estimate that improves with time. Nevertheless, as shown on simulations and on real incidence data, EpiInvert shows excellent agreement with EpiEstim. Its main improvement is the reduction of the time shift between the estimation and the actual value of R_t. If we compare EpiEstim with the EpiInvert estimate R_t^t (the estimation of R_t that uses the incidence values up to the time t), EpiInvert provides an estimate of R_t about 2 days in advance for the instantaneous reproduction number, and about 3 days in advance for the case reproductive number. This means that for both models, EpiInvert can anticipate by several days an estimate of R_t. This estimate is more precise than the EpiEstim estimate. In addition, we proved on a simulator that the EpiInvert estimate R_t^{t+8}, obtained 8 days later than the current date, t, is very close to the ground truth. Comparing it with the EpiEstim estimate confirms that the time delay of EpiEstim is about 3 days for the instantaneous reproduction number ($F = F_1$), and more than 8 days for the case reproductive number ($F = F_2$). Finally, comparing the EpiEstim and the EpiInvert estimated curves of R_t on real data confirms these 3 and 8 days delays between both estimates. These facts are extremely relevant, given that the control of social distancing policies requires a timely estimate of R_t.

	F_1	F_1	F_1	F_1	F_1	F_1	F_2	F_2	F_2	F_2	F_2	F_2	F_2	F_2	F_2	F_2	F_2
R_t	R_t^{Epi}	R_t^t	R_t^{t+1}	R_t^{t+2}	R_t^{t+3}	$R_t^{t_c}$	R_t^{Epi}	R_t^t	R_t^{t+1}	R_t^{t+2}	R_t^{t+3}	R_t^{t+4}	R_t^{t+5}	R_t^{t+6}	R_t^{t+7}	R_t^{t+8}	$R_t^{t_c}$
$\tilde{t}$	2.65	0.87	0.41	0.27	0.23	0.2	8.42	5.41	4.43	3.46	2.52	1.62	1.41	0.59	0.49	0.44	0.37
$\mathcal{S}(0)$	0.140	0.071	0.049	0.045	0.046	0.049	0.280	0.212	0.179	0.148	0.122	0.103	0.091	0.086	0.083	0.081	0.080
$\mathcal{S}(\tilde{t})$	0.053	0.044	0.040	0.041	0.043	0.047	0.108	0.078	0.077	0.077	0.076	0.075	0.075	0.075	0.075	0.075	0.075

Table 1. Statistical results for the different estimation results of R_t on 4800 simulations for the F1 and F2 models. First row: the renewal equation model used. Second Row: the R_t estimate. Third row: the optimal shift $\tilde{t}$ minimizing the RMSE between R_t^{GT} and the R_t estimate. Fourth row: the median of $\mathcal{S}(0)$, the RMSE without temporal shift. Fifth row: the RMSE $\mathcal{S}(\tilde{t})$ after applying the optimal temporal shift.

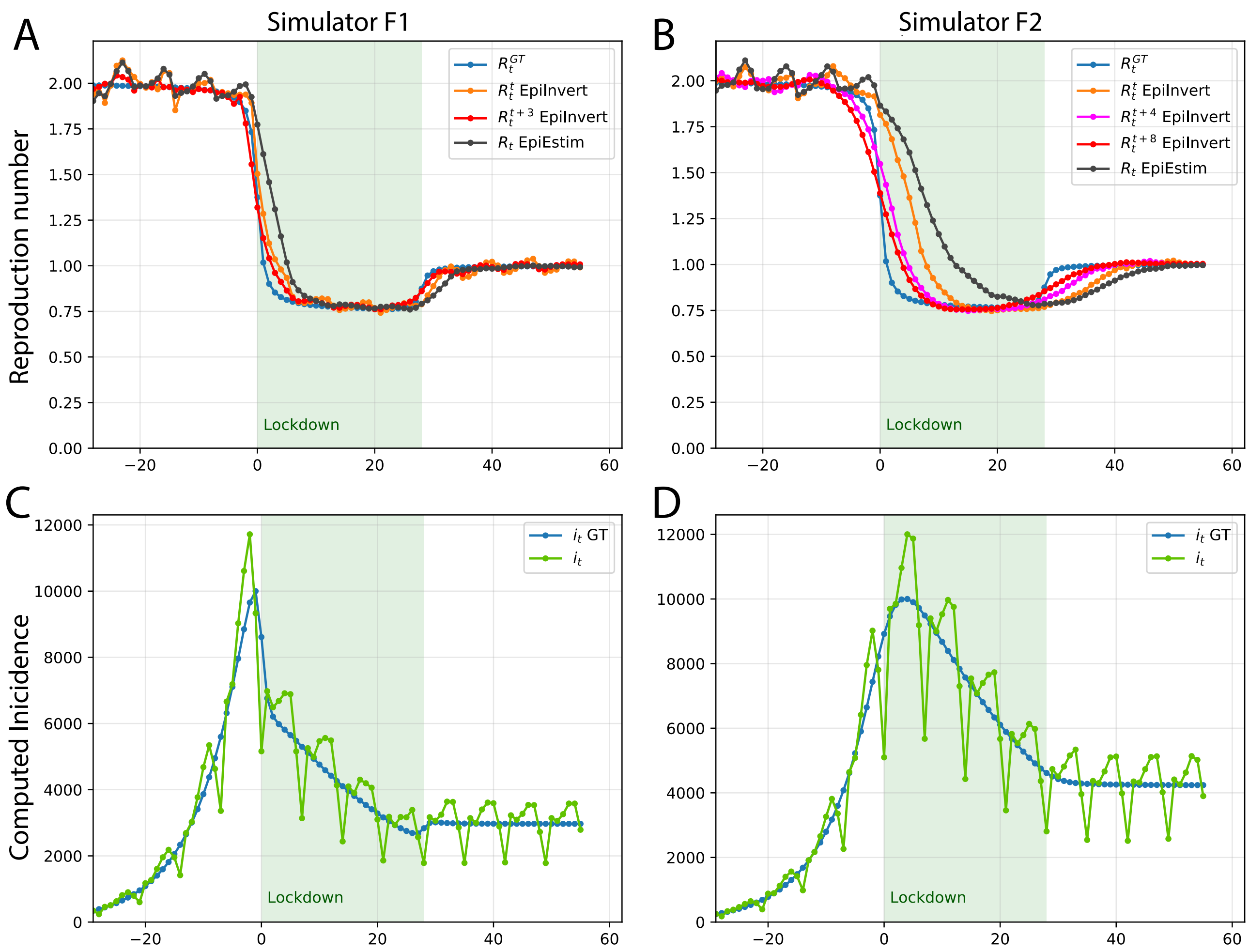

Fig. 9. Comparison on a lockdown scenario of the results of the R_t estimation methods. The R_t^{GT} ground truth parameters are $R_0 = 2$ (reproduction number before the lockdown), $R_i = 0.75$ (reproduction number after the lockdown), $t' = 28$ (lockdown duration), $s = 0.5$ (slope of the transition between R_0 and R_i), $i_{max} = 10000$ (maximum of the incidence curve), and a weekly periodic bias borrowed from the USA. The simulations and inversions were performed in A and C for $F \equiv F_1$ and in B and D for $F \equiv F_2$. Note that the same R_t^{GT} scenario (blue curve in A and B) leads to very different incidence curves (in green) in C an D. Hence, the results of the $F1$ and $F2$ inversions in A and B cannot be compared. In A and B, R_t^{t+k} denotes the EpiInvert estimate of R_t obtained k days later.

