Computing the daily reproduction number of COVID-19 by inverting the renewal equation

Luis Alvarez[1*], Miguel Colom[2], Jean-David Morel[3] and Jean-Michel Morel[2]

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

[*]Correspondence to: lalvarez@ulpgc.es

Abstract

The COVID-19 pandemic has undergone rapid changes in its infection rates driven by governmental measures. The reproduction number $R(t)$ indicates the number of cases generated by an infected person at time t and is a key indicator of the spread of an epidemic. The EpiEstim method is the most widely accepted method for estimating $R(t)$. Here, we propose a new method, EpiInvert, that estimates $R(t)$ 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. EpiInvert shows good agreement with EpiEstim, but provides estimates of $R(t)$ up to 9 days in advance. 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/ern`.

One-Sentence Summary: We propose a method computing the COVID-19 reproduction number up to 9 days closer to present than current methods.

The reproduction number R_t is a key epidemiological parameter evaluating transmission rate 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],

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

where t_c represents the current time (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.

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 a simplified version of the renewal

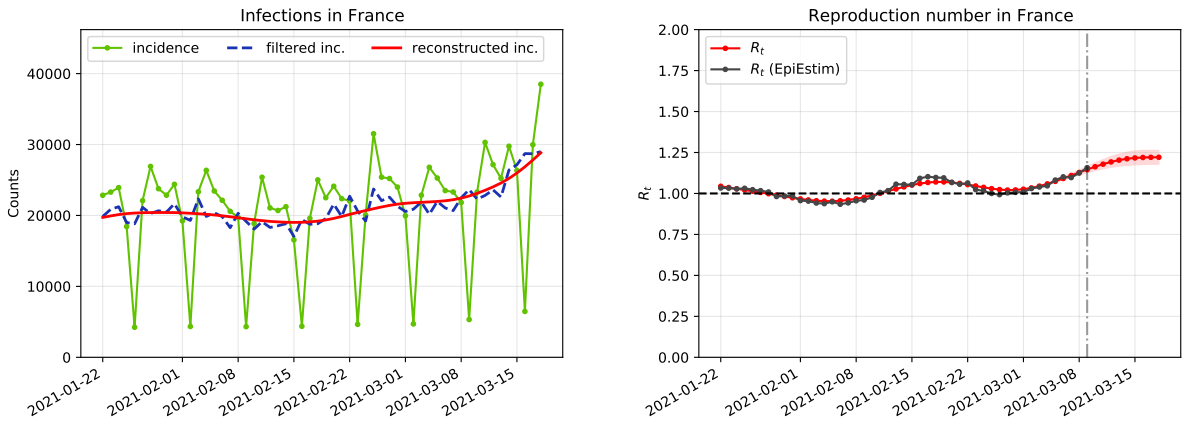

Figure 1: Illustration of the EpiInvert method on the France incidence curve of new cases. On the left: in green, the raw oscillating curve of incident cases up to March 26, 2021. In blue, the incidence curve after correction of the "week-end bias". In red, the incidence curve simulated from R_t after 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 [9]. Estimating its value every day guides the health policy of each country. Having R_t larger than 1, as it is the case for France on March 26, 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, predicts R_t nearly nine days closer to the present than EpiEstim.

equation proposed in Cori et al in [5].

$$i_t = R_t \sum_{f_0}^{f} i_{t-s} \Phi_s, \tag{2}$$

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

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

This method, implemented by the EpiEstim software, is highly recommended in a very recent review [10] signed by representatives from ten different epidemiological labs from several continents. A detailed description of EpiEstim can be found in the supplementary material. Equation (2) is the standard method, 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 [11, 12, 5, 13, 14]).

Comparing Equations (2) and (1) shows that the second equation is derived from the first by assuming R_t constant on the serial interval $[t-f, t-f_0]$. 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

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

where μ is a central value of the probability distribution of the serial interval Φ that could be the median or the mean. In the Ma et al. [15] 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 that EpiEstim estimates R_t with an average delay of more than 5 days.

In practice, the delay is even longer, due to the way the sliding average of the incidence is calculated. 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 the 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 raises three challenges:

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

2. A simplification of the renewal equation (1) leads the standard method to estimate R_t with a backward shift of more than 5 days.

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

These cumulative backward shifts cause a time delay of more than 8.5 days. In other terms, the value of R_t computed at time t refers approximately to R_{t-9}.

Here, we address these three issues by proposing a method that inverts Equation (1) without simplifying it. The result of EpiInvert, the inversion method developed here, is illustrated in Figure 1 (right), where the EpiEstim result (in black) is superposed with the estimate (in red) of R_t by EpiInvert. After registering both, the black EpiEstim curve stops nine days before EpiInvert, the red curve (our estimate). We found, using the incidence curve of 70 countries, that the optimal shift between the EpiEstim and EpiInvert R_t estimates is about 8.3 ± 0.5 days and that the RMSE approximation error between both estimates is just about $3.6\% \pm 1.9\%$.

Indeed, the general integral equation (1) is a functional equation in R. Integral equations have been previously used to estimate R_t: in [16], 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 [17] variational approach involving a regularization term. This method is widely used to solve integral equations and convolutional equations [18]. 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 (on the left, in red).

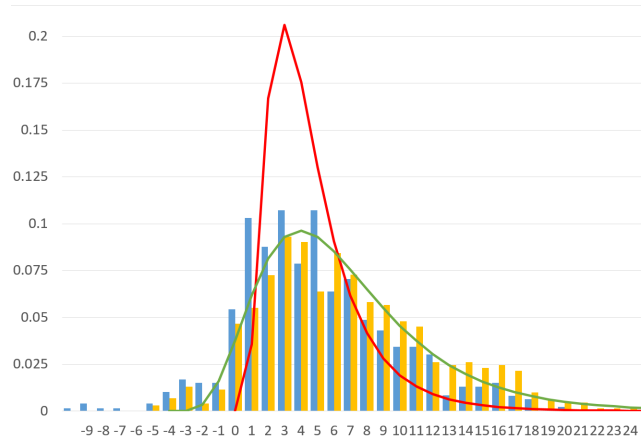

Figure 2: Serial intervals used in our experiments: the discrete one proposed by Du et al. in [19] (solid bars in blue), the serial interval proposed by Ma et al. [15] (solid bars in yellow) and its shifted log-normal approximation (in green), finally a log-normal approximation of the serial interval proposed by Nishiura et al. in [20] (in red).

Available serial interval functions for the COVID-19

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 [19] define it as "the time duration between a primary case (infector) developing symptoms and secondary case (infectee) developing symptoms."

Du et al. in [19] 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 [19] 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 [20] using 28 cases, which is approximated by a log-normal distribution, and a serial interval obtained by Ma et al. in [15] 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. [19] and Ma et al. [15][1]. 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 $[f_0, f]$.

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

Computing R_t by a variational method

We consider two versions of the general renewal equation (1) given by

$$i_t = F(R, i, \Phi, t) \quad \text{for } t = 0, .., t_c, \tag{5}$$

where

$$F_1 \equiv R_t \sum_{s=f_0}^{f} i_{t-s}\Phi_s; \quad F_2 \equiv \sum_{s=f_0}^{f} i_{t-s}R_{t-s}\Phi_s. \tag{6}$$

F_2 corresponds to the general renewal equation and F_1 to the simplifed Cori et al. [5] version. The very same formula can also be derived for the classic Wallinga Teunis method [4], as shown in the supplementary material. 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

$$E(\{R_t\};\mathbf{q}) = \sum_{t=0}^{t_c}\left(\frac{q_{t\%7}i_t - F(\{q_{t\%7}i_t\}, R, \Phi, t)}{p_{50}(i)}\right)^2 + w\sum_{t=1}^{t_c}(R_t - R_{t-1})^2, \tag{7}$$

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 current day.

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

$$\sum_{t=t_c-T+1}^{t_c} i_t = \sum_{t=t_c-T+1}^{t_c} \hat{i}_t = \sum_{t=t_c-T+1}^{t_c} q_{t\%7}i_t, \tag{8}$$

to the minimization problem (7).

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 (7), $p_{50}(i)$ is the 50th percentile (the median) of $\{i_t\}_{t=t_c-T+1,..,t_c}$ used to normalize the energy with respect to the size of i_t. The first term of E is a data fidelity term which forces the renewal equation (5) to be satisfied as much as possible. The second term is a classic Tikhonov-Arsenin regularizer of R_t.

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 $R_0 = R0$. (In the supplementary material we present a basic estimation of $R0$ from the initial exponential growth rate of the epidemic obtained as in [21]), therefore, to solve Equation (7), we add the boundary condition $R_0 = R0$. The proposed inversion model provides an estimation of R_t up to the current day t_c. Yet if $f_0 < 0$, the functional (7) 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 supplementary material, 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 (7).

Experimental setup

All of the experiments described here can be reproduced with the online interface available at `www.ipol.im/ern`. 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 date updated to the current date. We detail our daily sources in the supplementary material.

An empirical confidence interval for R_t and a measure for the oscillation of i_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:

1. Compute $\{R^k(t)\}_{t\in[0,t_c-k]}$ by minimizing (7) for $k = 1, 2, 3$, using the data sequence up to $t_c - k$.

2. Compute for each $t \in [0, t_c]$ a confidence bound of R_t with respect to its value $R^1(t)$, $R^2(t)$ and $R^3(t)$ in the three preceding days given by

$$\sigma(t) = \sqrt{\frac{\sum_{k=1}^{3}(R_t - R^k(t))^2}{3}}, \tag{9}$$

where $R^k(t)$ in $(t_c - k, t_c]$ are obtained by linear extrapolation.

We then define a conservative empirical confidence interval as $[R_t - 2\cdot\sigma(t), R_t + 2\cdot\sigma(t)]$. This interval is displayed for each t in the online algorithm `www.ipol.im/ern` and has the aspect of a fattened curve above and below R_t.

We also define the oscillation of i_t by

$$\mathcal{V}(i) \equiv \frac{\|i'\|_{L^1[t_c-T,t_c]}}{\|i\|_{L^1[t_c-T,t_c]}} \approx \frac{\sum_{t=t_c-T+1}^{t_c}|i_t - i_{t-1}|}{\sum_{t=t_c-T+1}^{t_c} i_t} \tag{10}$$

Efficiency measure of the weekly bias correction

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

$$\mathcal{I} = \sqrt{\frac{\sum_{t=t_c-T+1}^{t_c}\left(\hat{i}_t - F(\hat{i}, R, \Phi, t)\right)^2}{\sum_{t=t_c-T+1}^{t_c}\left(i_t - F(i, R1, \Phi, t)\right)^2}}. \tag{11}$$

$\mathcal{I}$ represents the reduction, after the correction of the weekly periodic bias, of the average distance between the corrected $\hat{i}_t = i_t q_{t\%7}$ and $F(\hat{i}, R, \Phi, t)$ for the optimal values of R_t and $\mathbf{q}$. Here $R1$ denotes the result of the same optimization 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^i the EpiEstim estimation of the reproduction number by Cori et al. in [5], detailed in the supplementary material. As we have argued above, we expect a significant temporal shift between R_t and R_t^i, 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/ern`. 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^i and R_t yielding the best RMSE between both estimates:

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

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 $F \equiv F_2$;
- w: regularization weight, with default values $w = 5.5$ for $F \equiv F_1$ and $w = 5$ for $F \equiv F_2$;
- Correction of the weekly periodic bias (option by default)

Note that the regularization weight w is the only numerical mandatory parameter.

Summary of the output displayed at `www.ipol.im/ern`

First we draw two charts. In the first one we draw R_t and R_t^i shifted back $\tilde{t}$ days where $\tilde{t}$ is defined in (12). R_t is surrounded by a shaded area that represents the above defined empirical confidence interval. 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. $R_{t_c}^i$: last available value of the EpiEstim estimate R_t^i.
3. $\tilde{t}$: optimal shift (in days) between R and R^i defined in (12).
4. $\mathcal{S}(\tilde{t})$: RMSE between R and R^i shifted back $\tilde{t}$ days (defined in (12)).
5. $\mathcal{V}(i)$: variability of the original incidence curve, i_t, given by (10).
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 i_t and $F(i, R, \Phi, t)$ before and after the correction of the weekly periodic bias (defined in (11)).
8. $\mathbf{q} = (q_0, .., q_6)$: the correction coefficients of the weekly periodic bias (q_6 corresponds to the current time t_c).

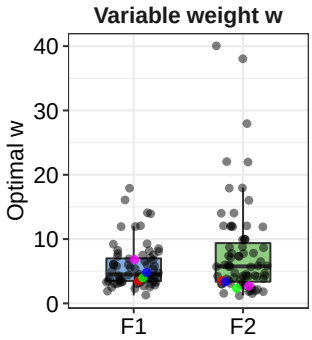

Figure 3: Distribution of w for F_1 and F_2 when the regularization weight w and the delay $\tilde{t}$ are optimized independently for each country to minimize the average error $\mathcal{S}(\tilde{t})$ between the EpiEstim and the EpinInvert methods on a time lapse of 56 days. France in blue, Japan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

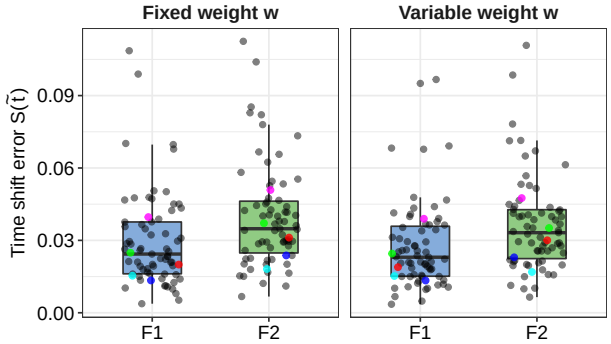

Figure 4: Average error $\mathcal{S}(\tilde{t})$ over 56 consecutive days of the error between the EpiEstim and the EpiInvert estimates of R_t for each country. France in blue, Japan in green, Peru in black, South Africa in magenta, USA in red.

Results

To estimate a reference value for the regularization parameter w we used the incidence data up to Saturday March 18, 2021 for the 70 countries showing the larger number of cases. For each country, we optimized the RMSE $\mathcal{S}(\tilde{t})$ between R and R^i shifted back $\tilde{t}$ days (defined in (12)). 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 country to minimize the average error over 56 days between the EpiEstim and the EpinInvert methods. The mean w was $5.5\pm2.4\%$ for F_1 and $5.9\pm2.9\%$ for F_2 which indicated that a common value of w could be fixed for all countries. Here and in all figures to follow, each plot represents a country.

In Fig 4, we show, for the versions F_1 and F_2 of the renewal 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 overall average error is $2.9\% \pm 1.7\%$ for F_1 and $3.6\% \pm 1.9\%$ for F_2. 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 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 its median for all countries namely $w = 5$ for $F \equiv F_1$, and $w = 5.5$ for $F \equiv F_2$. 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

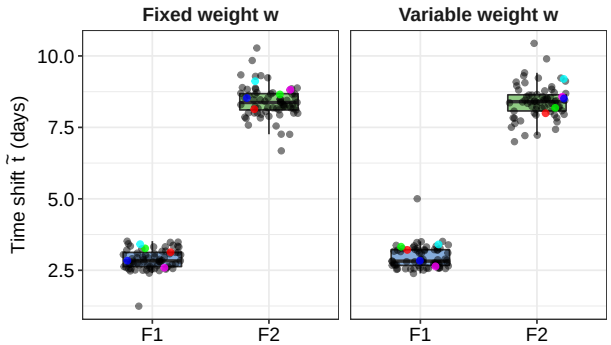

Figure 5: Optimal time shift $\tilde{t}$ obtained by minimizing the mean error $\tilde{S}(t)$ over 56 days between the EpiEstim and the EpiInvert estimates of R_t for each country. The time shift is as predicted by our theoretical analysis close to 3 days for F_1 and slightly above 8 days for F_2. On the left w is fixed and on the right it is the optimal weight per country.

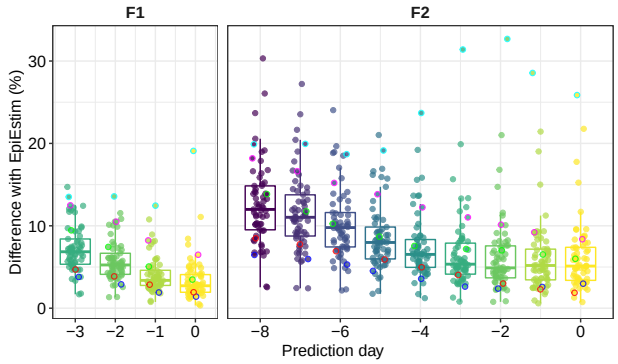

Figure 6: Relative error between the EpiInvert and EpiEstim estimations, depending on the anticipation day. EpiInvert anticipates the value of R_t by 0 to 3 days in the F_1 formulation and by 0 to 8 days in the F_2 formulation. Each plot represents one country. France in blue, Japan in green, Peru in black, South Africa in magenta, USA in red.

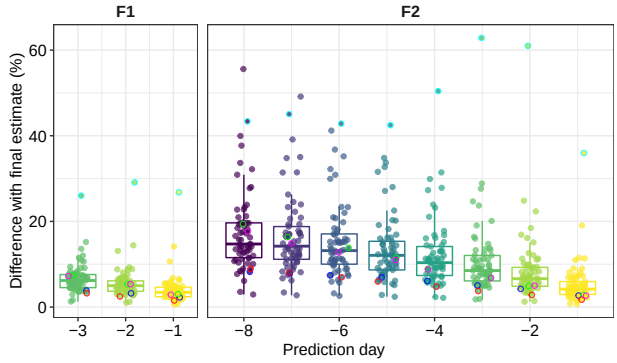

Figure 7: Internal relative error between the EpiInvert estimations depending on the anticipation day. Each plot represents one country. The mean difference for each prediction day is marked by a horizontal bar. The standard deviation of the relative error is half the height of each box. France in blue, Japan in green, Peru in black, South Africa in magenta, USA in red.

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 $\tilde{t} = 2.96 \pm 0.42$ for variable w and $\tilde{2}.86 \pm 0.43$ for F_1 with fixed w, and similarly for F_2: $\tilde{t} = 8.33 \pm 0.55$ and $\tilde{t} = 8.38 \pm 0.52$.