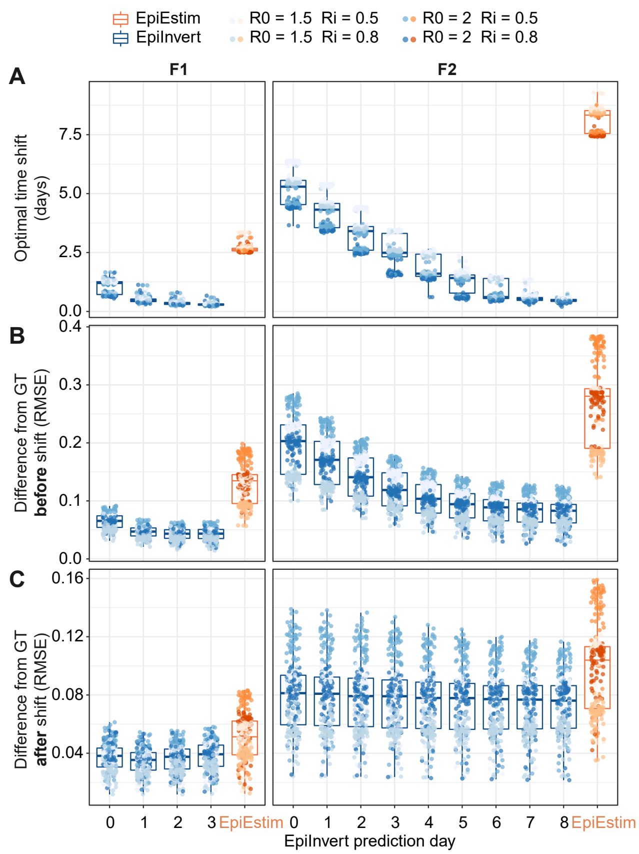

Fig. 10. Distributions of the optimal time shift (A), the RMSE before the time shift (B) and the RMSE after the time shift (C) between the ground truth R_t^{GT} and its approximation obtained by EpiEstim and by EpiInvert, as a function of the number of days k after t used in the prediction. Note that columns F1 and F2 cannot be compared. Indeed, as illustrated in Figure 9, the incidence simulations for a same ground truth R_t^{GT} are quite different. The result of EpinInvert, which evolves with time, converges near-perfectly to the ground truth after 8 days (resp. 3 days) for F2 and F1 respectively. The EpiEstim result is static and stays 30 to 40% away from the ground truth. As argued in section 1, this large relative error in the F2 model can be compensated by shifting back the estimate by 8.5 days.

1. P Rodpothong, P Auewarakul, Viral evolution and transmission effectiveness. *World J. Virol.* **1**, 131 (2012).
2. X He, et al., Temporal dynamics in viral shedding and transmissibility of covid-19. *Nat. medicine* **26**, 672–675 (2020).
3. P Ashcroft, et al., Covid-19 infectivity profile correction. *arXiv preprint arXiv:2007.06602* (2020).
4. J Wallinga, P Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. epidemiology* **160**, 509–516 (2004).
5. A Cori, NM Ferguson, C Fraser, S Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. journal epidemiology* **178**, 1505–1512 (2013).
6. AJ Lotka, Relation between birth rates and death rates. *Science* **26**, 21–22 (1907).
7. H Nishiura, Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918–19. *Theor. Biol. Med. Model.* **4**, 20 (2007).
8. H Nishiura, G Chowell, *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends*, eds. G Chowell, JM Hyman, LMA Bettencourt, C Castillo-Chavez. (Springer Netherlands, Dordrecht), pp. 103–121 (2009).
9. C Fraser, Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS ONE* **2**, 1–12 (2007).
10. K Gostic, et al., Practical considerations for measuring the effective reproductive number, Rt. *MedRxiv* (2020).
11. KM Gostic, et al., Practical considerations for measuring the effective reproductive number, r t. *PLoS computational biology* **16**, e1008409 (2020).
12. R Thompson, et al., Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29**, 100356 (2019).
13. QH Liu, et al., Measurability of the epidemic reproduction number in data-driven contact networks. *Proc. Natl. Acad. Sci.* **115**, 12680–12685 (2018).
14. T Obadia, R Haneef, PY Boëlle, The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC medical informatics decision making* **12**, 147 (2012).
15. TZ Boulmezaoud, L Alvarez, M Colom, JM Morel, A Daily Measure of the SARS-CoV-2 Effective Reproduction Number for all Countries. *Image Processing On Line* **10**, 191–210 (2020) https://doi.org/10.5201/ipol.2020.304.
16. S Ma, et al., Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *Medrxiv* (2020).
17. J Demongeot, K Oshinubi, H Seligmann, F Thuderoz, Estimation of daily reproduction rates in covid-19 outbreak. *medRxiv* (2021).
18. AN Tikhonov, VY Arsenin, Solutions of ill-posed problems. *New York* **1**, 30 (1977).
19. M Benning, M Burger, Modern regularization methods for inverse problems. *Acta Numer.* **27**, 1–111 (2018).
20. J Asher, Forecasting ebola with a regression transmission model. *Epidemics* **22**, 50–55 (2018) The RAPIDD Ebola Forecasting Challenge.
21. Z Du, et al., The serial interval of COVID-19 from publicly reported confirmed cases. *medRxiv* (2020).
22. H Nishiura, NM Linton, AR Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections. *Int. journal infectious diseases* (2020).
23. SW Park, et al., Forward-looking serial intervals correctly link epidemic growth to reproduction numbers. *Proc. Natl. Acad. Sci.* **118** (2021).
24. NG Davies, et al., Estimated transmissibility and impact of sars-cov-2 lineage b.1.1.7 in england. *Science* **372** (2021).
25. L Alvarez, Comparative analysis of the first wave of the COVID-19 pandemic in South Korea, Italy, Spain, France, Germany, the United Kingdom, the USA and the New-York state. MedRxiv (2020).

Supporting Information

In this section we describe and analyze the EpiEstim method and its parameters (Section A). In Section B the Wallinga-Teunis method. Section C presents implementation details of EpiInvert. Section D shows some technical details on our experiments on simulated data. Section D makes a case study of the USA, France, Japan, Peru and South Africa.

A. The EpiEstim method. One of the most widely used methods to estimate the instantaneous reproduction number is the EpiEstim method proposed by Cori et al. (5). In what follows, we will denote by R_t^{Epi} the EpiEstim estimation. The authors show that if i_t follows a Poisson distribution with expectation $\lambda = \mathbf{E}[i_t] = R_t^{Epi} \sum_{s=1}^{t} i_{t-s}\Phi_s$ and R_t^{Epi} is assumed to follow a gamma prior distribution $\Gamma(a,b)$, then the following analytical expression can be obtained for the posterior distribution of R_t^{Epi}:

$$R_{t,\tau}^{Epi} = \frac{a + \sum_{s=t-\tau+1}^{t} i_s}{b^{-1} + \sum_{s=t-\tau+1}^{t} \sum_{k=1}^{f} i_{s-k}\Phi_k}, \qquad \text{[A]}$$

where R_t^{Epi} is assumed to be locally constant in a time window of size τ ending at time t. However, i_t does not follow a Poisson distribution as its local variance in most states much higher than its mean, being dominated by the weekend effect. In this method, implemented in the EpiEstim R package, a regularization of the estimation is introduced by assuming that R_t^{Epi} is constant in a time window of size τ ending at time t. We found that the parameters a and b of the prior Gamma distribution $\Gamma(a,b)$, have very little influence on the current estimation of R_t^{Epi}. Cori et al. in (5) proposed to use $a = 1$ and $b = 5$. Taking into account the magnitude of the current number of daily cases in countries affected by Covid-19, the contribution of a and b to the expression [A] can be neglected. As shown in (15), assuming that the mean ab of the prior Gamma distribution $\Gamma(a,b)$ satisfies

$$ab = \frac{\sum_{s=t-\tau+1}^{t} i_s}{\sum_{s=t-\tau+1}^{t} \sum_{k=1}^{f} i_{s-k}\Phi_k}, \qquad \text{[B]}$$

equation [A] becomes

$$R_{t,\tau}^{Epi} = \frac{\bar{i}_{t,\tau}}{\sum_{k=1}^{f} \bar{i}_{t-k,\tau}\Phi_k} \qquad \text{[C]}$$

which corresponds to the usual R_t^{Epi} estimate obtained directly from equation [2] applied to $\bar{i}_t$, where $\bar{i}_t$ is the average of i_t in the interval $[t-\tau, t]$. Therefore, if we remove the parameters a and b from the estimation of R_t^{Epi}, the main difference between the EpiEstim estimation and the one proposed here for $F \equiv F_1$ is that in EpiEstim, a serial interval with non-positive values is not allowed and that the regularity is forced by a backward seven day average of the incidence curve. This is replaced by a regularity term in the proposed variational formulation. Notice that due to the backward averaging of the incidence curve, we can expect a time shift between both estimations.

B. The Wallinga and Teunis computation of R_t. The Wallinga-Teunis (4) method is also implemented in the EpiEstim package and widely considered as a reliable method to compute the case reproduction number, R_t^c, retrospectively (11). Its formulas to estimate R_t^c at time t require the knowledge of i_t for $t = 0, \cdots, t+f$. Starting from the original definitions of the authors, we give a mathematical proof that their method is actually computing R_t by the F_1 form of the renewal equation. The method is based on the following estimation of the "relative likelihood, $p_{k,l}$, that a case k has been infected by case l",

$$p_{k,l} = \frac{\Phi(t_k - t_l)}{\sum_{m=1, m\neq k}^{n} \Phi(t_k - t_m)}$$

where n represents the reported cases and t_k is the time of infection for the case k. Wallinga and Teunis define the *case reproduction number* by

$$R_l = \sum_k p_{k,l}. \qquad \text{[D]}$$

Since R_l only depends on the time of infection t_l, it is actually an estimation of the reproduction number at time $t = t_l$, so the Wallinga and Teunis estimate, R_t^c, of the reproduction number can be expressed as:

$$R_t^c = \sum_k \frac{\Phi(t_k - t)}{\sum_{m=1, m\neq k}^{n} \Phi(t_k - t_m)} \qquad \text{[E]}$$

It remains to establish a relation of R_t^c with the instantaneous reproduction number R_t^{Epi} obtained by the renewal equation with $F \equiv F_1$,

$$R_t^{Epi} = \frac{i_t}{\sum_s i_{t-s}\Phi_s}. \qquad \text{[F]}$$

Grouping in the sum in [E] the cases k such that $t_k = \bar{t}$ and taking into account that there are $i_{\bar{t}}$ such cases, R_t^c can be rewritten as

$$R_t^c = \sum_{\bar{t}} \frac{\Phi(\bar{t} - t) i_{\bar{t}}}{\sum_{s>0} i_{\bar{t}-s}\Phi_s} = \sum_{\bar{t}} \Phi(\bar{t}-t) R^{Epi}{}_{\bar{t}} = \sum_s \Phi(s) R^{Epi}{}_{s+t}. \qquad \text{[G]}$$

We can therefore interpret R_t^c as the forward convolution of the initial estimate R_t^{Epi} with the kernel given by Φ_s. This relation between the instantaneous and case reproduction numbers has also been proven in (9) (equation (10)). On the other hand, as explained above, the EpiEstim estimate R_t^{Epi} can be interpreted (if we neglect the parameters a and b of the Gamma distribution) as the application of Equation [F] to the incidence curve filtered by sliding average on $[t-\tau+1, t]$. In conclusion the Cori et al. and the Wallinga and Teunis methods use the renewal equation $F \equiv F_1$. Note, however, that the Wallinga and Teunis method computes the reproduction number only retrospectively. Indeed, the computation of R_t^c requires the values of $i_{\tilde{t}}$ for any $\tilde{t} > t$ such that $\Phi(\tilde{t} - t) > 0$. This fact was observed in Cori et al.: (in the WT approach), "estimates are right censored, because the estimate of R at time t requires incidence data from times later than t."

C. Implementation details of EpiInvert.

Boundary condition for $[t > t_c]$. The proposed inversion model provides an estimation of R_t up to the the date,t_c, of the last available incidence value. An obvious objection is that if $f_0 < 0$, the functional [9] involves a few future values of R_t and i_t for $t_c \leq t \leq t_c - f_0$. These values are unknown at present time t_c. We use a basic linear regression to extrapolate the values of i_t beyond t_c. To compute the regression line $(i = m_7 \cdot t + n_7)$ we use the last seven values of i_t. In summary, the extension of i_t beyond the observed interval $[0, t_c]$ is defined by

$$i_t = \begin{cases} I_0 e^{at} - I_0 e^{a(t-1)} & if \quad t < 0; \\ m_7 \cdot t + n_7 & if \quad t > t_c. \end{cases} \qquad \text{[H]}$$

The above defined boundary conditions has a very minor influence in the final estimation of R_t in the last days when minimizing [9]. Indeed, the extension of i_t for $t < 0$ is only relevant at the beginning of the epidemic spread. On the other hand, the extension of i_t for $t > t_c$ is only required when the serial interval has negative values. For instance, to evaluate the renewal equation in the energy at the current time t_c using this approach for $F \equiv F_2$ we use the expression

$$i_{t_c} = \sum_{s=0}^{f} i_{t_c-s} R_{t_c-s} \Phi_s + \sum_{s=f_0}^{-1} i_{t_c-s} R_{t_c} \Phi_s,$$

and the extension of i_t for $t > t_c$ is only used in the last term of the above expression where the values of Φ_s are usually very small. Hence the influence of this extension procedure for i_t is also almost negligible. To confirm this claim, we compared, using the shifted log-normal approximation of the serial interval proposed by Ma et al., the estimate of R_{t_c} using the extrapolation based on a linear regression of the last 7 days, with the basic extrapolation given by $i_t = i_{t_c}$ for $t > t_c$. Computing the absolute value of the difference of both estimates for 81 countries we obtain that the quartiles of such distribution of values are $Q_0 = 6.6 \cdot 10^{-6}$, $Q_1 = 1.3 \cdot 10^{-4}$, $Q_2 = 3.1 \cdot 10^{-4}$, $Q_3 = 5.7 \cdot 10^{-4}$ and $Q_4 = 4.9 \cdot 10^{-3}$. We conclude that extrapolation of i_t beyond t_c is a valid strategy to estimate R_t up to $t = t_c$.