As shown in Fig. 4, the agreement between R_t and R_t^i shifted back by the optimal delay $\tilde{t}$ is overwhelming. Indeed, for $F \equiv F_1$, we obtain that the median of their relative difference $\mathcal{S}(\tilde{t})$ is just 2.43%. For $F \equiv F_2$, it is 3.3%. The median of the shift $\tilde{t}$ is given by 2.89 (for $F \equiv F_1$) and 8.33 (for $F \equiv F_2$). 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 about 9 days for $F \equiv F_2$. The difference between the predicted time delay and the observed one therefore is 0.5 days. This is easily explained by the regularization term in Equation (7), 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 reliability of the EpiInvert estimate, which as we saw anticipates by 8 days the EpiEstim estimate of R_t. Let us denote by $R_t(t-k)$ the EpinInvert estimate at time $t-k$ using the incidence curve up to the date t. According to the estimated shift between R_t and R_t^i, for $F \equiv F_1$, $R_t(t-3)$ should be similar to R_t^i. The first estimation of $R_t(t-3)$ is obtained 3 days before when we compute $R_{t-3}(t-3)$.

In Fig. 6 we show a box plot of the relative error between the EpiInvert and EpiEstim estimations, depending on the anticipation day. EpiInvert anticipates the value of R_t by 0 to 3 days in the F_1 formulation and by 0 to 8 days in the F_2 formulation. Each dot represents one country. On the left of Fig. 6 we compare for $k = 3, 2, 1, 0$ the means of the relative errors $|R_k(t-3) - R^i(t)|$ for $t \in [t_c - T + 1, t_c]$ ($T = 56$) and for the 70 countries selected as the ones with higher incidence. The difference $|R_k(t-3) - R^i(t)|$ cannot be considered as an approximation error. A good agreement is expected between $R_k(t-3)$ and $R^i(t)$, but the method underlying both estimations is different. Our goal was not to approximate the EspiEstim method, but to solve the renewal equation in a more exact formulation. Nevertheless, it was important to verify that EpiInvert finds very similar values to EpiEstim on their common interval of definition. These values are predicted by EpiInvert up to 9 days in advance. As expected, this relative error grows for the early predictions. Nevertheless, for the renewal equation F_2, one observes a plateau of this error between days -4 and 0 with a mean difference of about 5.5%. Even with an 8 days anticipation, the average relative error on R_t stays below 12%.

Finally, we are obviously interested in the internal coherence of the EpiInvert predictions. Indeed, contrarily to EpiEstim, the estimate $R_{t+k}(t)$ or R at time t evolves for $k \geq 0$ and becomes more accurate at later dates. Fig. 7 gives a box plot of the internal relative error between the EpiInvert estimations depending on the anticipation day. Since the estimate of EpiEstim at each day evolves with the knowledge of the incidence in later days, we can see how this estimate evolves as time passes. Each plot represents one country. We see that the relative error on R at a given date t goes down almost linearly from 14% in an early prediction (8 days ahead) to 4.4% (1 day ahead). The robustness of the prediction is positively affected by incidence numbers.

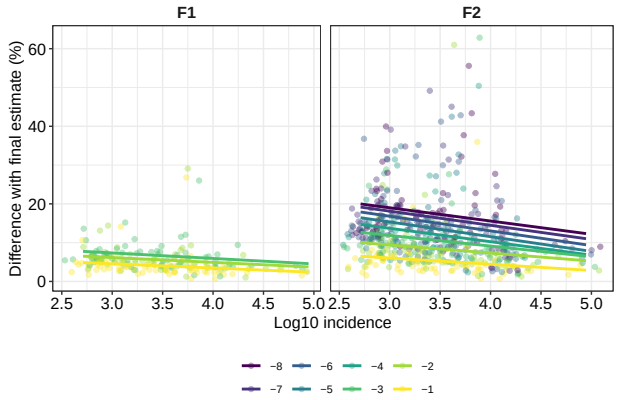

Figure 8: Linear regression of the internal relative error between the EpiInvert estimation as a function of the mean incidence. The regression lines are clearly decreasing, which means that a higher incidence favor a better estimate of R_t.

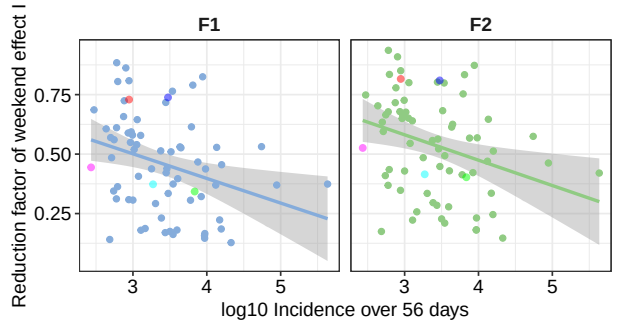

Figure 9: Reduction factor $\mathcal{I}$ (see (11)) obtained by applying correcting coefficients to reduce the "weed end effect". This reduction decreases from about 0.5 to less than 0.25. The plots are ordered in increasing order of average incidence.

Fig. 8 indeed shows, for each anticipation day $-k = -1, \cdots -8$, the linear regression of the internal relative error between the EpiInvert estimations at days 0 and $-k$, viewed as a function of the mean incidence of the country. These eight 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. 9 shows a regression plot of the reduction factor of the oscillation of i_t obtained by applying correcting coefficients to reduce the "week-end effect". This reduction decreases from about 0.5 to less than 0.25, 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 and 7 are presented in interactive mode with tooltip detailed statistics on each country.

CONCLUSION

The reproduction number R_t can be estimated by solving a renewal equation linking R_t, i_t and Φ_s. We considered the formulations of the renewal equation providing the named instantaneous reproduction number ($F \equiv F_1$) and the named effective reproduction number ($F \equiv F_2$). The daily incidence data i_t recorded by health administrations show a strong non-Poisson 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 an empirical confidence interval. In contrast to former methods, EpiInvert can use serial intervals with distributions containing negative days (as it is the case for the COVID-19). Thus, it avoids an artificial truncation of the distribution. EpiInvert shows excellent agreement with EpiEstim. Its main improvement is to anticipate by several days the estimate of R_t : 3 days for the F_1 formulation of the renewal equation, and more than 8 days for its F_2 formulation. This last fact is extremely relevant, given that the control of social distancing policies requires a timely estimate of R_t.

References

[1] P. Rodpothong, P. Auewarakul, *World Journal of Virology* **1**, 131 (2012).

[2] X. He, *et al.*, *Nature medicine* **26**, 672 (2020).

[3] P. Ashcroft, *et al.*, *arXiv preprint arXiv:2007.06602* (2020).

[4] J. Wallinga, P. Teunis, *American Journal of epidemiology* **160**, 509 (2004).

[5] A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, *American journal of epidemiology* **178**, 1505 (2013).

[6] A. J. Lotka, *Science* **26**, 21 (1907).

[7] H. Nishiura, *Theoretical Biology and Medical Modelling* **4**, 20 (2007).

[8] H. Nishiura, G. Chowell, *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends* (Springer Netherlands, Dordrecht, 2009), pp. 103–121.

[9] K. Gostic, *et al.*, *MedRxiv* (2020).

[10] K. M. Gostic, *et al.*, *PLoS computational biology* **16**, e1008409 (2020).

[11] R. Thompson, *et al.*, *Epidemics* **29**, 100356 (2019).

[12] Q.-H. Liu, *et al.*, *Proceedings of the National Academy of Sciences* **115**, 12680 (2018).

[13] T. Obadia, R. Haneef, P.-Y. Boëlle, *BMC medical informatics and decision making* **12**, 147 (2012).

[14] T. Z. Boulmezaoud, L. Alvarez, M. Colom, J.-M. Morel, *Image Processing On Line* **10**, 191 (2020). https://doi.org/10.5201/ipol.2020.304.

[15] S. Ma, *et al.*, *Medrxiv* (2020).

[16] J. Demongeot, K. Oshinubi, H. Seligmann, F. Thuderoz, *medRxiv* (2021).

[17] A. N. Tikhonov, V. Y. Arsenin, *New York* **1**, 30 (1977).

[18] M. Benning, M. Burger, *Acta Numerica* **27**, 1–111 (2018).

[19] Z. Du, *et al.*, *medRxiv* (2020).

[20] H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, *International journal of infectious diseases* (2020).

[21] 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 (2020). MedRxiv.

Supplementary materials

Material and Methods

We describe and analyze the EpiEstim method and its parameters. We prove that the Wallinga-Teunis method is actually computing R_t by the F_1 form of the renewal equation. Next we detail the boundary conditions used in EpiInvert and some special operations on the incidence curve for countries that do not provide an estimate each day. We show a case study of Cuba, France, Spain and the USA and finally we present 86 results for a collection of countries and US states in alphabetic order.

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^i the EpiEstim estimation. The authors show that if i_t follows a Poisson distribution with expectation $\lambda = \mathbf{E}[i_t] = R_t^i \sum_{s=1}^{t} i_{t-s}\Phi_s$ and R_t^i 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^i:

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

where R_t^i 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^i 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^i. 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 [14], 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

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

which corresponds to the usual R_t^i 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^i, 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.

The Wallinga and Teunis computation of R_t

The Wallinga-Teunis [4] method is also implemented in EpiEstim and widely considered as a reliable method to compute R_t retrospectively [10]. Its formulas require the knowledge of i_t for $t = t_c, \cdots, t + f$. 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 $t = t_l$,

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

It remains to establish a relation of $R^{WT}(t)$ with the solution $\tilde{R}_t$ obtained by the renewal equation with $F \equiv F_1$,

$$\tilde{R}_t = \frac{i_t}{\sum_{s>0} i_{t-s}\Phi_s}. \qquad \text{(F)}$$

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^{WT} can be rewritten as

$$R_t^{WT} = \sum_{\bar{t}} \frac{i_{\bar{t}}\Phi(\bar{t} - t)}{\sum_{s>0} i_{\bar{t}-s}\Phi_s} = \sum_{\bar{t}} \Phi(\bar{t} - t)\tilde{R}_{\bar{t}}. \qquad \text{(G)}$$

We can therefore interpret R_t^{WT} as the forward convolution of the initial estimate $\tilde{R}_t$ with the kernel given by Φ_s.

On the other hand, as explained above, the EpiEstim estimate R_t^i 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^{WT} 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."

Implementation details of EpiInvert

Alternate minimization of the energy (7)

To minimize the energy (7), 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 (7) 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 (7) with respect to $\mathbf{q}$ can be expressed, in matrix form, as the constrained least square minimization problem:

$$min_{\mathbf{q}} \left\| A\mathbf{q} - \mathbf{b} \right\|^2, \tag{H}$$

where $\|.\|$ is the usual Euclidean norm, $A \in R^{T\times 7}$ is defined by

$$A_{k,m} = \begin{cases} 0 & if \quad m \neq k\%7 \\ i_k & if \quad m = k\%7, \end{cases}$$

$\mathbf{b} \in R^T$ is defined by $b_k = F(i, R, \Phi, k)$. The constraint (8) is expressed as an additional linear equation,

$$\mu_0 q_0 + \mu_1 q_1 + \mu_2 q_2 + \mu_3 q_3 + \mu_4 q_4 + \mu_5 q_5 + \mu_6 q_6 = \sum_{t=0}^{t_c} i_t$$

where $\mu_k = \sum_{t=0}^{k+7t \leq t_c} i_{k+7t}$.

This linear constraint is easily included in the minimization procedure. 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 (7) 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 $MaxIter = 100$.

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 (7) with respect to R_t.
for $n = 1, 2, .., MaxIter$ **do**
  compute $\mathbf{q}^n = \mathbf{q}(R^{n-1})$ minimizing (7) with respect to $\mathbf{q}$.
  compute $i_t^n = q_{t\%7}^n i_t$.
  compute $\mathcal{I}^n$ using (11).
  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 (7) with respect to R_t.
    $R = R^n$.
  end
end
return $\hat{i}$, R, $\mathbf{q}$.

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

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

$$a = median(\{log\left(\frac{I_{t+1}}{I_t}\right) \ : \ t = 0, .., 14\}). \tag{I}$$

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

$$i_0 = I_0(1 - e^{-a}) = I_0 R0 \sum_{k=f_0}^{f} (e^{-ka} - e^{-(k+1)a})\Phi_k. \tag{J}$$

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

$$R0 = \frac{1 - e^{-a}}{\sum_{k=f_0}^{f} (e^{-ka} - e^{-(k+1)a})\Phi_k}. \tag{K}$$

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 [21] 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.

Boundary condition for $[t > t_c]$

The proposed inversion model provides an estimation of R_t up to the current day t_c. An obvious objection is that if $f_0 < 0$, the functional (7) 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

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

The above defined boundary conditions has a very minor influence in the final estimation of R_t in the last days when minimizing (7). 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 $Q_0 = 6.55872E-06$, $Q_1 = 0.000126075$, $Q_2 = 0.000313458$, $Q_3 = 0.000565911$ and $Q_4 = 0.004890481$. 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 missing days. We do not allow negative numbers in the incidence curve. By default, we replace by zero any negative value of the incidence curve.

Case studies: Cuba, France, Spain, USA and World

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. For the US states, the data are obtained from the New York Times report available at `https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv`.

In Fig. S1 we show the charts obtained for the world population with $F \equiv F_1$ and $F \equiv F_2$. Table S1 contains a summary of the values computed for each experiment. To compute the EpiEstim estimation R_t^i, 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 total world population shows a clear weekly periodic bias. The correction of this bias works quite well, as the RMSE reduction reaches $\mathcal{I} = 0.337$ for $F \equiv F_1$ and $\mathcal{I} = 0.380$ for $F \equiv F_2$. The oscillation of the incidence curve is strongly reduced, passing from $\mathcal{V}(i) = 0.115$ to $\mathcal{V}(\hat{i}) = 0.063$. The agreement with EpiEstim is also excellent as $\mathcal{S}(\tilde{t}) = 0.01$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.014$ for $F \equiv F_2$. The daily bias correction factors are similar for $F \equiv F_1$ and $F \equiv F_2$. On Sundays and Mondays the number of cases is underestimated and overestimated on Wednesdays, Thursdays and Fridays.

France also displays a clear weekly periodic bias: on Mondays the number of cases is strongly underestimated, and on Wednesdays it is strongly overestimated. The correction of the periodic bias works well, as $\mathcal{I} = 0.481$ for $F \equiv F_1$ and $\mathcal{I} = 0.513$ for $F \equiv F_2$. The oscillation of the incidence curve is therefore reduced, passing from $\mathcal{V}(i) = 0.329$ to $\mathcal{V}(\hat{i}) = 0.202$. The agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.026$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.025$ for $F \equiv F_2$.

Spain is special: it does not provide data on weekends or holidays. In that case a constant value is being assigned to i_t in the affected days. Despite this, the correction of the weekly periodic bias works again well and yields $\mathcal{I} = 0.171$ for $F \equiv F_1$ and $\mathcal{I} = 0.290$ for $F \equiv F_2$. The oscillation of the incidence curve reduces from $\mathcal{V}(i) = 0.135$ to $\mathcal{V}(\hat{i}) = 0.087$. The agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.025$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.046$ for $F \equiv F_2$. Observe how the incidence curve is underestimated on Sundays, Mondays and Tuesdays, and overestimated on Thursdays, Fridays and Saturdays.

In the USA we obtain $\mathcal{I} = 0.450$ for $F \equiv F_1$ and $\mathcal{I} = 0.569$ for $F \equiv F_2$. The oscillation of the incidence curve

is reduced from $\mathcal{V}(i) = 0.130$ to $\mathcal{V}(\hat{i}) = 0.085$. The agreement with EpiEstim is again very good with $\mathcal{S}(\tilde{t}) = 0.014$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.023$ for $F \equiv F_2$. On Sundays the number of cases is underestimated, and overestimated on Fridays.

Although in general countries present a clear weekly periodic pattern in the incidence curve this is not the case for Cuba. In this country we obtain $\mathcal{I} = 0.890$ for $F \equiv F_1$ and $\mathcal{I} = 0.928$ for $F \equiv F_2$. The incidence curve oscillation is slightly reduced after the correction of the periodic bias. Finally, the agreement with the EpiEstim method is good, with $\mathcal{S}(\tilde{t}) = 0.034$ for $F \equiv F_1$ and $\mathcal{S}(\tilde{t}) = 0.041$ for $F \equiv F_2$.

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.

The optimal shift $\tilde{t}$ between R_t is R_t^i fits in the range obtained by a joint analysis of the 70 countries. Indeed, for $F \equiv F_1$ $\tilde{t}$ ranges from 2.72 to 3.50 and for $F \equiv F_2$ $\tilde{t}$ ranges from 8.00 to 9.7.

Additional experiments

We can start this large set of experiments with a recent example in France showing how EpiInvert gives a valuable extension to EpiEstim. In Fig. S2 we observe a very good agreement between the EpiEstim estimate of $R(t)$ by March 26 ($R(t) = 1.239$) and the EpiInvert estimate 8 days in advance ($R(t) = 1.221$). But the EpiInvert estimate is more regular and it does not produce the singularity observed in the EpiEstim estimate by March 15.

Next, for $F \equiv F_2$, using the data of incidence curve up to March 26, 2021, we present a collection of 64 experiments on different countries and separately 24 experiments on some US states. The regions are sorted in alphabetic order. For each experiment we show the charts and the following selection of numerical values:

1. $\mathcal{I}$: reduction factor of the RMSE error between i_t and $F(i, R, \Phi, t)$ before and after the correction of the weekly periodic bias defined as:

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

   where $\hat{i}$ represents the incidence curve after correction and $R1(.)$ represents the initial R_t estimate without correcting the periodic bias. In the case we do not apply the correction of the periodic bias, this value does not appear in the experiment.
2. $\tilde{t}$: optimal shift (in days) between our estimate of R and the one obtained by EpiEstim.
3. $\mathcal{S}(\tilde{t})$: RMSE between our estimate of R and the one obtained by EpiEstim shifted back $\tilde{t}$ days.
4. $R_{t_c}^i$: last available value of the EpiEstim estimate R_t^i.
5. R_{t_c} : last available value of our R_t estimate.

The default value for the regularization parameter is $w = 5$ for $F \equiv F_1$ and $w = 5.5$ for $F \equiv F_2$, otherwise it is explicitly written in the experiment.