Pre-processing the incidence curve. Some countries do not provide data on holidays or weekends and only provide the cumulative total of cases on the next working day. To avoid the strong discontinuity in the data sequence produced by the lack of data, we automatically divide the case numbers of the first non-missing day, between the number of days affected. We do not allow negative numbers in the incidence curve. By default, we replace by zero any negative value of the incidence curve.

Alternate minimization of the energy [9]. To minimize the energy [9], we use an alternate minimization algorithm with respect to R_t and $\mathbf{q}$. Indeed, if $\mathbf{q}$ is fixed, then the optimization of the energy [9] with respect to R_t leads to a linear system of equations that is easily solved. In what follows, we will denote by $R(t, i, \mathbf{q})$ the result of this minimization. On the other hand, when R_t is fixed, the minimization of [9] with respect to $\mathbf{q}$ also leads to a linear system of equations. The constraint [10] is expressed as an additional linear equation,

$$\mu_0 q_0 + \mu_1 q_1 + \mu_2 q_2 + \mu_3 q_3 + \mu_4 q_4 + \mu_5 q_5 + \mu_6 q_6 = \sum_{t=t_c-T+1}^{t_c} i_t, \quad \text{[I]}$$

where $\mu_k = \sum_{t=t_c-T+1}^{k+7t \le t_c} i_{k+7t}$. This linear constraint is easily included in the minimization procedure using, for instance, Lagrange multipliers. So $\mathbf{q}$ is computed as the unique solution of the associated linear system. In what follows we will denote by $\mathbf{q}(R)$ the result of this minimization. Let us denote by R_t^n and $\mathbf{q}^n$ the estimation of R_t and $\mathbf{q}$ in the iteration n of the alternate minimization algorithm. We also denote by $i_t^n = i_t \cdot q_{t\%7}^n$ the filtered incidence curve at iteration n. We initialize $n = 0$, $i^0 \equiv i$, $\mathbf{q}^0 \equiv 1$ and we compute $R_t^0 = R(t, i^0, \mathbf{q}^0)$ as the minimizer of the energy [9] with respect to R_t for $\mathbf{q} \equiv \mathbf{q}^0$.

The whole method is summarized in Algorithm 1, where the maximum number of iterations is fixed to $MaxIter = 100$.

Algorithm 1 Estimation of $\hat{i}$, R, $\mathbf{q}$ from i and Φ.

Initialization: $i^0 \equiv i$, $\mathcal{I}^0 = 1$, $\mathbf{q}^0 \equiv 1$. compute $R_t^0 = R(t, i^0, \mathbf{q}^0)$ minimizing [9] with respect to R_t.
for $n = 1, 2, .., MaxIter$ **do**
  compute $\mathbf{q}^n = \mathbf{q}(R^{n-1})$ minimizing [9] with respect to $\mathbf{q}$.
  compute $i_t^n = q_{t\%7}^n i_t$.
  compute $\mathcal{I}^n$ using [13].
  if $\mathcal{I}^n > \mathcal{I}^{n-1}$ **then**
    stop the iteration
  else
    $\hat{i} \equiv i^n$.
    $\mathbf{q} \equiv \mathbf{q}^n$.
    compute $R_t^n = R(t, i^n, \mathbf{q}^n)$ minimizing [9] with respect to R_t.
    $R = R^n$.
  end
end
return $\hat{i}$, R, $\mathbf{q}$.

Initial boundary condition, for $t = 0$ **.** The evaluation of $F_2(i, R, \Phi, t)$ requires values of R_t and i_t beyond the interval $[0, t_c]$. Given the boundary conditions established, we assume that $R_t = R0$ for $t < 0$ and $R_t = R_{t_c}$ for $t > t_c$. Concerning i_t, for $t < 0$ we will assume, as usual, that at the beginning of the epidemic spread the virus is in free circulation and the cumulative number of infected detected $I_t \equiv \sum_{k=0}^{t} i_k$ follows an exponential growth for $t < 0$, that is $I_t = I_0 e^{at}$, where a represents the initial exponential growth rate of I_t at the beginning of the infection spread. We now naturally estimate a by

$$a = median(\{log\left(\frac{I_{t+1}}{I_t}\right) \ : \ t = 0, .., 14\}). \qquad \text{[J]}$$

If we assume that $I_t = I_0 e^{at}$ follows initially an exponential growth and that $R_t = R0$ is initially constant, then we can compute $R0$ from the exponential growth a and the renewal equation taking into account that

$$i_0 = I_0(1 - e^{-a}) = I_0 R0 \sum_{k=f_0}^{f} (e^{-ka} - e^{-(k+1)a}) \Phi_k. \qquad \text{[K]}$$

Hence, we can compute an approximation of $R0$ as

$$R0 = \frac{1 - e^{-a}}{\sum_{k=f_0}^{f} (e^{-ka} - e^{-(k+1)a}) \Phi_k}. \qquad \text{[L]}$$

This estimation of $R0$ is a discrete version of the formula given in (9) where the incidence curve is assumed to follow an exponential growth. Note that this estimation strongly depends on the serial interval used. For instance, if we assume that $a = 0.250737$ (the exponential growth rate obtained in (25) when the coronavirus is in free circulation), we obtain that $R0 = 2.700635$ for the Nishiura et al. serial interval, $R0 = 3.084528$ for the Ma et al. serial interval and $R0 = 1.839132$ for the Du et al. serial interval.

D. Experiments using simulated data. We describe here in more detail a simulator $R_t \to i_t$. The simulator starts from a realistic scenario on the evolution of R_t depending on parameters fixed by the user. Then, using a choice of serial intervals and a realistic weekly bias borrowed from real examples, the simulator samples the incidence curve i_t as the realization of Poisson variable. The simulated "ground truths" for R_t will be denoted by R_t^{GT}. They are similar to those proposed in Gostic et al. (11). R_t^{GT} goes from a user selected initial value $R_0 > 1$ to an intermediate value $R_i < 1$, and finally goes back to 1. This hypothesis for an evolution corresponds to a typical lockdown scenario where initially the number of cases grows exponentially, then a lockdown is implemented during a time that we denote by t', and finally social-distancing measures relax but try to stabilize R_t around 1. The parameters defining the simulated ground truth R_t^{GT} therefore are R_0, R_i, t' and s, which determines the slope of the transitions between R_0 and R_i and between R_i and 1. the larger s, the sharper the transitions. To define R_t^{GT} we use the following function:

$$R_{y_0,y_1,s,t'}(t) = y_0 + \frac{y_1 - y_0}{2}\left(1 + \frac{2}{\pi}\arctan\left(s\frac{\pi(t-t')}{|y_0 - y_1|}\right)\right) \quad [M]$$

where y_0, y_1, s and t' are the function parameters. This function satisfies : $\lim_{t\to-\infty} R_{y_0,y_1,s,t'}(t) = y_0$, $\lim_{t\to\infty} R_{y_0,y_1,s,t'}(t) = y_1$. The maximum of the absolute value of its derivative is equal to s and is attained at $t = t'$. Next, we define R_t^{GT} by

$$R_t^{GT} = \begin{cases} R_{R_0,R_i,s,0}(t) & \text{if } R_{R_0,R_i,s,0}(t) \geq R_{R_i,1,s',t'}(t); \\ R_{R_i,1,s',t'}(t) & \text{if } R_{R_0,R_i,s,0}(t) < R_{R_i,1,s',t'}(t). \end{cases} \quad [N]$$

To reduce the number of parameters we assume that $s' = s/5$, reflecting the fact that the relaxation of social distancing measures is more progressive than a lockdown.

The ground truth of the incidence curve, that we denote by i_t^{GT}, is computed from the renewal equation using R_t^{GT} as reproduction number and a user selected serial interval among three proposed (Du, Ma, Nishiura). We take an initial value for i_0^{GT} and iteratively compute i_t^{GT} from $\{i_{t'}^{GT} : t' < t\}$ using the renewal equation and the boundary conditions explained above. Then, we improve the accuracy of the estimation of i_t^{GT} by applying a Newton method until convergence. Indeed, given R_t^{GT}, the renewal equation is a fixed point equation in i_t. Since the ground truth of the incidence curve is defined up to the multiplication by a constant factor, rather than fixing the initial number of cases, we add a more intuitive parameter i_{max} which allows the user to fix the maximum value of the incidence curve in the whole period. This value impacts the noise inherent to a Poisson process: the smaller i_{max}, the larger the stochastic oscillation of i_t. We then simulate the observed incidence curve i_t assuming that $i_t = Pois(i_t^{GT} q'_{t\%7})$, that is, i_t follows a Poisson distribution of mean $i_t^{GT} q'_{t\%7}$ where $\mathbf{q}' = (q'_0, .., q'_6)$ is the vector with the weekly bias correction factors. The weekly bias proposes several real bias correction factors options $\mathbf{q} = (q_0, .., q_6)$, borrowed from the EpinInvert analysis of the incidence curve of 19 countries. To obtain $\mathbf{q}'$ from $\mathbf{q}$, we simply invert the weekly bias correction coefficients by setting $q'_k = 1/(q_k \lambda)$, where λ is a normalization factor preserving the cumulative number of cases in the period of analysis. More precisely, λ is derived from the equation

$$\sum_{t=t_c-T+1}^{t_c} \frac{i_t^{GT}}{q_{t\%7}\lambda} = \sum_{t=t_c-T+1}^{t_c} i_{t'}^{GT}. \quad [O]$$

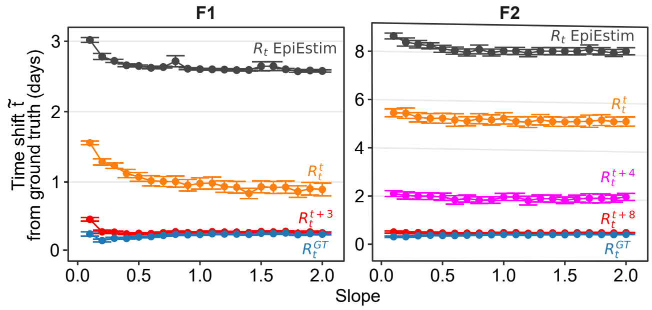

Fig. S1. Time delay between the reproduction number ground truth R_t^{GT} and its various estimates when computing the time shift that minimizes the RMSE between both curves. The horizontal axis is the slope of R_t at lockdown time ($t = 0$). Using the simulator, these estimates confirm the 8.5 days delay of the EpiEstim estimate with respect to the ground truth for F2, and a 3 days delay for the F1 form of the equation. The EpiInvert delay is also important on the first evaluation day, but decreases as days pass by.

From a sample of the stochastic simulation of i_t, the demo finally computes R_t using EpiInvert and EpiEstim.

In Fig. S1 we show, as a function of the slope in the lockdown transition, the distributions of the optimal time delay between R_t^{GT} and its various estimates. The optimal time shift between an R_t curve and R_t^{GT} is the one that minimizes their RMSE. We observe that the time shift is slightly larger when the slope of the transition is small.

E. Case studies: USA, France, Japan, Peru and South Africa . The country data about the registered daily infected are taken from https://ourworldindata.org. In the particular cases of France, Spain and Germany we use the official data reported by the countries. We shall use the incidence data up to Friday, July 23, 2021 (so the last weekly bias correction factor q_6 corresponds to a Friday). For the US states, the data are obtained from the New York Times report [§].

Table S2 contains a summary of the values computed for each experiment. To compute the EpiEstim estimation R_t^{Epi}, we used $\tau = 7$, that is, we assumed that R_t is constant in $[t-7, t]$. As proposed by Cori et al. in (5) we used $a = 1$ and $b = 5$ for the parameters of the $\Gamma(a, b)$ prior distribution for R_t. Yet, as explained above, these values could be neglected in the EpiEstim estimation, given the magnitude of the incidence data in these regions. The values of the bias correction coefficients q_k obtained for $F \equiv F_1$ and $F \equiv F_2$ are quite similar. So it seems that the choice of the renewal equation has no significant influence on the estimation of the bias correction coefficients.

In Fig. S2 we show the charts obtained for the USA with $F \equiv F_1$ and $F \equiv F_2$. The USA shows a clear weekly periodic bias. The correction of this bias works quite well, as the RMSE reduction reaches $\mathcal{I} = 0.409$ for $F \equiv F_1$ and $\mathcal{I} = 0.381$ for $F \equiv F_2$. The oscillation of the incidence curve is strongly reduced, passing from $\mathcal{V}(i) = 0.542$ to $\mathcal{V}(\hat{i}) = 0.267$. The agreement with EpiEstim is also excellent as $\mathcal{S}(\tilde{t}) = 0.053$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.048$ for $F \equiv F_2$. The daily bias correction factors are similar for $F \equiv F_1$ and $F \equiv F_2$. On Sundays the number of cases is strongly underestimated ($q_1 = 3.205$ for $F \equiv F_2$) and overestimated on Fridays ($q_6 = 0.569$ for

[§] https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv

	USA	USA	France	France	Japan	Japan	Peru	Peru	S.Africa	S.Africa
F	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
$\tilde{t}$	2.38	7.53	3.42	8.73	2.92	9.61	2.50	7.00	2.64	8.39
$\mathcal{S}(\tilde{t})$	0.053	0.048	0.052	0.065	0.014	0.022	0.068	0.070	0.027	0.032
$\mathcal{I}$	0.409	0.381	0.425	0.456	0.265	0.274	0.773	0.770	0.347	0.345
q_0	1.916	1.981	0.905	0.893	0.885	0.880	1.267	1.248	0.836	0.838
q_1	3.205	3.382	1.022	0.996	1.128	1.124	0.875	0.867	1.115	1.118
q_2	0.848	0.879	3.836	3.682	1.624	1.618	0.666	0.678	1.539	1.539
q_3	1.014	1.033	0.858	0.844	1.051	1.049	0.791	0.803	1.300	1.298
q_4	0.985	0.970	0.825	0.827	0.851	0.851	1.184	1.199	0.864	0.864
q_5	1.093	1.048	0.921	0.942	0.846	0.849	1.346	1.346	0.872	0.871
q_6	0.569	0.541	0.933	0.974	0.960	0.968	1.330	1.308	0.853	0.853

Table S2. Numerical results obtained by EpiInvert for the USA, France, Japan, Peru and South Africa using the data up to July 23, 2021 and the renewal equations $F = F_1$ and $F = F_2$.