In general the correction of the weekly periodic bias works quite well. We can highlight the following regions where such correction works extremely well: Germany ($\mathcal{I} = 0.166$), Croatia ($\mathcal{I} = 0.174$), Sweden ($\mathcal{I} = 0.192$), Switzerland ($\mathcal{I} = 0.206$), Poland ($\mathcal{I} = 0.227$), Portugal ($\mathcal{I} = 0.23$), and Utah ($\mathcal{I} = 0.292$). On the other side, there is also a number of regions where the correction of the weekly bias does not work well as Uruguay ($\mathcal{I} = 0.822$), China ($\mathcal{I} = 0.826$), Peru ($\mathcal{I} = 0.83$), Ethiopia ($\mathcal{I} = 0.831$), Indonesia ($\mathcal{I} = 0.89$), Cuba ($\mathcal{I} = 0.928$), Cyprus ($\mathcal{I} = 0.94$), Washington ($\mathcal{I} = 0.816$), New York ($\mathcal{I} = 0.86$) and Connecticut ($\mathcal{I} = 0.936$).

For some regions where the value of $\mathcal{S}(\tilde{t})$ is high, we repeat the experiment without correcting the weekly periodic bias. In general, in such cases we observe that profile of the R_t estimate is similar in both cases and the experimental variability is lower (or similar) in the case of using the bias correction. That is the case of Canada, China, Cuba, Cyprus, Ethiopia, Peru, New York or Washington. In the case of Denmark, both estimations of R_t are quite different and the experimental variability are very high due to some high oscillations of the incidence curve in the last days of the sequence.

A very special case is Kansas. As can be observed in Fig. S21, although the correction of the weekly periodic bias is not bad ($\mathcal{I} = 0.595$) the obtained R_t estimate is very inaccurate. The reason is that the incidence curve of Kansas is extremely oscillating ($\mathcal{V}(i) = 1.728$ (notice that in the USA $\mathcal{V}(i) = 0.130$)) but the oscillations are not 7-day periodic and the correction of the weekly periodic bias produces high distortions of the incidence curve when we approach the last day of the sequence. In this very particular case it is clearly better to do not use the correction of the weekly bias. Moreover, due to the extremely oscillating behaviour of the incidence curve, as shown in the experiments of Fig. S21 a high value of the regularization parameter ($w = 40$) is required in order to properly regularize the estimate of R_t.

Concerning the agreement with EpiEstim we observe that countries with small oscillations in the incidence curve like Iran ($\mathcal{V}(i) = 0.023$) or Russia ($\mathcal{V}(i) = 0.031$) show excellent agreement with EspiEstim ($\mathcal{S}(\tilde{t}) = 0.006$ for Iran and $\mathcal{S}(\tilde{t}) = 0.01$ for Russia). On the other hand, countries with small number of cases like China have no good agreement with EpiEstim ($\mathcal{S}(\tilde{t}) = 0.142$) with the default value of the regularization parameter w. In such cases our R_t estimate is much more regular than the one of EpiEstim.

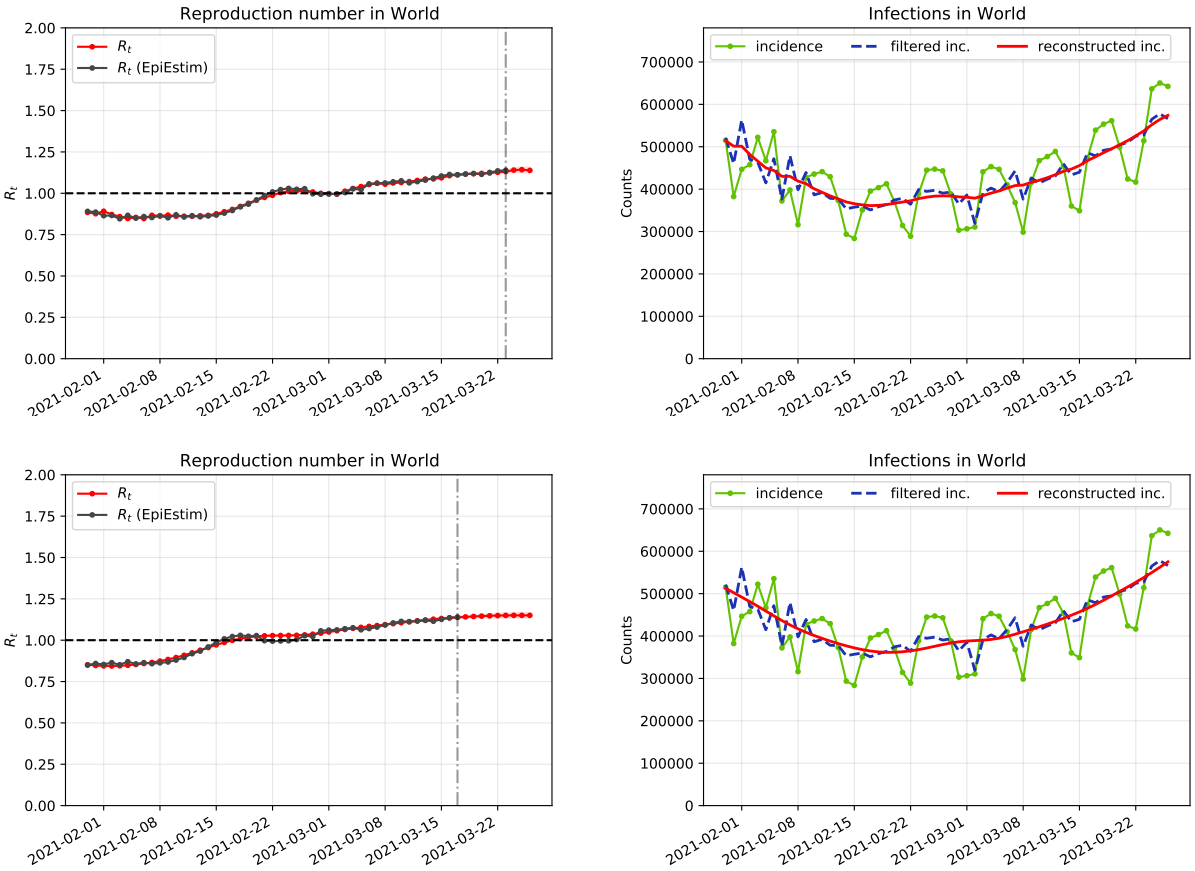

Figure S1: Results obtained for the world population up to March 26, 2021 using: (top) $F \equiv F_1$ and (down) $F \equiv F_2$.

	World	World	France	France	Spain	Spain	USA	USA	Cuba	Cuba
F	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
$\tilde{t}$	2.72	8.94	3.50	9.05	3.38	9.70	2.72	8.76	3.09	8.00
$\mathcal{S}(\tilde{t})$	0.010	0.014	0.026	0.025	0.025	0.046	0.014	0.023	0.034	0.041
$\mathcal{I}$	0.337	0.380	0.481	0.513	0.171	0.290	0.450	0.569	0.890	0.928
q_1	1.011	1.012	0.931	0.932	1.263	1.266	1.008	1.005	1.005	1.006
q_2	1.204	1.204	1.073	1.078	1.227	1.208	1.262	1.250	0.945	0.945
q_3	1.260	1.259	3.201	3.180	1.177	1.149	1.095	1.083	1.049	1.049
q_4	1.027	1.026	1.062	1.062	1.031	1.009	1.053	1.057	0.923	0.921
q_5	0.887	0.888	0.698	0.691	0.857	0.858	0.925	0.936	1.025	1.025
q_6	0.888	0.889	0.889	0.886	0.802	0.817	0.909	0.915	1.037	1.037
q_7	0.881	0.881	0.944	0.955	0.834	0.863	0.867	0.869	1.026	1.027

Table S1: Numerical results obtained by EpiInvert for the world population, France, Spain, the USA and Cuba using the data up to March 26, 2021 and the renewal equations $F = F_1$ and $F = F_2$.

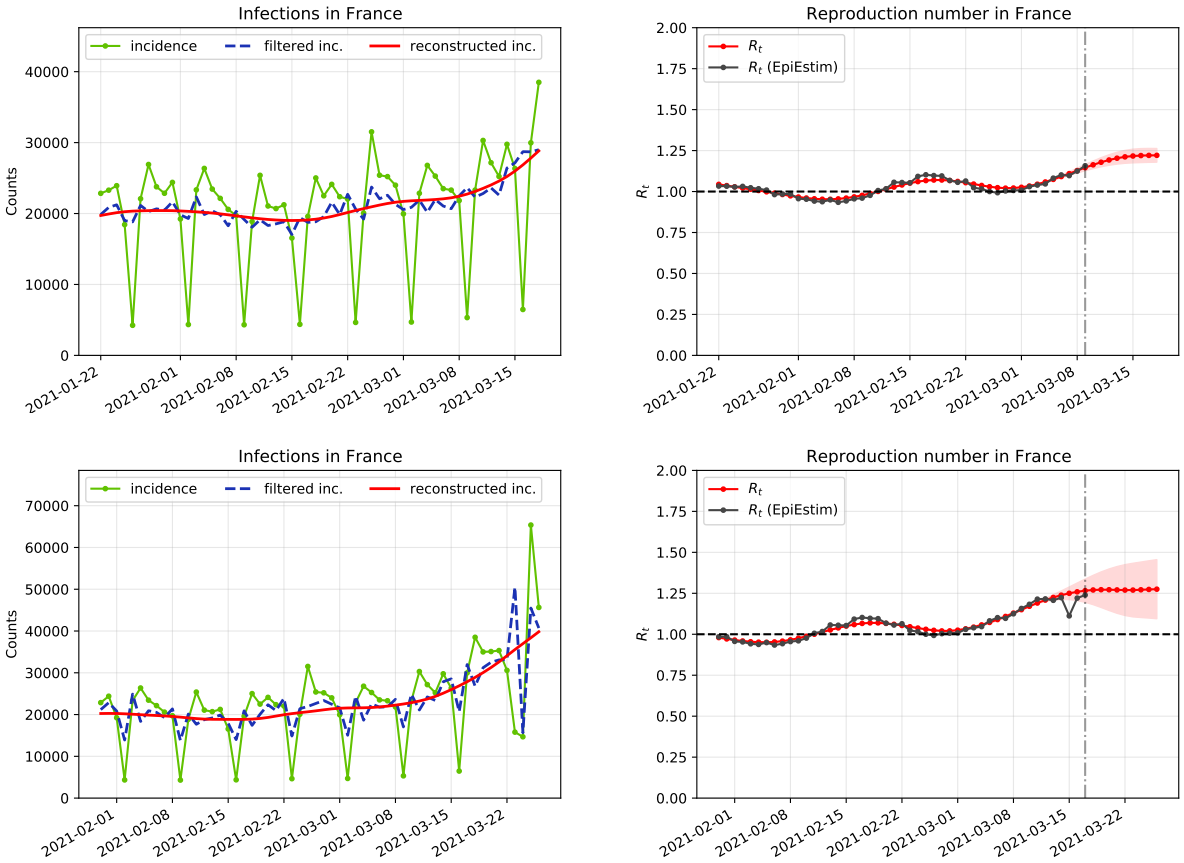

Figure S2: Comparison of the EpiEstim estimate of $R(t)$ in France by March 26 ($R(t) = 1.239$) and the EpiInvert estimate 8 days in advance ($R(t) = 1.221$). Notice the singularity observed in the EpiEstim estimate by March 15.

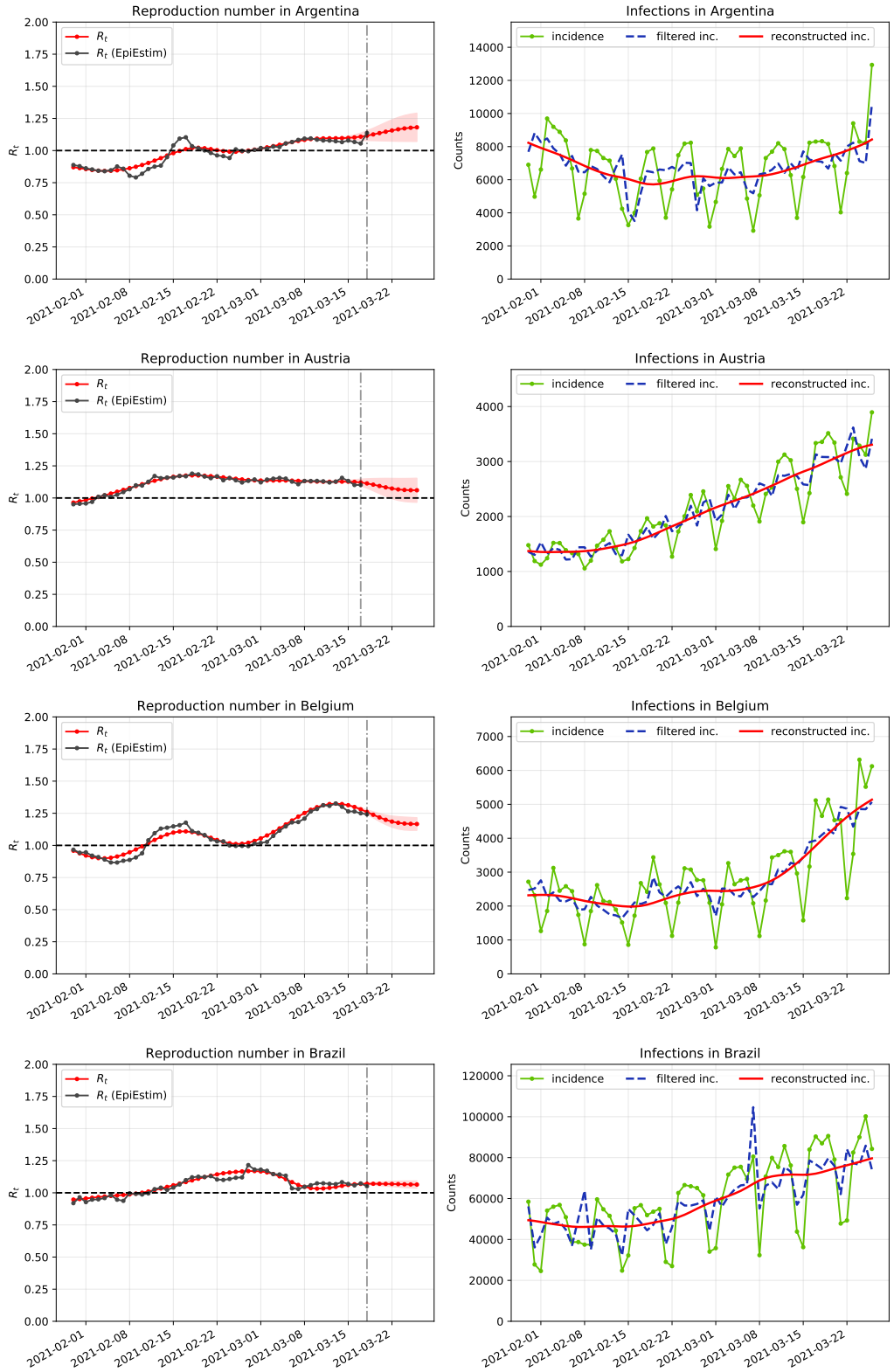

Figure S3: From top to down: Argentina ($\mathcal{I} = 0.451$, $\tilde{t} = 8.11$, $\mathcal{S}(\tilde{t}) = 0.033$, $R^i(t_c) = 1.137$, $R(t_c) = 1.181$), Austria ($\mathcal{I} = 0.464$, $\tilde{t} = 9.37$, $\mathcal{S}(\tilde{t}) = 0.014$, $R^i(t_c) = 1.101$, $R(t_c) = 1.060$), Belgium ($\mathcal{I} = 0.293$, $\tilde{t} = 8.49$, $\mathcal{S}(\tilde{t}) = 0.031$, $R^i(t_c) = 1.242$, $R(t_c) = 1.166$) and Brazil ($\mathcal{I} = 0.560$, $\tilde{t} = 8.21$, $\mathcal{S}(\tilde{t}) = 0.027$, $R^i(t_c) = 1.051$, $R(t_c) = 1.064$).