$F \equiv F_2$).

In Fig. S3 we show the charts obtained for France with $F \equiv F_1$ and $F \equiv F_2$. France also displays a clear weekly periodic bias: on Mondays the number of cases is strongly underestimated ($q_2 = 3.682$ for $F \equiv F_2$). The correction of the periodic bias works well, as $\mathcal{I} = 0.425$ for $F \equiv F_1$ and $\mathcal{I} = 0.456$ for $F \equiv F_2$. The oscillation of the incidence curve is therefore reduced, passing from $\mathcal{V}(i) = 0.346$ to $\mathcal{V}(\hat{i}) = 0.087$. The agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.052$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.065$ for $F \equiv F_2$.

In Fig. S4 we show the charts obtained for Japan with $F \equiv F_1$ and $F \equiv F_2$. In Japan, the weekly bias correction works very well and yields $\mathcal{I} = 0.265$ for $F \equiv F_1$ and $\mathcal{I} = 0.274$ for $F \equiv F_2$. The oscillation of the incidence curve reduces from $\mathcal{V}(i) = 0.189$ to $\mathcal{V}(\hat{i}) = 0.069$. The agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.014$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.022$ for $F \equiv F_2$. Observe how the incidence curve is underestimated on Mondays ($q_2 = 1.618$).

In Fig. S5 we show the charts obtained for Peru with $F \equiv F_1$ and $F \equiv F_2$. Although in general countries present a clear weekly periodic pattern in the incidence curve this is not the case for Peru: we obtain $\mathcal{I} = 0.773$ for $F \equiv F_1$ and $\mathcal{I} = 0.770$ for $F \equiv F_2$. The oscillation of the incidence curve is not reduced, going from $\mathcal{V}(i) = 0.355$ to $\mathcal{V}(\hat{i}) = 0.369$. The agreement with EpiEstim is good with $\mathcal{S}(\tilde{t}) = 0.068$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.070$ for $F \equiv F_2$.

In Fig. S6 we show the charts obtained for South Africa with $F \equiv F_1$ and $F \equiv F_2$. The correction of the periodic bias works well, as $\mathcal{I} = 0.347$ for $F \equiv F_1$ and $\mathcal{I} = 0.345$ for $F \equiv F_2$. The oscillation of the incidence curve is reduced, passing from $\mathcal{V}(i) = 0.191$ to $\mathcal{V}(\hat{i}) = 0.087$. On Mondays the number of cases is underestimated ($q_2 = 1.539$ for $F \equiv F_2$). The agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.027$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.032$ for $F \equiv F_2$.

The optimal shift $\tilde{t}$ between R_t is R_t^{Epi} obtained for the different countries fits in the range obtained by a joint analysis of the 55 countries. Indeed, for $F \equiv F_1$ $\tilde{t}$ ranges from 2.38 to 3.42 and for $F \equiv F_2$ $\tilde{t}$ ranges from 7.00 to 9.61.

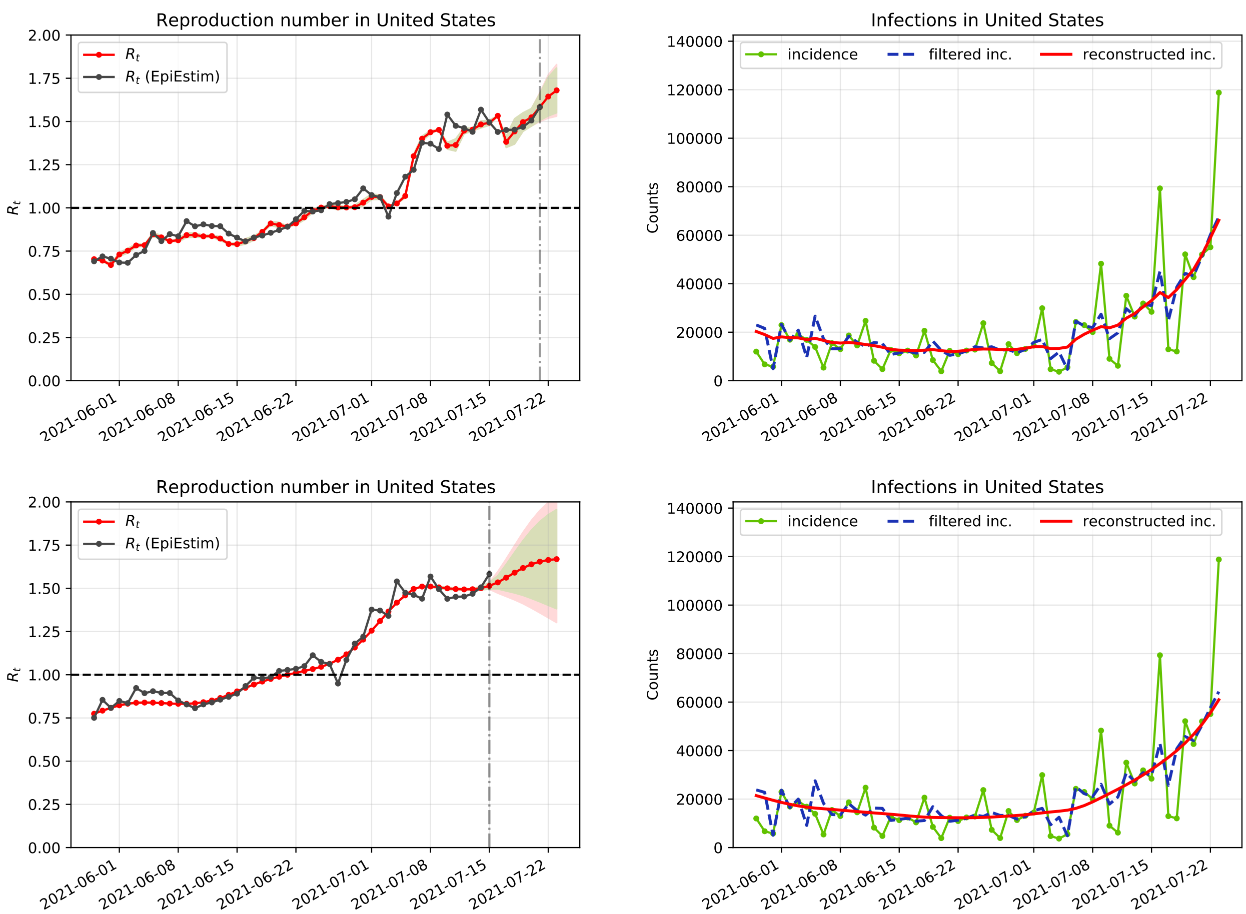

Fig. S2. Results obtained for the USA up to July 23, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.

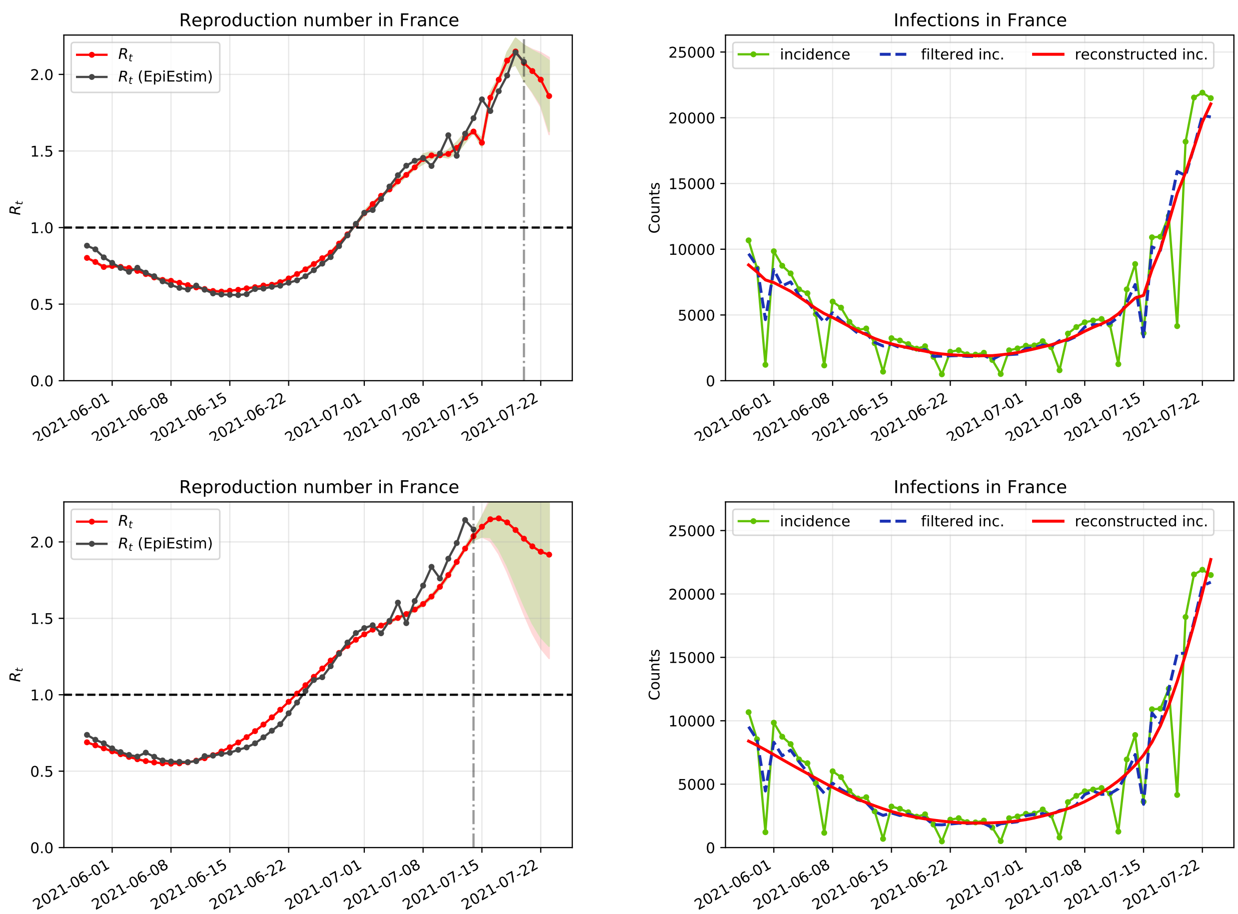

Fig. S3. Results obtained for France up to July 23, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.

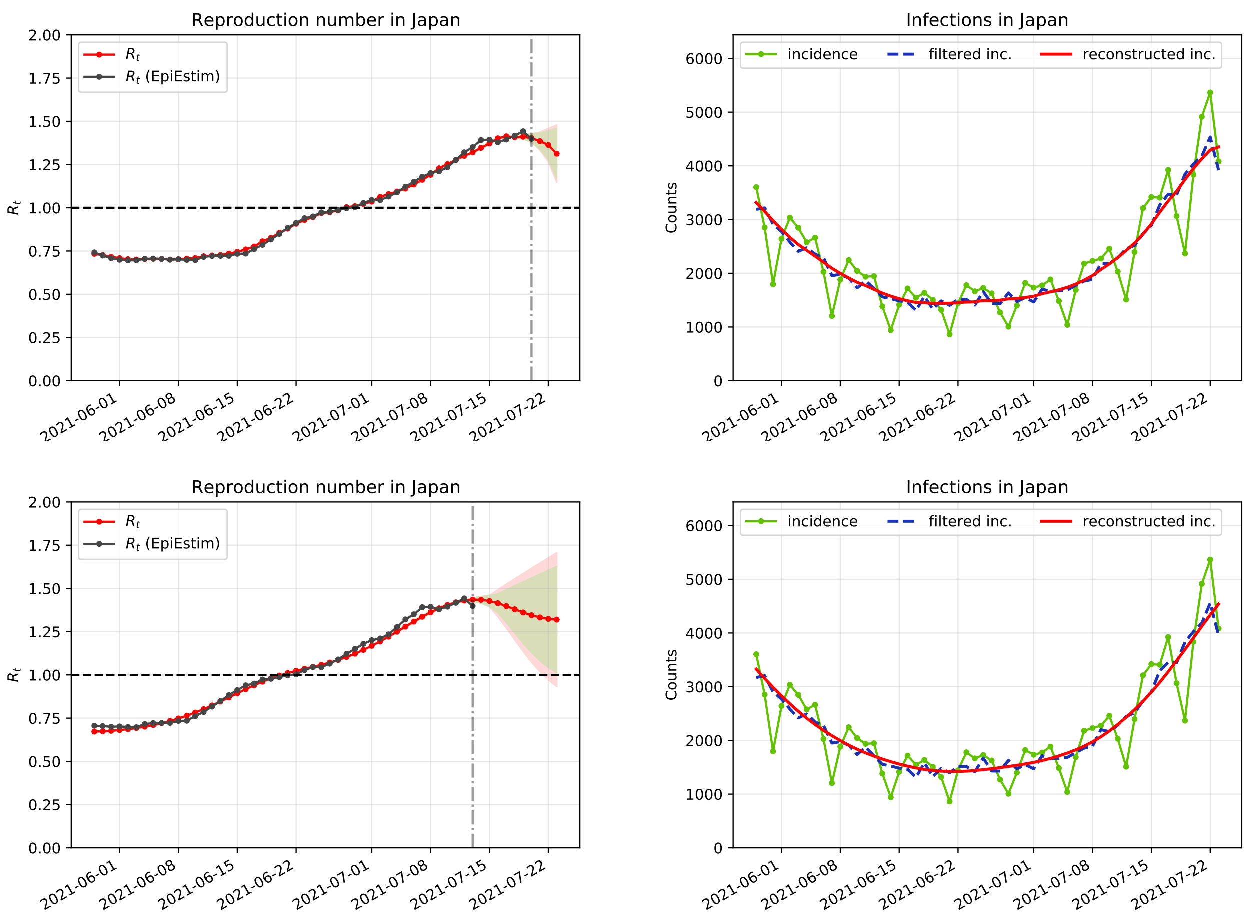

Fig. S4. Results obtained for Japan up to July 23, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.