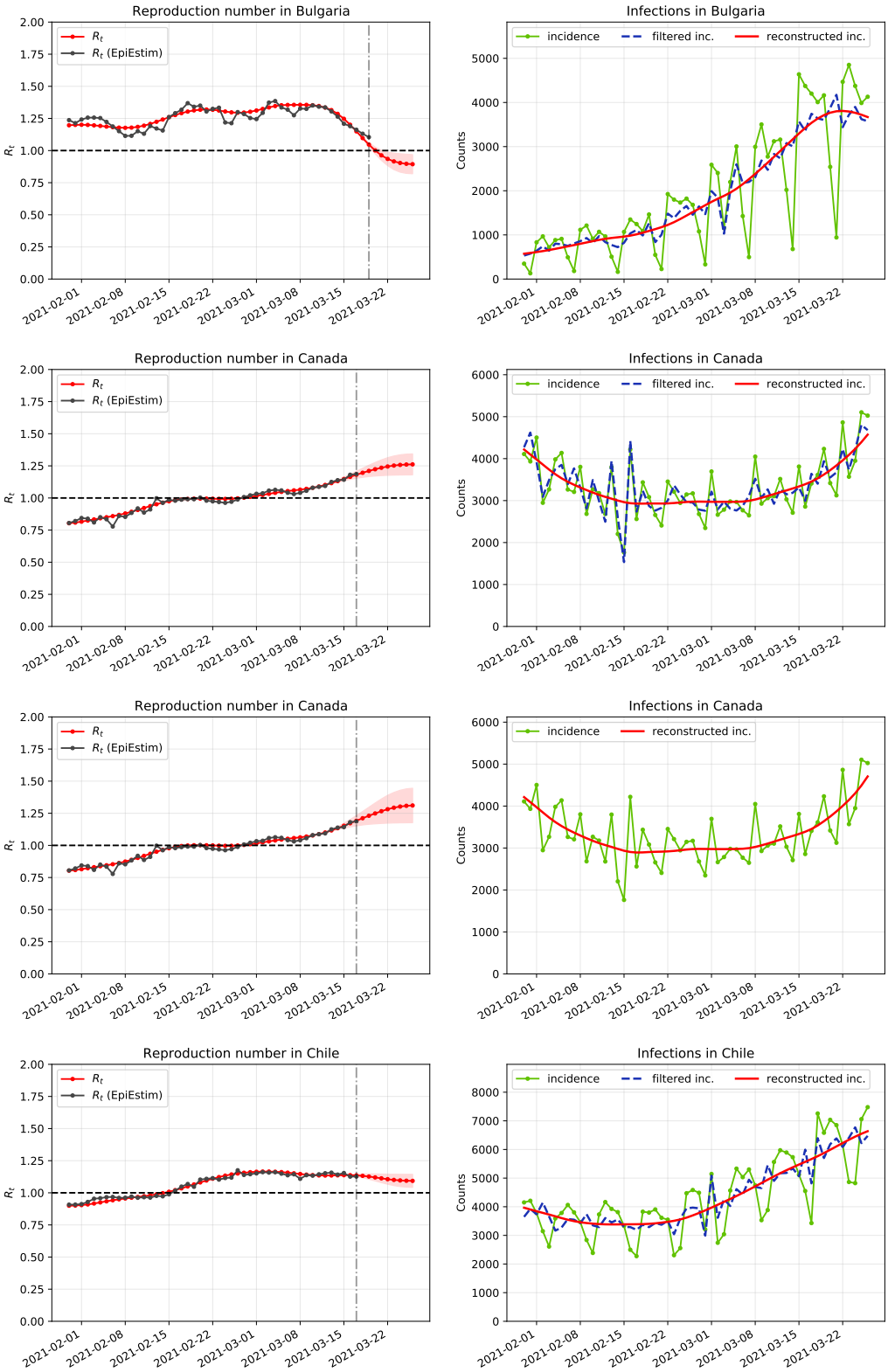

Figure S4: From top to down: Bulgaria ($\mathcal{I} = 0.245$, $\tilde{t} = 7.23$, $\mathcal{S}(\tilde{t}) = 0.041$, $R^i(t_c) = 1.105$, $R(t_c) = 0.894$), Canada ($\mathcal{I} = 0.780$, $\tilde{t} = 9.10$, $\mathcal{S}(\tilde{t}) = 0.019$, $R^i(t_c) = 1.187$, $R(t_c) = 1.262$), Canada ($\tilde{t} = 9.07$, $\mathcal{S}(\tilde{t}) = 0.019$, $R^i(t_c) = 1.187$, $R(t_c) = 1.311$) and Chile ($\mathcal{I} = 0.385$, $\tilde{t} = 8.95$, $\mathcal{S}(\tilde{t}) = 0.017$, $R^i(t_c) = 1.127$, $R(t_c) = 1.093$).

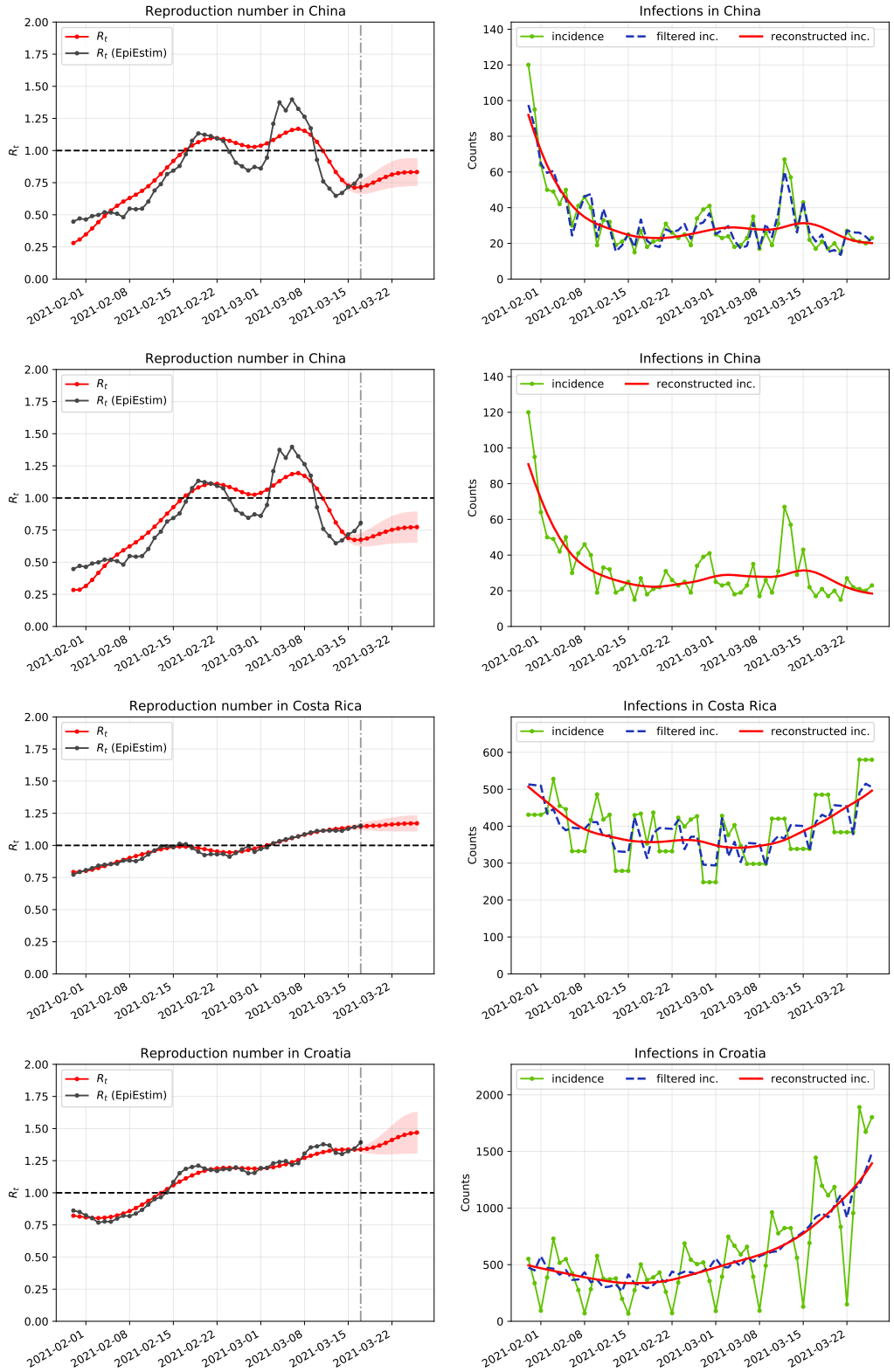

Figure S5: From top to down: China ($\mathcal{I} = 0.826$, $\tilde{t} = 9.30$, $\mathcal{S}(\tilde{t}) = 0.142$, $R^i(t_c) = 0.805$, $R(t_c) = 0.833$), China ($\tilde{t} = 9.11$, $\mathcal{S}(\tilde{t}) = 0.131$, $R^i(t_c) = 0.805$, $R(t_c) = 0.773$), Costa Rica ($\mathcal{I} = 0.3799$, $\tilde{t} = 8.97$, $\mathcal{S}(\tilde{t}) = 0.018$, $R^i(t_c) = 1.154$, $R(t_c) = 1.172$) and Croatia ($\mathcal{I} = 0.174$, $\tilde{t} = 8.88$, $\mathcal{S}(\tilde{t}) = 0.033$, $R^i(t_c) = 1.393$, $R(t_c) = 1.469$).

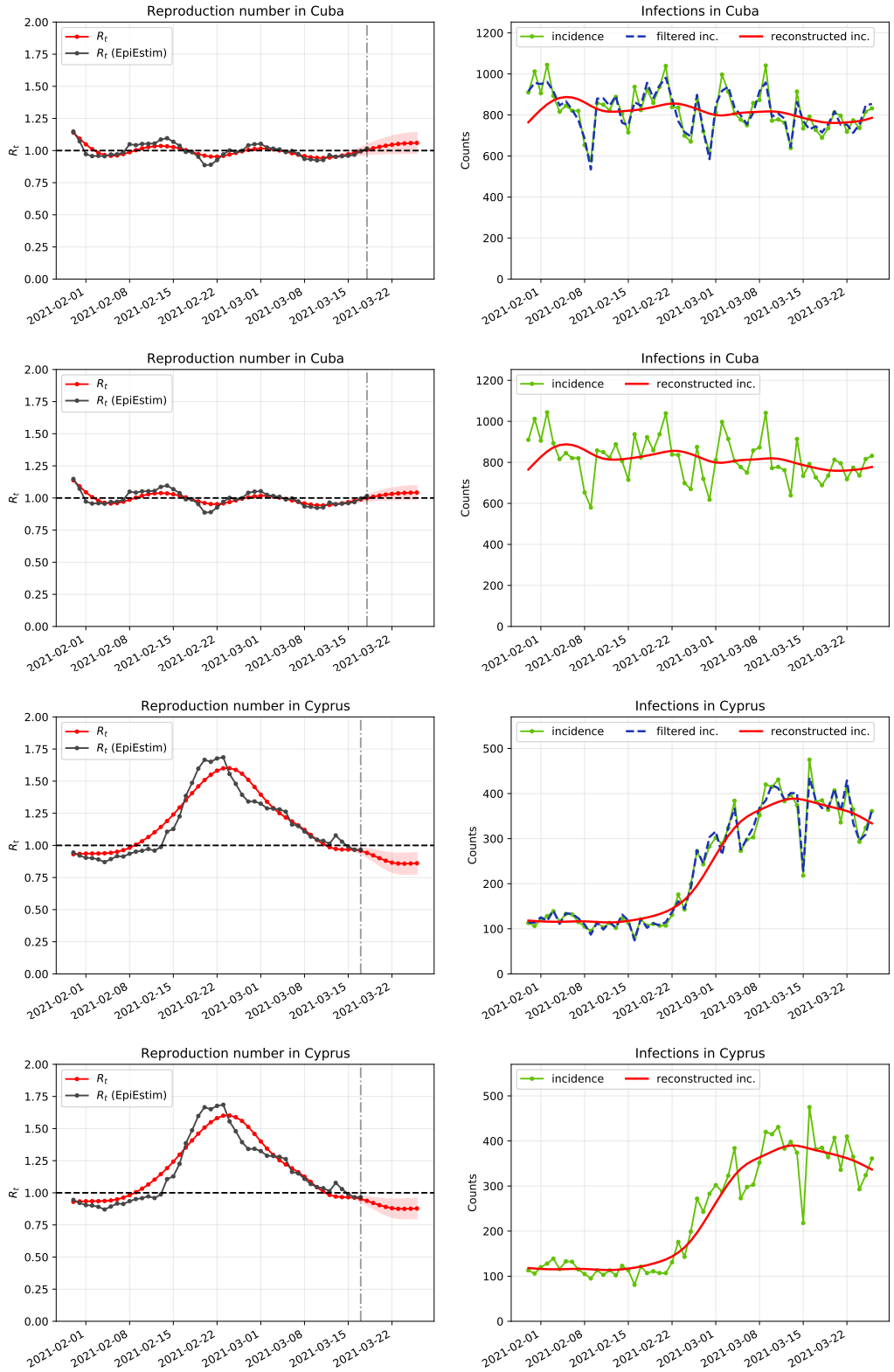

Figure S6: From top to down: Cuba ($\mathcal{I} = 0.928$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.041$, $R^i(t_c) = 1.016$, $R(t_c) = 1.060$), Cuba ($\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.040$, $R^i(t_c) = 1.016$, $R(t_c) = 1.042$), Cyprus ($\mathcal{I} = 0.940$, $\tilde{t} = 8.77$, $\mathcal{S}(\tilde{t}) = 0.074$, $R^i(t_c) = 0.965$, $R(t_c) = 0.861$) and Cyprus ($\tilde{t} = 8.77$, $\mathcal{S}(\tilde{t}) = 0.074$, $R^i(t_c) = 0.965$, $R(t_c) = 0.878$).

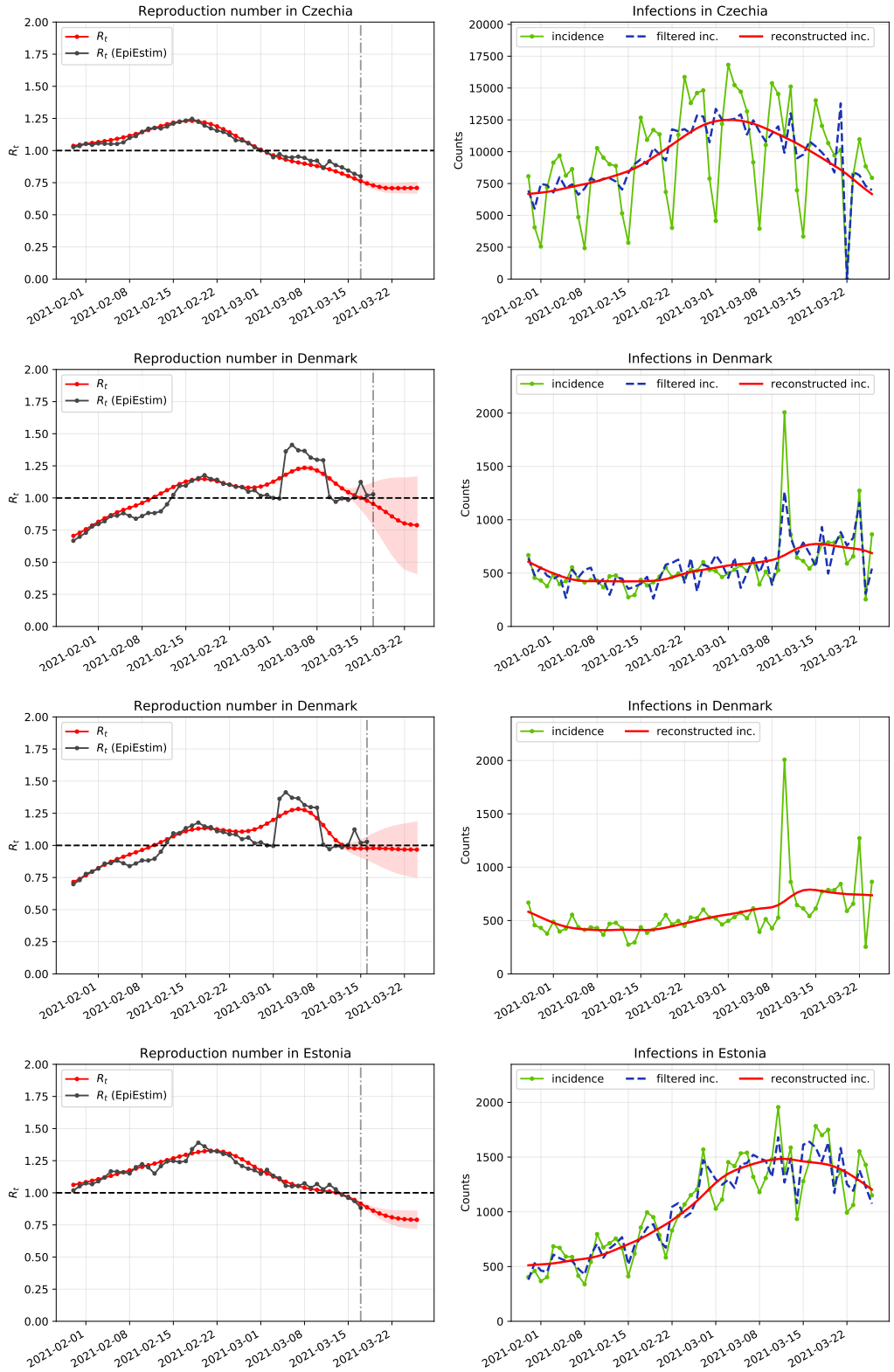

Figure S7: From top to down: Czechia ($\mathcal{I} = 0.3733$, $\tilde{t} = 9.39$, $\mathcal{S}(\tilde{t}) = 0.026$, $R^i(t_c) = 0.795$, $R(t_c) = 0.709$), Denmark ($\mathcal{I} = 0.717$, $\tilde{t} = 7.42$, $\mathcal{S}(\tilde{t}) = 0.080$, $R^i(t_c) = 1.029$, $R(t_c) = 0.788$) and Denmark ($\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.074$, $R^i(t_c) = 1.029$, $R(t_c) = 0.967$), Estonia ($\mathcal{I} = 0.661$, $\tilde{t} = 8.81$, $\mathcal{S}(\tilde{t}) = 0.029$, $R^i(t_c) = 0.883$, $R(t_c) = 0.790$).