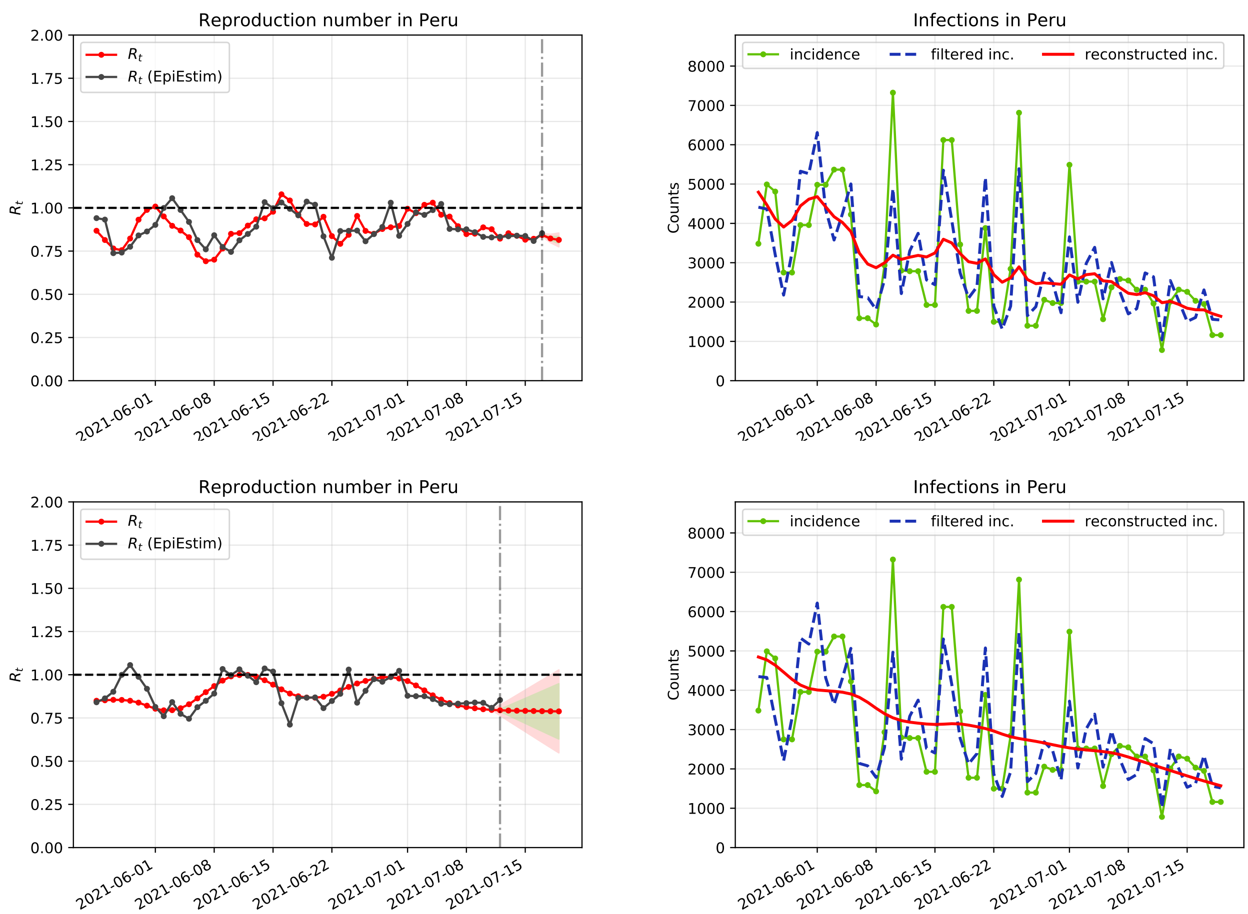

Fig. S5. Results obtained for Peru up to July 23, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.

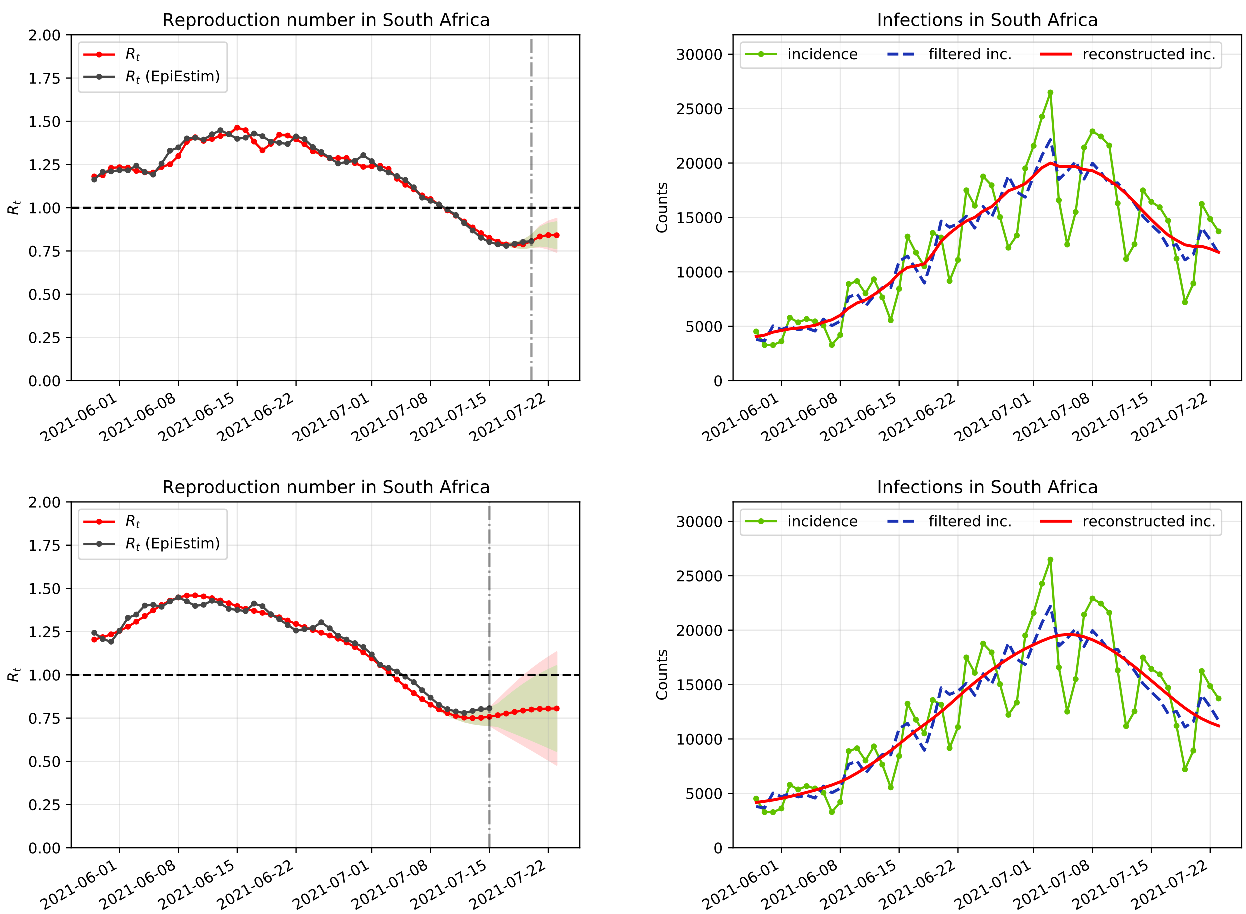

Fig. S6. Results obtained for South Africa up to July 23, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.