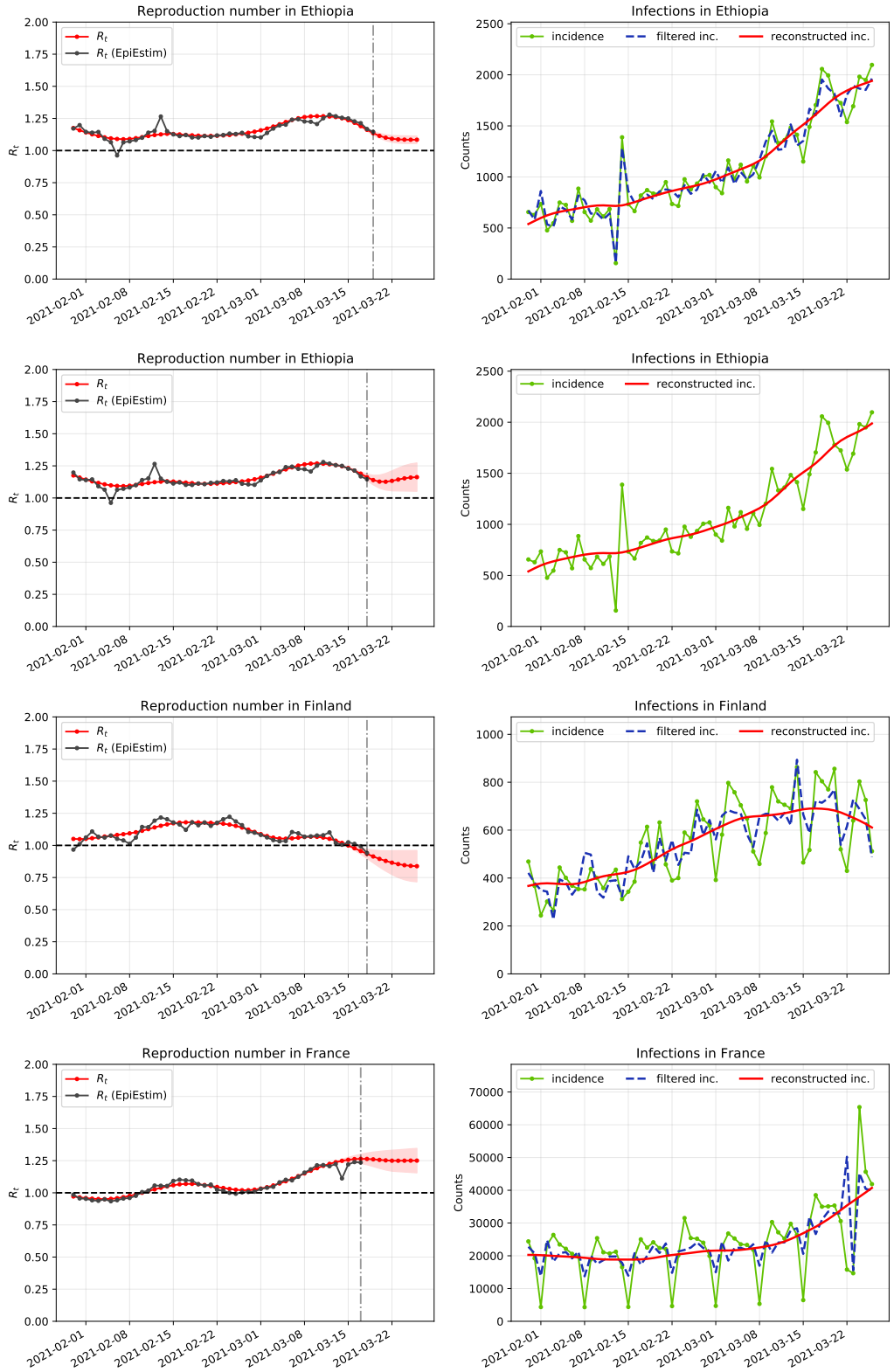

Figure S8: From top to down: Ethiopia ($\mathcal{I} = 0.831$, $\tilde{t} = 7.43$, $\mathcal{S}(\tilde{t}) = 0.049$, $R^i(t_c) = 1.146$, $R(t_c) = 1.084$), Ethiopia ($\tilde{t} = 7.50$, $\mathcal{S}(\tilde{t}) = 0.050$, $R^i(t_c) = 1.146$, $R(t_c) = 1.162$), Finland ($\mathcal{I} = 0.630$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.038$, $R^i(t_c) = 0.941$, $R(t_c) = 0.838$) and France ($\mathcal{I} = 0.513$, $\tilde{t} = 9.05$, $\mathcal{S}(\tilde{t}) = 0.025$, $R^i(t_c) = 1.236$, $R(t_c) = 1.251$).

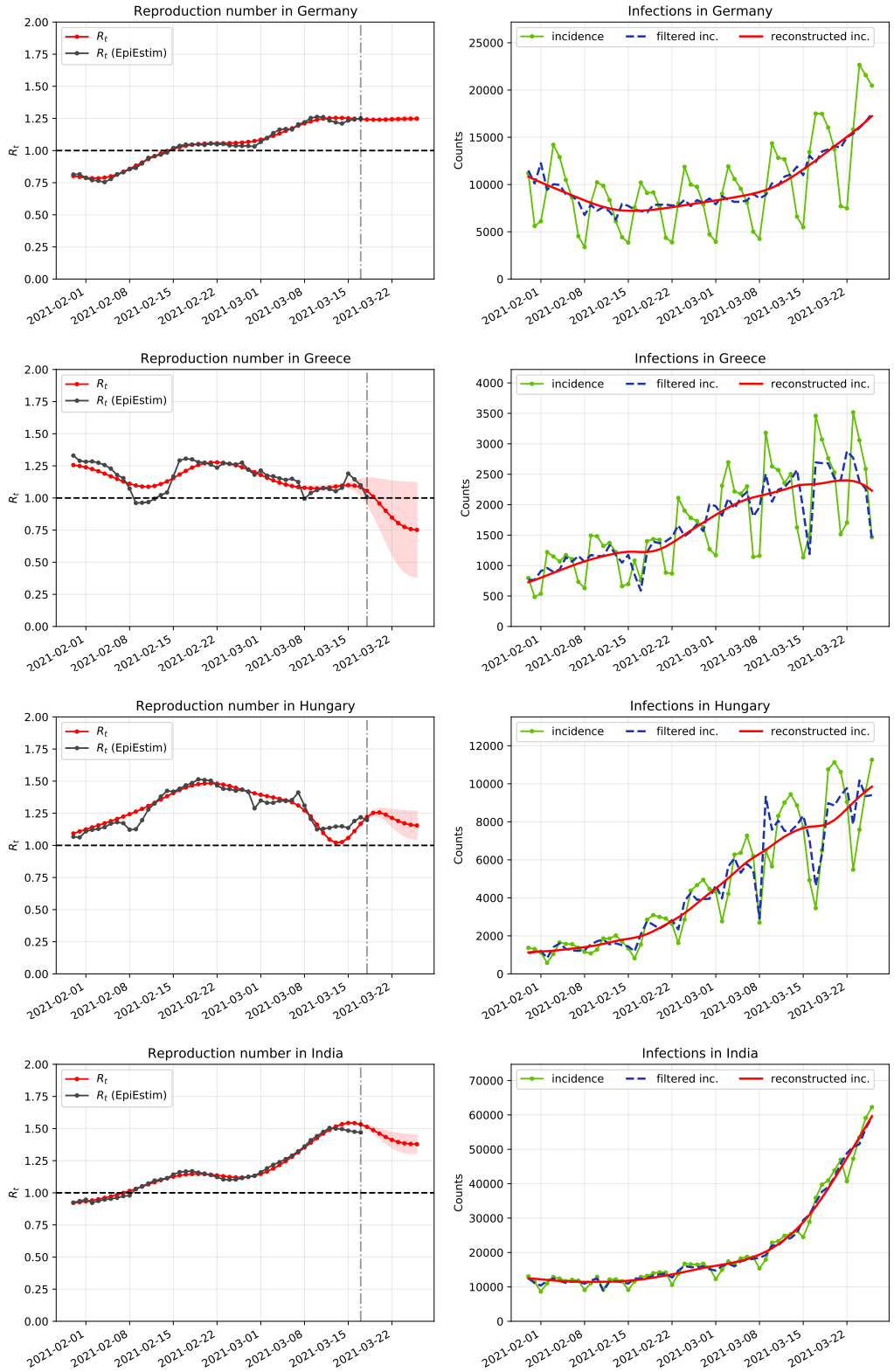

Figure S9: From top to down: Germany ($\mathcal{I} = 0.166$, $\tilde{t} = 9.00$, $\mathcal{S}(\tilde{t}) = 0.018$, $R^i(t_c) = 1.251$, $R(t_c) = 1.249$), Greece ($\mathcal{I} = 0.507$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.061$, $R^i(t_c) = 1.007$, $R(t_c) = 0.752$), Hungary ($\mathcal{I} = 0.644$, $\tilde{t} = 8.15$, $\mathcal{S}(\tilde{t}) = 0.052$, $R^i(t_c) = 1.198$, $R(t_c) = 1.155$) and India ($\mathcal{I} = 0.3734$, $\tilde{t} = 8.59$, $\mathcal{S}(\tilde{t}) = 0.022$, $R^i(t_c) = 1.470$, $R(t_c) = 1.378$).

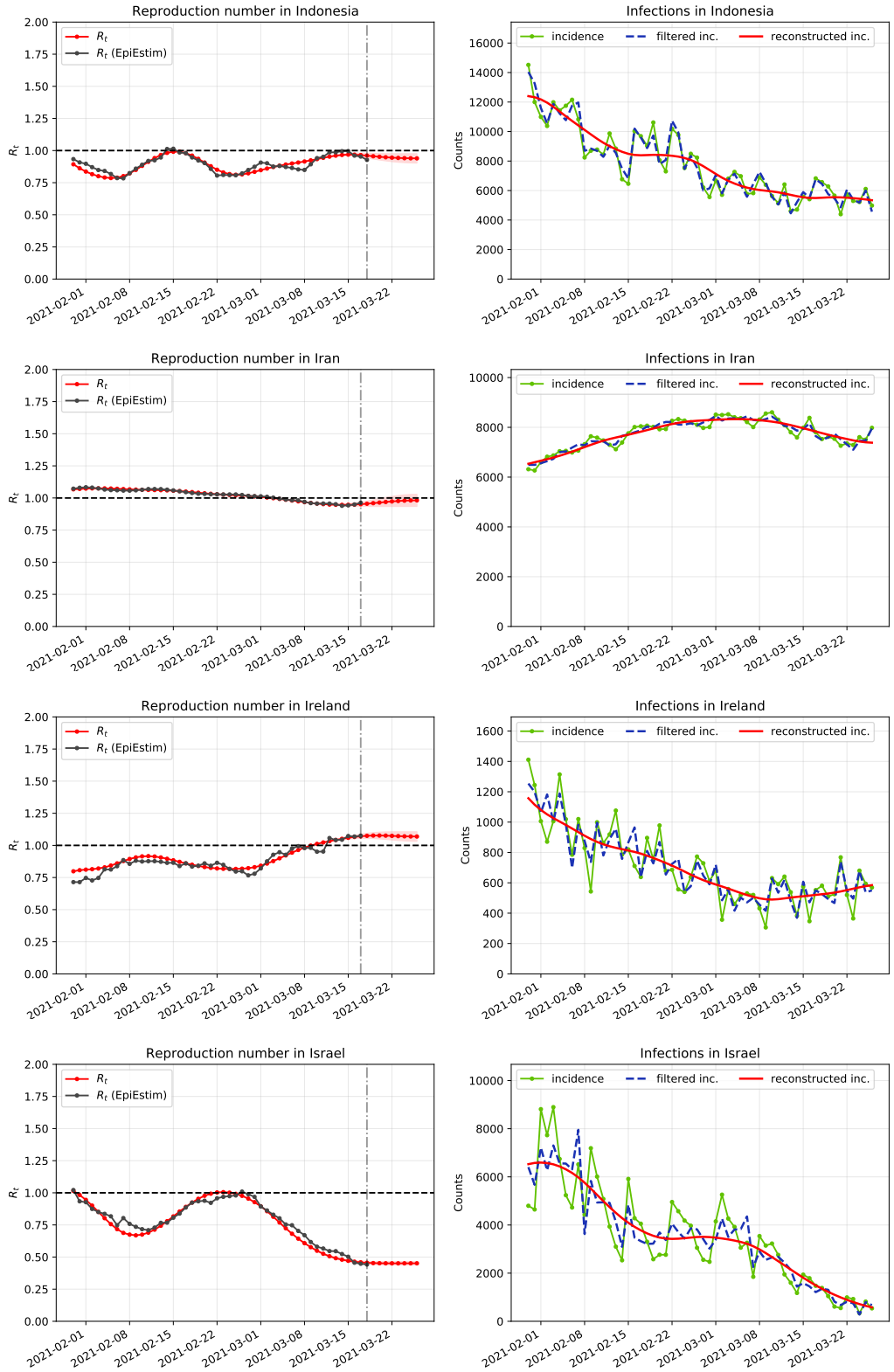

Figure S10: From top to down: Indonesia ($\mathcal{I} = 0.890$, $\tilde{t} = 8.18$, $\mathcal{S}(\tilde{t}) = 0.029$, $R^i(t_c) = 0.927$, $R(t_c) = 0.939$), Iran ($\mathcal{I} = 0.676$, $\tilde{t} = 9.11$, $\mathcal{S}(\tilde{t}) = 0.006$, $R^i(t_c) = 0.963$, $R(t_c) = 0.982$), Ireland ($\mathcal{I} = 0.730$, $\tilde{t} = 8.57$, $\mathcal{S}(\tilde{t}) = 0.045$, $R^i(t_c) = 1.075$, $R(t_c) = 1.069$) and Israel ($\mathcal{I} = 0.570$, $\tilde{t} = 8.45$, $\mathcal{S}(\tilde{t}) = 0.042$, $R^i(t_c) = 0.3742$, $R(t_c) = 0.451$).

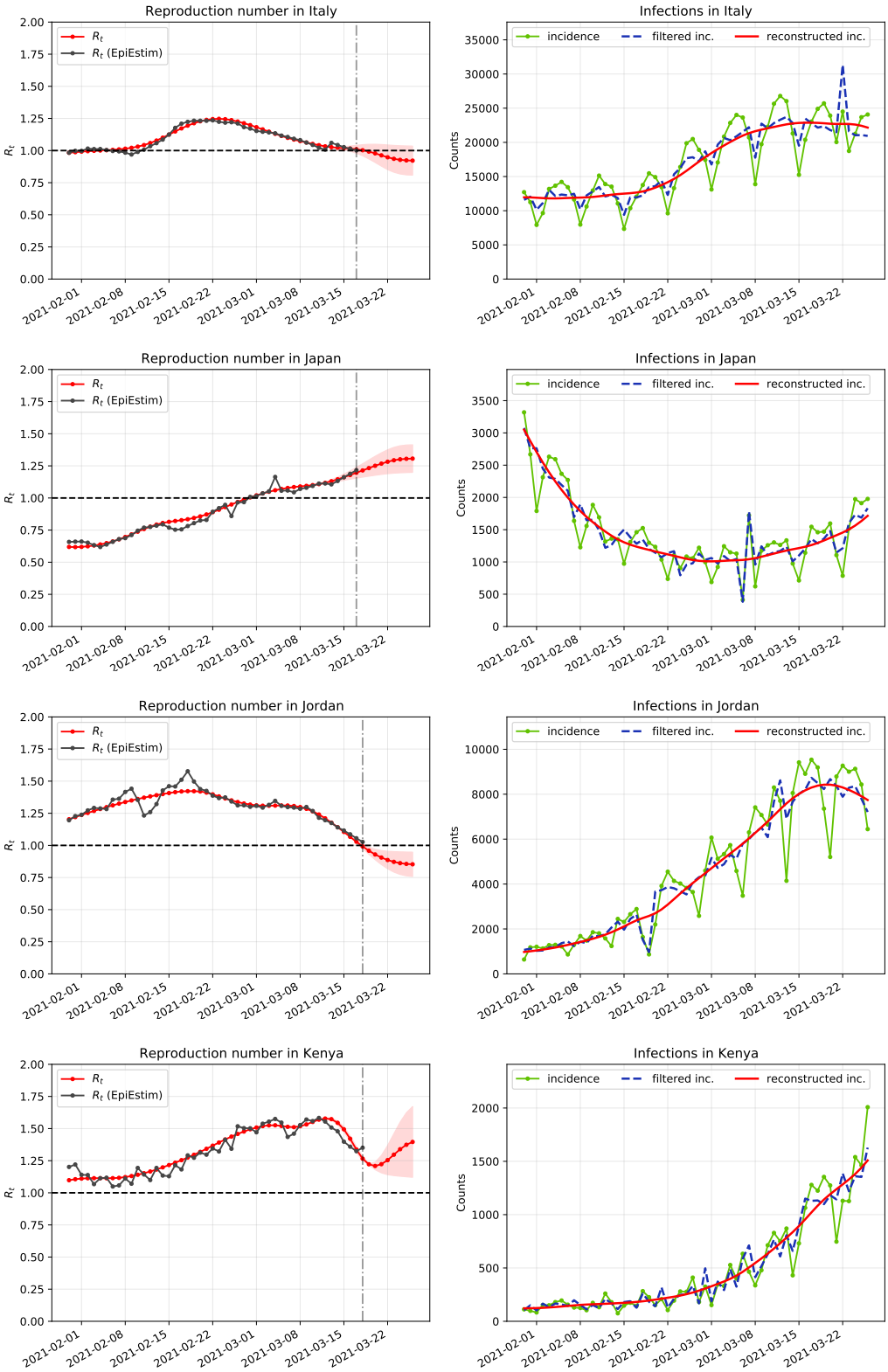

Figure S11: From top to down: Italy ($\mathcal{I} = 0.584$, $\tilde{t} = 8.74$, $\mathcal{S}(\tilde{t}) = 0.020$, $R^i(t_c) = 1.001$, $R(t_c) = 0.922$), Japan ($\mathcal{I} = 0.570$, $\tilde{t} = 8.93$, $\mathcal{S}(\tilde{t}) = 0.041$, $R^i(t_c) = 1.216$, $R(t_c) = 1.307$), Jordan ($\mathcal{I} = 0.437$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.047$, $R^i(t_c) = 1.027$, $R(t_c) = 0.853$) and Kenya ($\mathcal{I} = 0.574$, $\tilde{t} = 7.91$, $\mathcal{S}(\tilde{t}) = 0.056$, $R^i(t_c) = 1.351$, $R(t_c) = 1.396$).

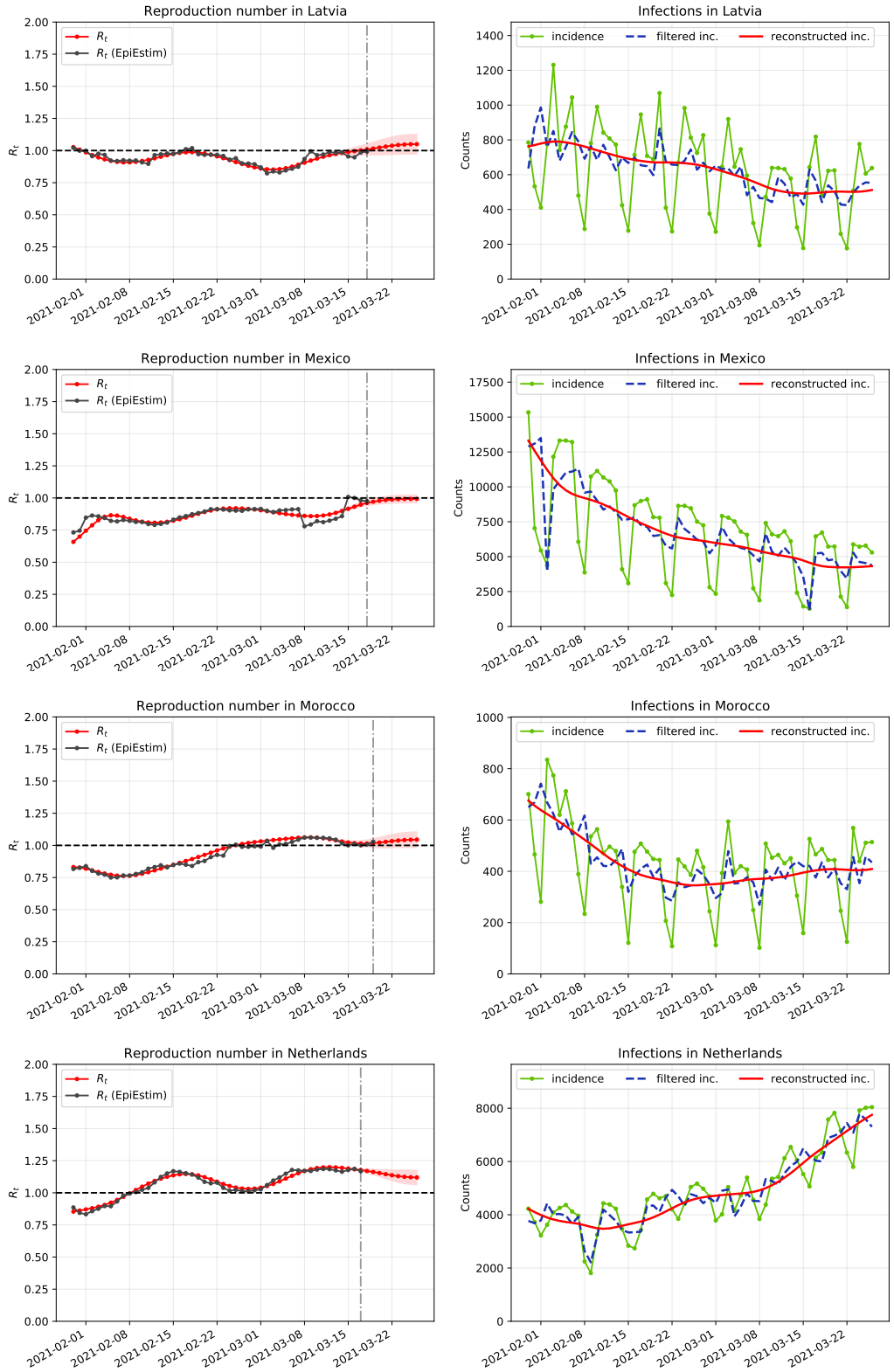

Figure S12: From top to down: Latvia ($\mathcal{I} = 0.299$, $\tilde{t} = 8.43$, $\mathcal{S}(\tilde{t}) = 0.026$, $R^i(t_c) = 0.993$, $R(t_c) = 1.050$), Mexico ($\mathcal{I} = 0.449$, $\tilde{t} = 8.46$, $\mathcal{S}(\tilde{t}) = 0.051$, $R^i(t_c) = 0.976$, $R(t_c) = 0.993$), Morocco ($\mathcal{I} = 0.329$, $\tilde{t} = 7.32$, $\mathcal{S}(\tilde{t}) = 0.025$, $R^i(t_c) = 1.030$, $R(t_c) = 1.045$) and Netherlands ($\mathcal{I} = 0.601$, $\tilde{t} = 8.55$, $\mathcal{S}(\tilde{t}) = 0.023$, $R^i(t_c) = 1.168$, $R(t_c) = 1.120$).

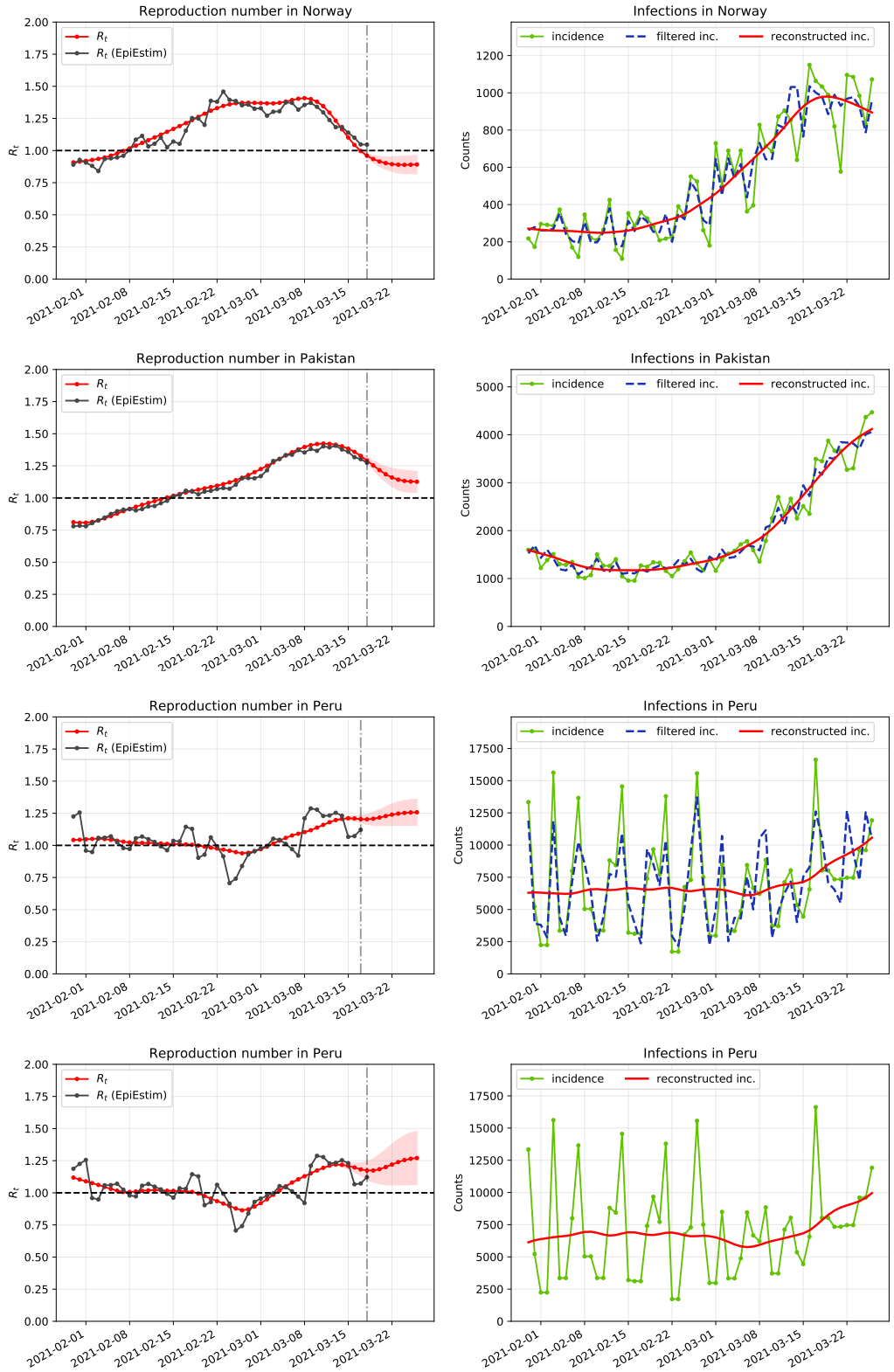

Figure S13: From top to down: Norway ($\mathcal{I} = 0.609$, $\tilde{t} = 8.48$, $\mathcal{S}(\tilde{t}) = 0.053$, $R^i(t_c) = 1.046$, $R(t_c) = 0.891$), Pakistan ($\mathcal{I} = 0.557$, $\tilde{t} = 8.37$, $\mathcal{S}(\tilde{t}) = 0.027$, $R^i(t_c) = 1.274$, $R(t_c) = 1.127$), Peru ($\mathcal{I} = 0.830$, $\tilde{t} = 8.68$, $\mathcal{S}(\tilde{t}) = 0.105$, $R^i(t_c) = 1.122$, $R(t_c) = 1.258$) and Peru ($\tilde{t} = 8.02$, $\mathcal{S}(\tilde{t}) = 0.088$, $R^i(t_c) = 1.122$, $R(t_c) = 1.270$).

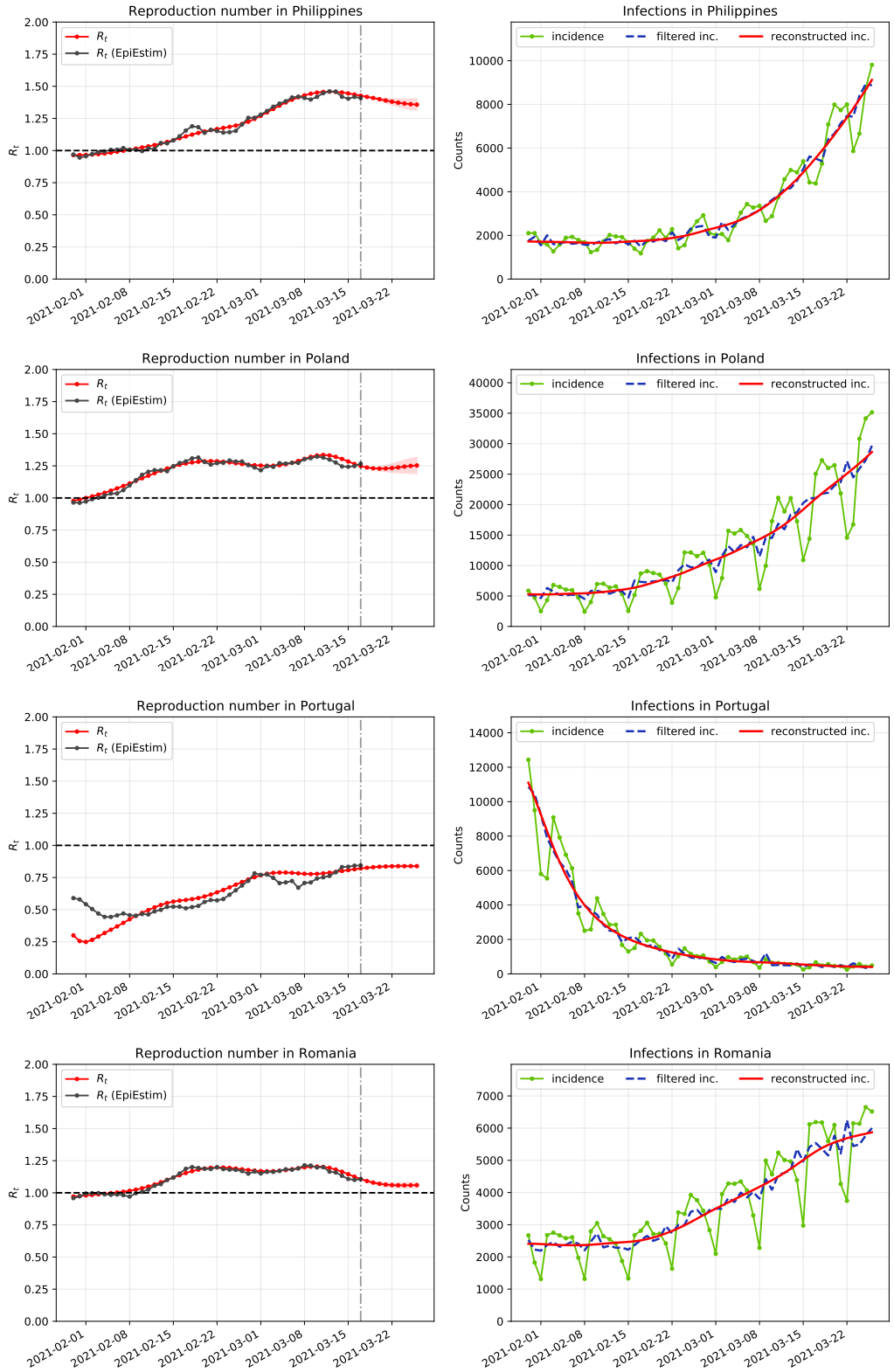

Figure S14: From top to down: Philippines ($\mathcal{I} = 0.313$, $\tilde{t} = 8.55$, $\mathcal{S}(\tilde{t}) = 0.026$, $R^i(t_c) = 1.408$, $R(t_c) = 1.358$), Poland ($\mathcal{I} = 0.227$, $\tilde{t} = 8.77$, $\mathcal{S}(\tilde{t}) = 0.022$, $R^i(t_c) = 1.267$, $R(t_c) = 1.253$), Portugal ($\mathcal{I} = 0.230$, $\tilde{t} = 9.49$, $\mathcal{S}(\tilde{t}) = 0.103$, $R^i(t_c) = 0.844$, $R(t_c) = 0.839$) and Romania ($\mathcal{I} = 0.265$, $\tilde{t} = 8.57$, $\mathcal{S}(\tilde{t}) = 0.019$, $R^i(t_c) = 1.107$, $R(t_c) = 1.060$).

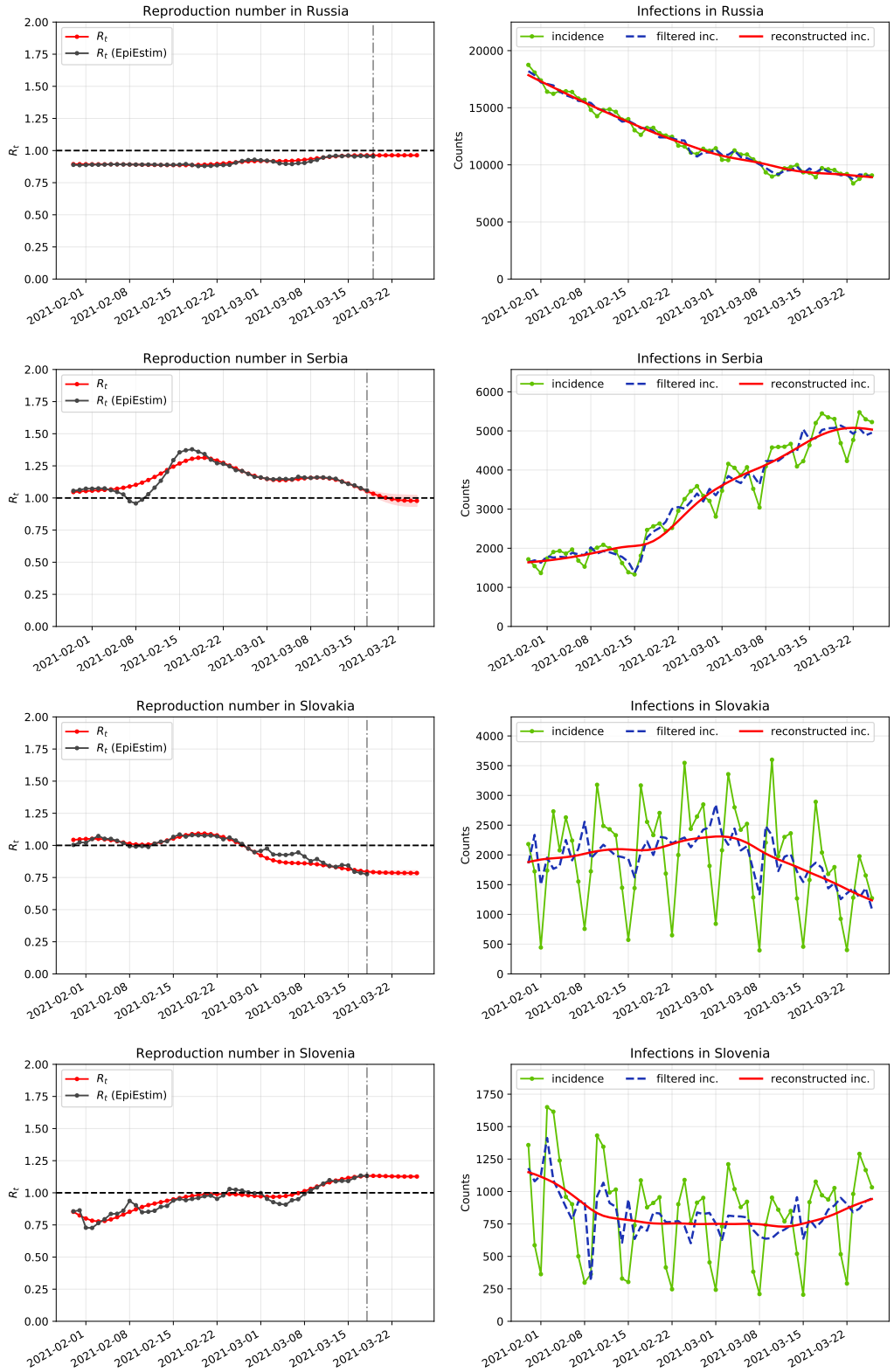

Figure S15: From top to down: Russia ($\mathcal{I} = 0.621$, $\tilde{t} = 7.25$, $\mathcal{S}(\tilde{t}) = 0.010$, $R^i(t_c) = 0.954$, $R(t_c) = 0.963$), Serbia ($\mathcal{I} = 0.548$, $\tilde{t} = 8.41$, $\mathcal{S}(\tilde{t}) = 0.043$, $R^i(t_c) = 1.059$, $R(t_c) = 0.978$), Slovakia ($\mathcal{I} = 0.304$, $\tilde{t} = 8.31$, $\mathcal{S}(\tilde{t}) = 0.029$, $R^i(t_c) = 0.776$, $R(t_c) = 0.785$) and Slovenia ($\mathcal{I} = 0.354$, $\tilde{t} = 7.88$, $\mathcal{S}(\tilde{t}) = 0.034$, $R^i(t_c) = 1.132$, $R(t_c) = 1.127$).

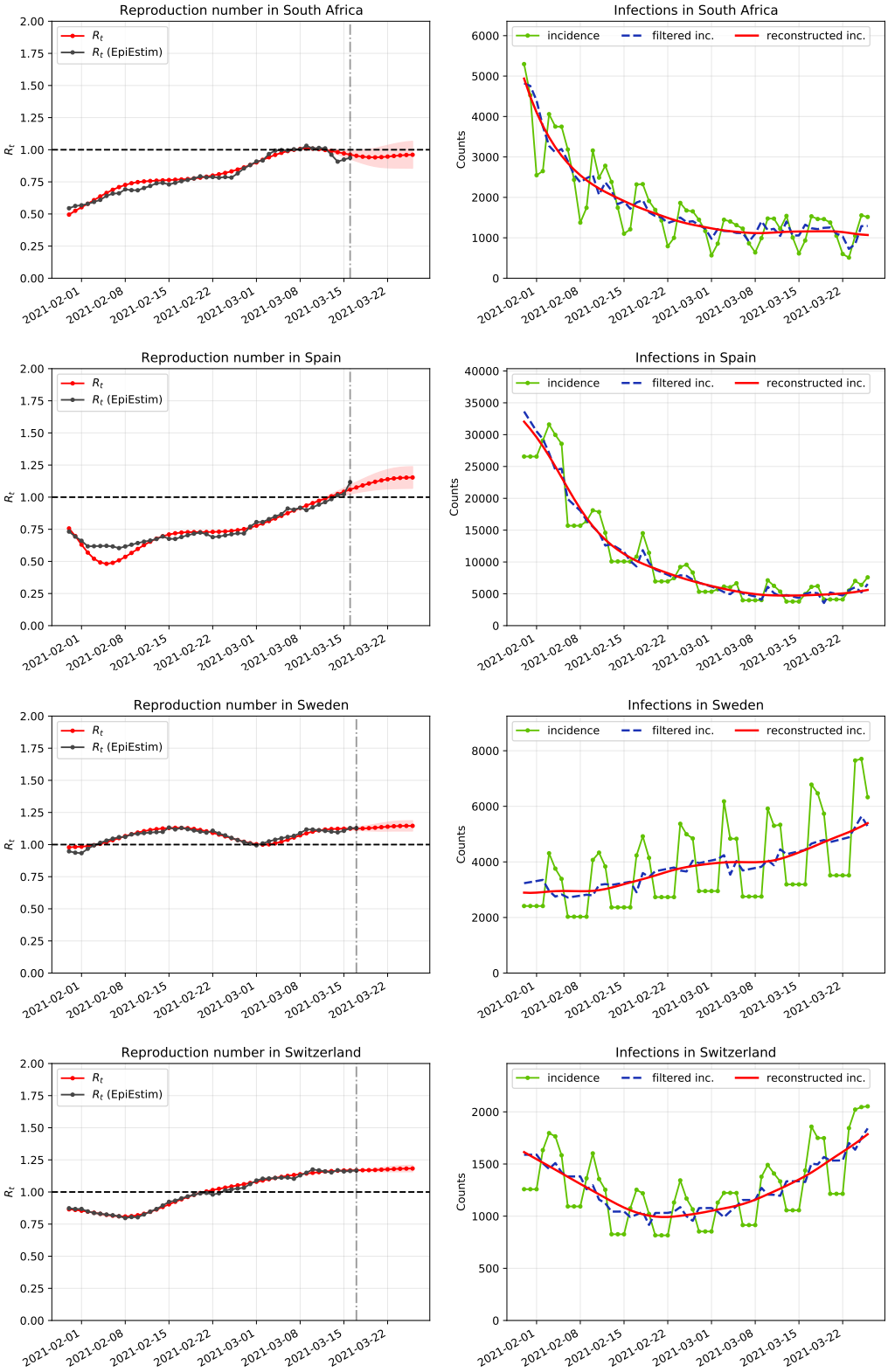

Figure S16: From top to down: South Africa ($\mathcal{I} = 0.299$, $\tilde{t} = 10.42$, $\mathcal{S}(\tilde{t}) = 0.055$, $R^i(t_c) = 0.937$, $R(t_c) = 0.961$), Spain ($\mathcal{I} = 0.290$, $\tilde{t} = 9.70$, $\mathcal{S}(\tilde{t}) = 0.046$, $R^i(t_c) = 1.117$, $R(t_c) = 1.153$), Sweden ($\mathcal{I} = 0.192$, $\tilde{t} = 9.45$, $\mathcal{S}(\tilde{t}) = 0.022$, $R^i(t_c) = 1.128$, $R(t_c) = 1.146$) and Switzerland ($\mathcal{I} = 0.206$, $\tilde{t} = 9.10$, $\mathcal{S}(\tilde{t}) = 0.016$, $R^i(t_c) = 1.167$, $R(t_c) = 1.183$).

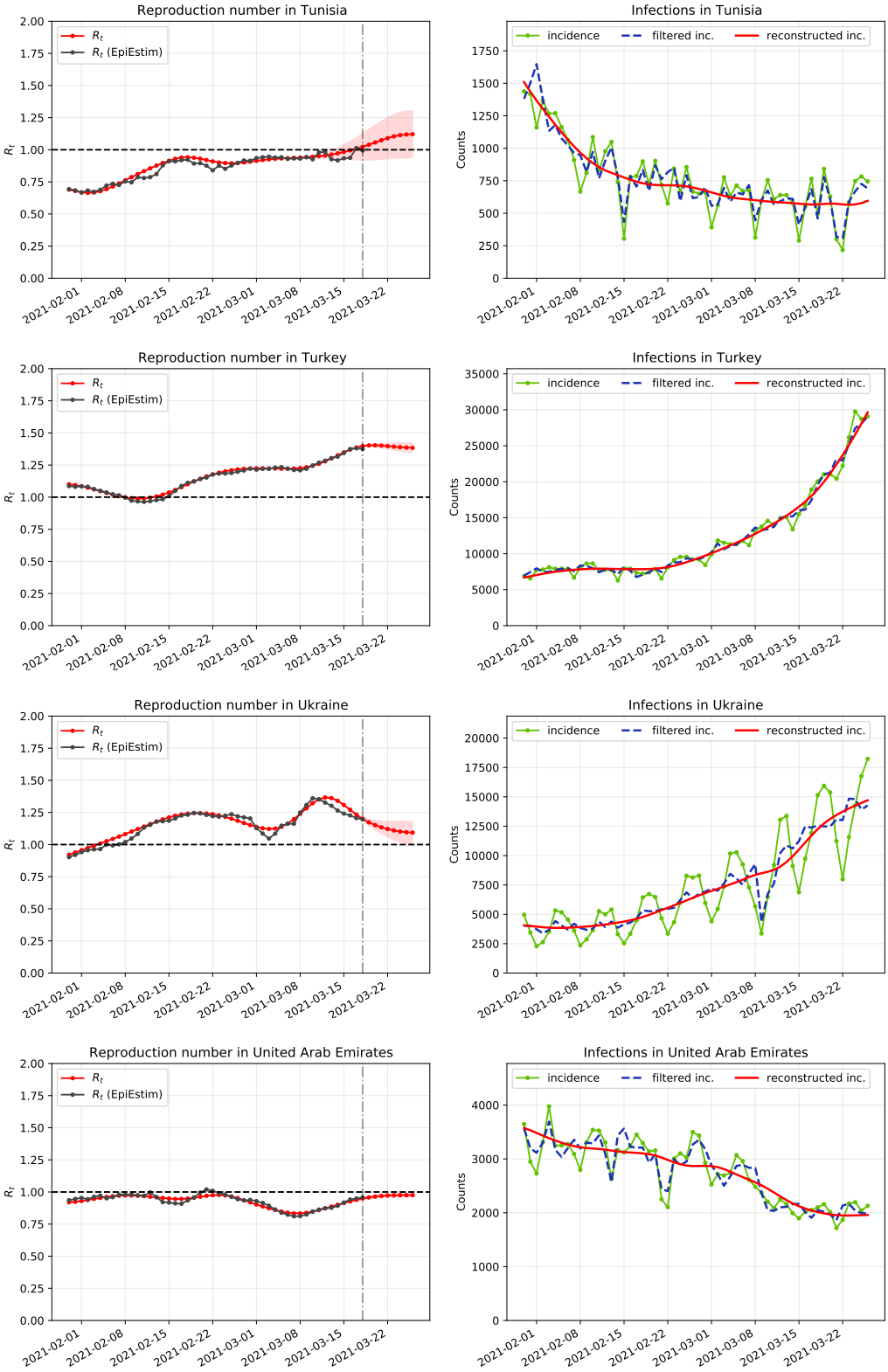

Figure S17: From top to down: Tunisia ($\mathcal{I} = 0.738$, $\tilde{t} = 8.19$, $\mathcal{S}(\tilde{t}) = 0.038$, $R^i(t_c) = 0.995$, $R(t_c) = 1.121$), Turkey ($\mathcal{I} = 0.554$, $\tilde{t} = 8.32$, $\mathcal{S}(\tilde{t}) = 0.013$, $R^i(t_c) = 1.375$, $R(t_c) = 1.384$), Ukraine ($\mathcal{I} = 0.384$, $\tilde{t} = 8.07$, $\mathcal{S}(\tilde{t}) = 0.034$, $R^i(t_c) = 1.197$, $R(t_c) = 1.093$) and United Arab Emirates ($\mathcal{I} = 0.773$, $\tilde{t} = 8.41$, $\mathcal{S}(\tilde{t}) = 0.021$, $R^i(t_c) = 0.957$, $R(t_c) = 0.975$).

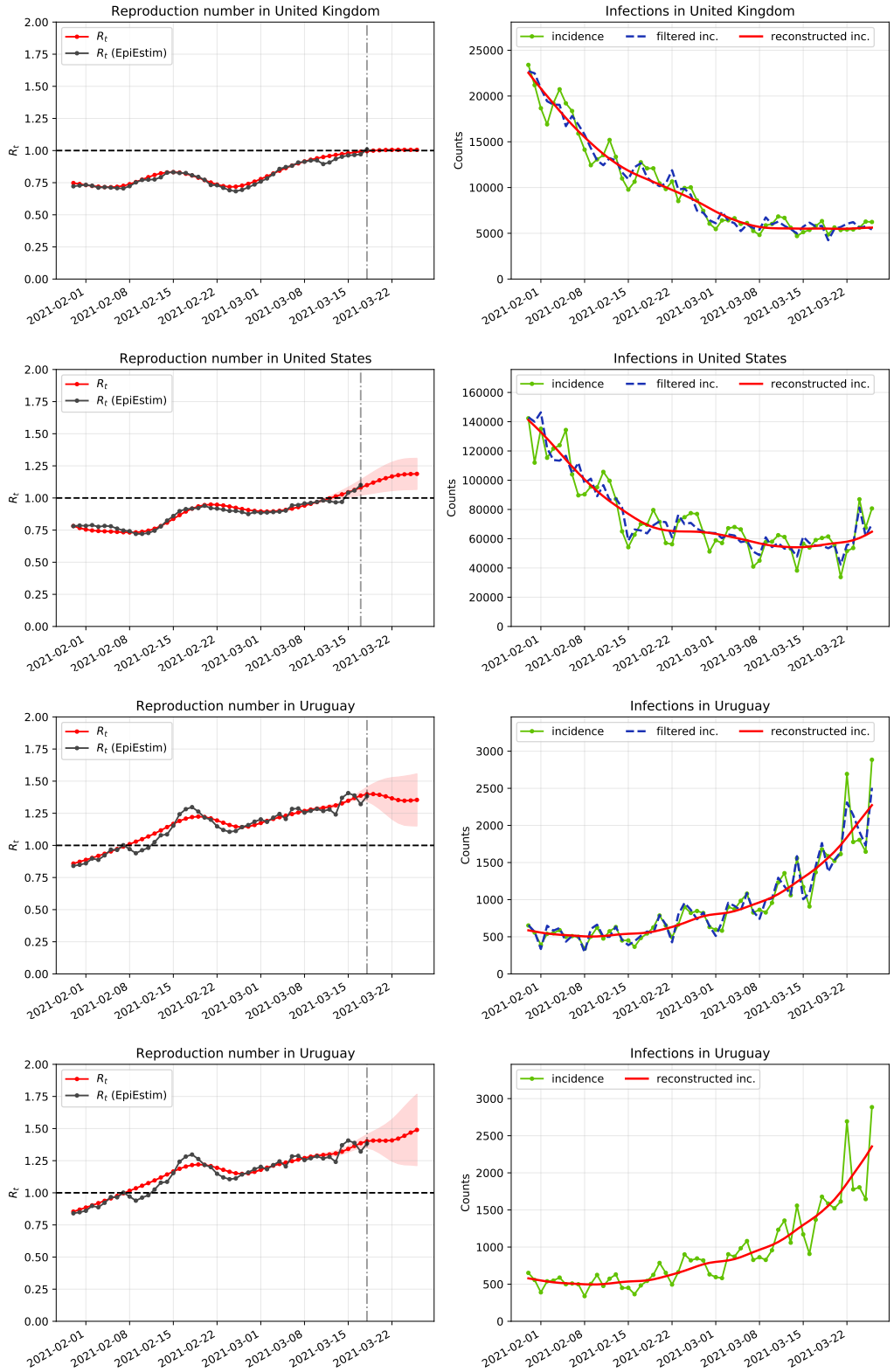

Figure S18: From top to down: United Kingdom ($\mathcal{I} = 0.557$, $\tilde{t} = 8.26$, $\mathcal{S}(\tilde{t}) = 0.024$, $R^i(t_c) = 1.009$, $R(t_c) = 1.006$), USA ($\mathcal{I} = 0.569$, $\tilde{t} = 8.76$, $\mathcal{S}(\tilde{t}) = 0.023$, $R^i(t_c) = 1.100$, $R(t_c) = 1.188$), Uruguay ($\mathcal{I} = 0.822$, $\tilde{t} = 8.26$, $\mathcal{S}(\tilde{t}) = 0.039$, $R^i(t_c) = 1.384$, $R(t_c) = 1.354$) and Uruguay ($\tilde{t} = 8.34$, $\mathcal{S}(\tilde{t}) = 0.041$, $R^i(t_c) = 1.384$, $R(t_c) = 1.489$).

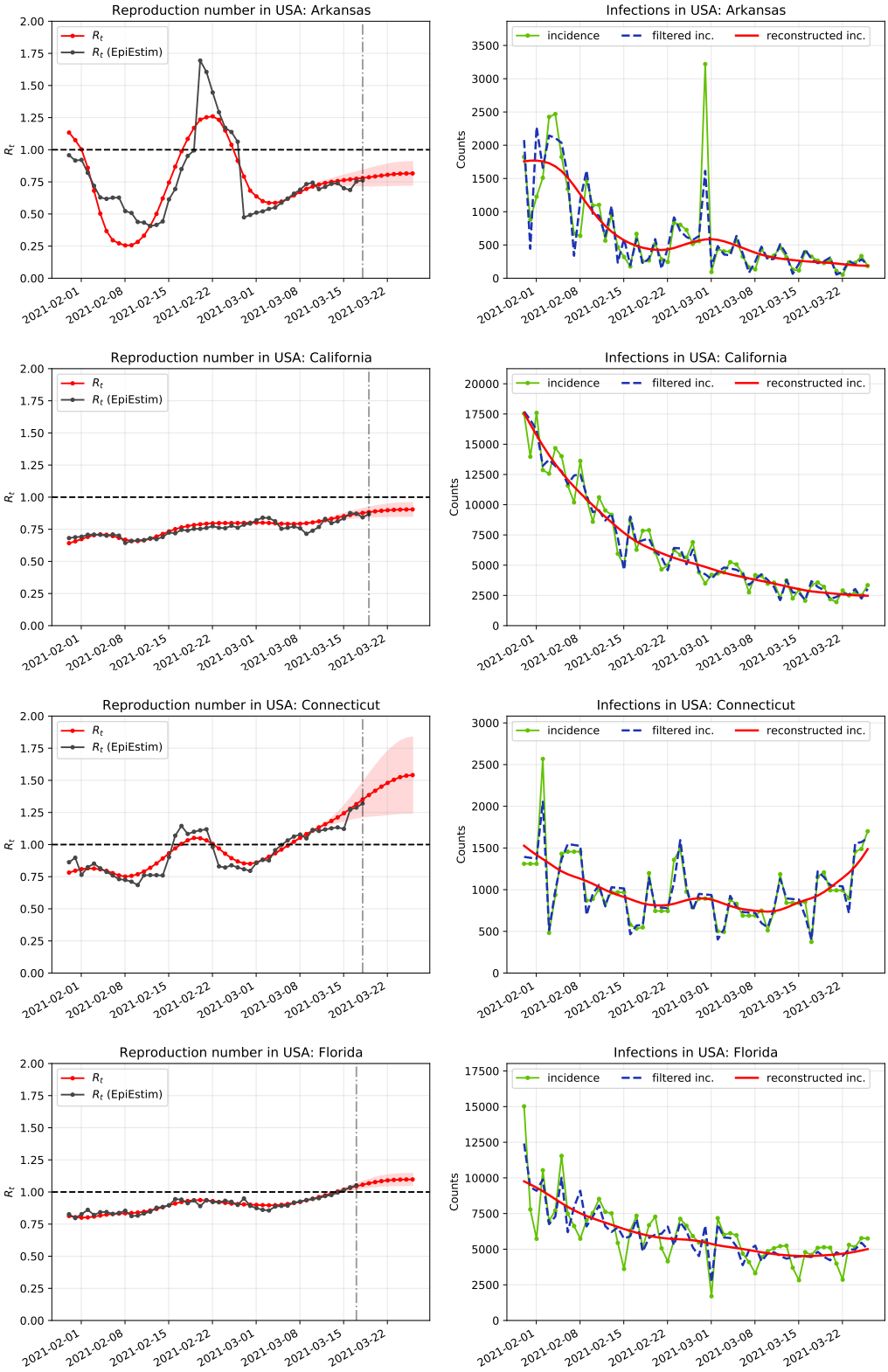

Figure S19: From top to down: Arkansas ($\mathcal{I} = 0.753$, $\tilde{t} = 8.46$, $\mathcal{S}(\tilde{t}) = 0.155$, $R^i(t_c) = 0.762$, $R(t_c) = 0.816$), California ($\mathcal{I} = 0.692$, $\tilde{t} = 6.72$, $\mathcal{S}(\tilde{t}) = 0.033$, $R^i(t_c) = 0.868$, $R(t_c) = 0.904$), Connecticut ($\mathcal{I} = 0.936$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.061$, $R^i(t_c) = 1.320$, $R(t_c) = 1.541$) and Florida ($\mathcal{I} = 0.606$, $\tilde{t} = 8.74$, $\mathcal{S}(\tilde{t}) = 0.021$, $R^i(t_c) = 1.052$, $R(t_c) = 1.098$).

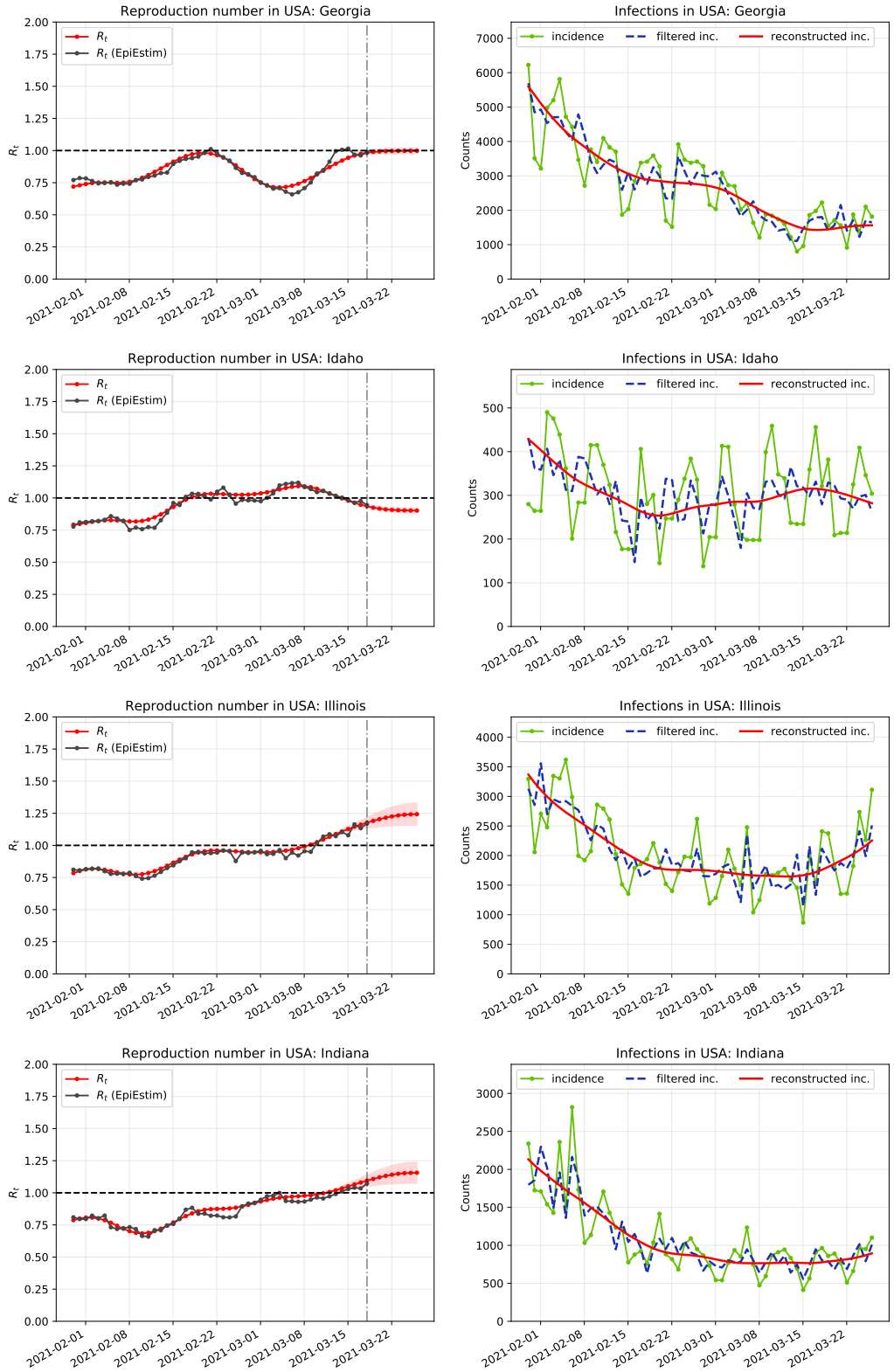

Figure S20: From top to down: Georgia ($\mathcal{I} = 0.463$, $\tilde{t} = 8.13$, $\mathcal{S}(\tilde{t}) = 0.035$, $R^i(t_c) = 0.988$, $R(t_c) = 0.998$), Idaho ($\mathcal{I} = 0.450$, $\tilde{t} = 8.02$, $\mathcal{S}(\tilde{t}) = 0.034$, $R^i(t_c) = 0.944$, $R(t_c) = 0.902$), Illinois ($\mathcal{I} = 0.516$, $\tilde{t} = 8.25$, $\mathcal{S}(\tilde{t}) = 0.026$, $R^i(t_c) = 1.170$, $R(t_c) = 1.243$) and Indiana ($\mathcal{I} = 0.586$, $\tilde{t} = 8.31$, $\mathcal{S}(\tilde{t}) = 0.031$, $R^i(t_c) = 1.073$, $R(t_c) = 1.156$).

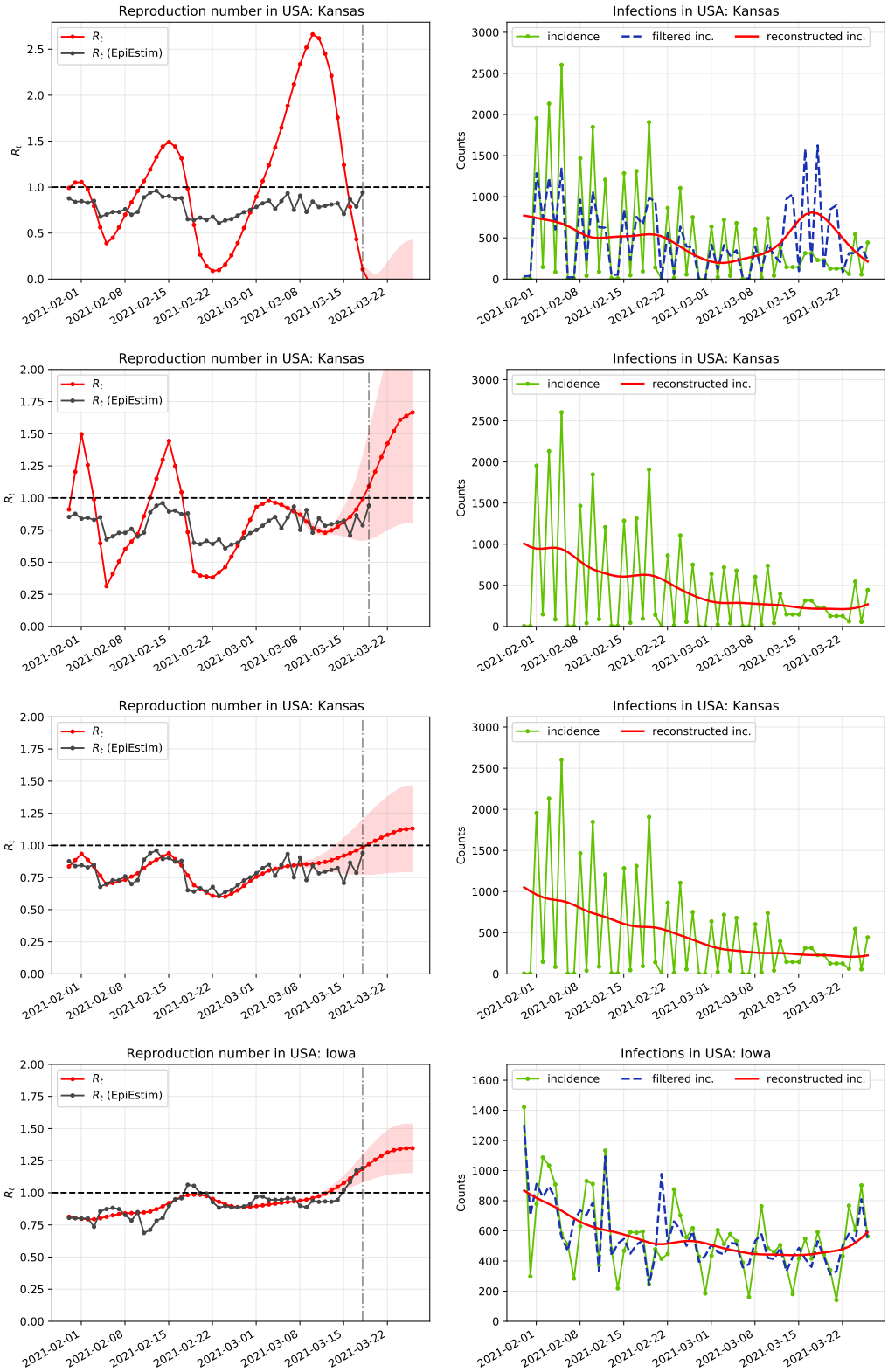

Figure S21: From top to down: Kansas ($\mathcal{I} = 0.595$, $\tilde{t} = 8.50$, $\mathcal{S}(\tilde{t}) = 0.692$, $R^i(t_c) = 0.940$, $R(t_c) = -0.022$, $\mathcal{V}(i) = 1.728$), Kansas ($\tilde{t} = 7.42$, $\mathcal{S}(\tilde{t}) = 0.205$, $R^i(t_c) = 0.940$, $R(t_c) = 1.666$), Kansas ($\tilde{t} = 7.59$, $\mathcal{S}(\tilde{t}) = 0.065$, $R^i(t_c) = 0.940$, $R(t_c) = 1.131$, $w = 40$) and Iowa ($\mathcal{I} = 0.707$, $\tilde{t} = 8.49$, $\mathcal{S}(\tilde{t}) = 0.055$, $R^i(t_c) = 1.195$. $R(t_c) = 1.347$)

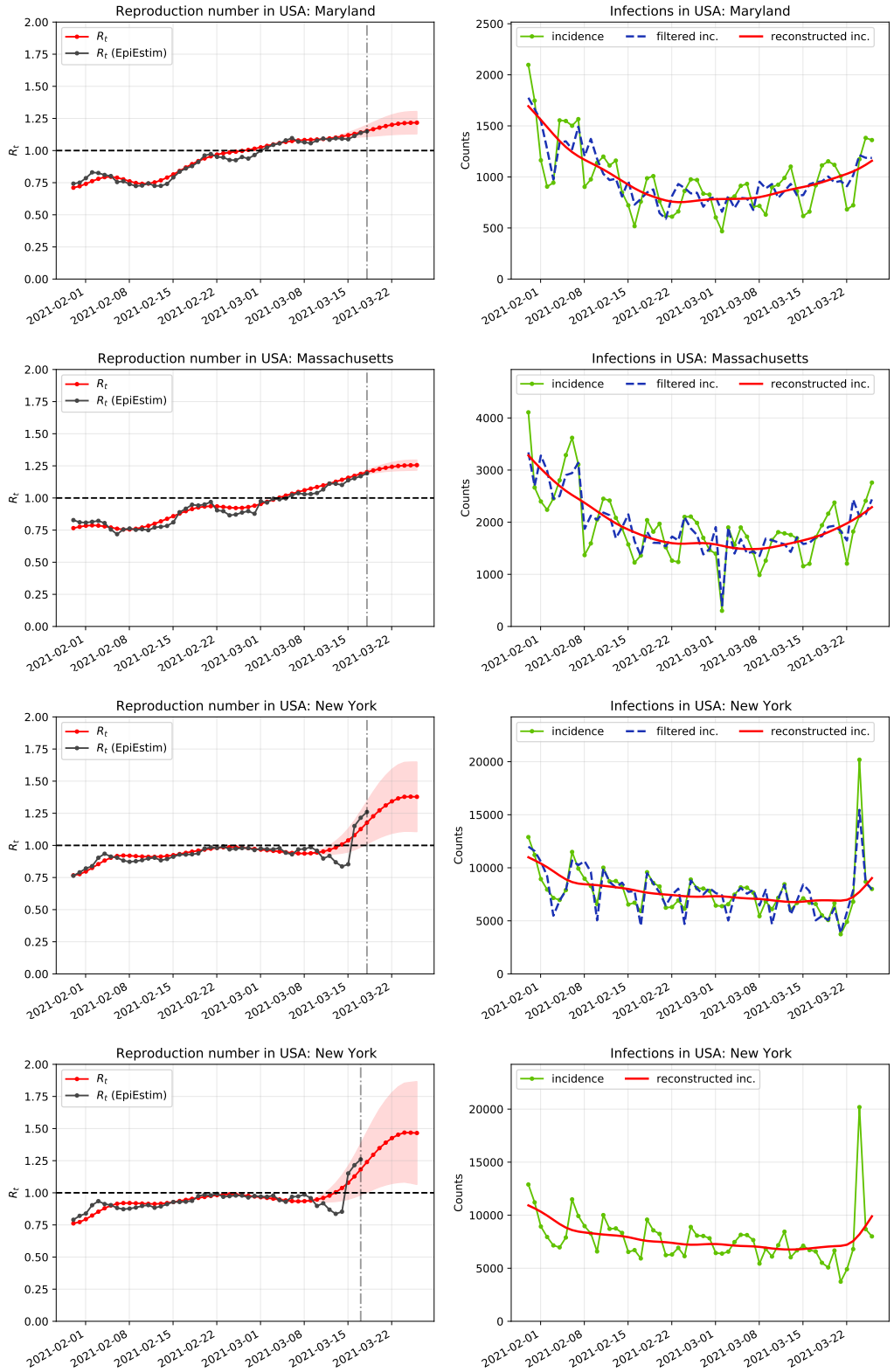

Figure S22: From top to down: Maryland ($\mathcal{I} = 0.525$, $\tilde{t} = 8.20$, $\mathcal{S}(\tilde{t}) = 0.028$, $R^i(t_c) = 1.150$, $R(t_c) = 1.217$), Massachusetts ($\mathcal{I} = 0.630$, $\tilde{t} = 8.35$, $\mathcal{S}(\tilde{t}) = 0.031$, $R^i(t_c) = 1.192$, $R(t_c) = 1.256$), New York ($\mathcal{I} = 0.860$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.048$, $R^i(t_c) = 1.260$, $R(t_c) = 1.378$) and New York ($\tilde{t} = 8.76$, $\mathcal{S}(\tilde{t}) = 0.049$, $R^i(t_c) = 1.260$, $R(t_c) = 1.465$).

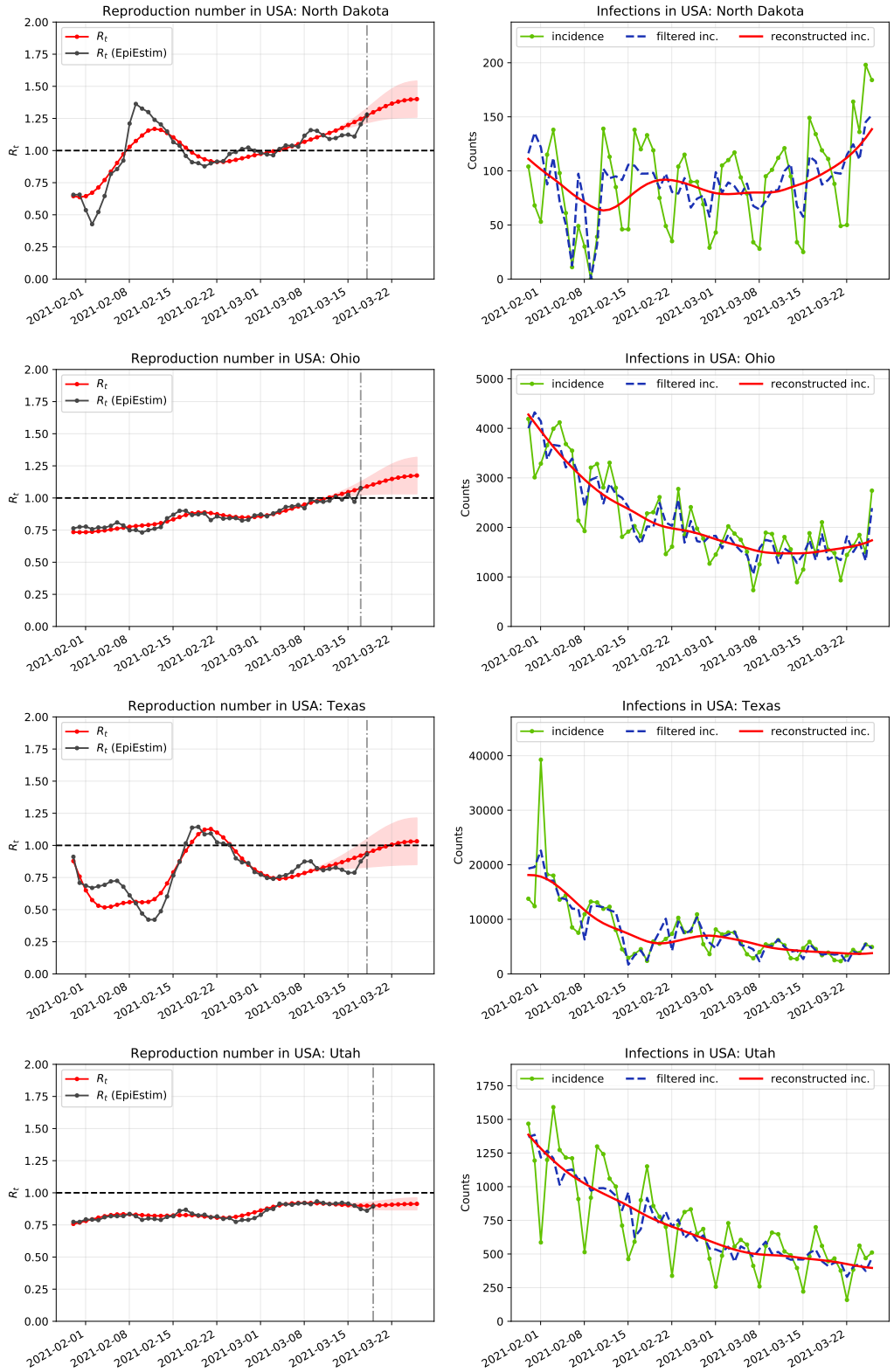

Figure S23: From top to down: North Dakota ($\mathcal{I} = 0.537$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.086$, $R^i(t_c) = 1.279$, $R(t_c) = 1.401$), Ohio ($\mathcal{I} = 0.542$, $\tilde{t} = 8.55$, $\mathcal{S}(\tilde{t}) = 0.032$, $R^i(t_c) = 1.077$, $R(t_c) = 1.175$), Texas ($\mathcal{I} = 0.576$, $\tilde{t} = 7.93$, $\mathcal{S}(\tilde{t}) = 0.081$, $R^i(t_c) = 0.931$, $R(t_c) = 1.032$) and Utah ($\mathcal{I} = 0.292$, $\tilde{t} = 7.06$, $\mathcal{S}(\tilde{t}) = 0.023$, $R^i(t_c) = 0.895$, $R(t_c) = 0.914$).

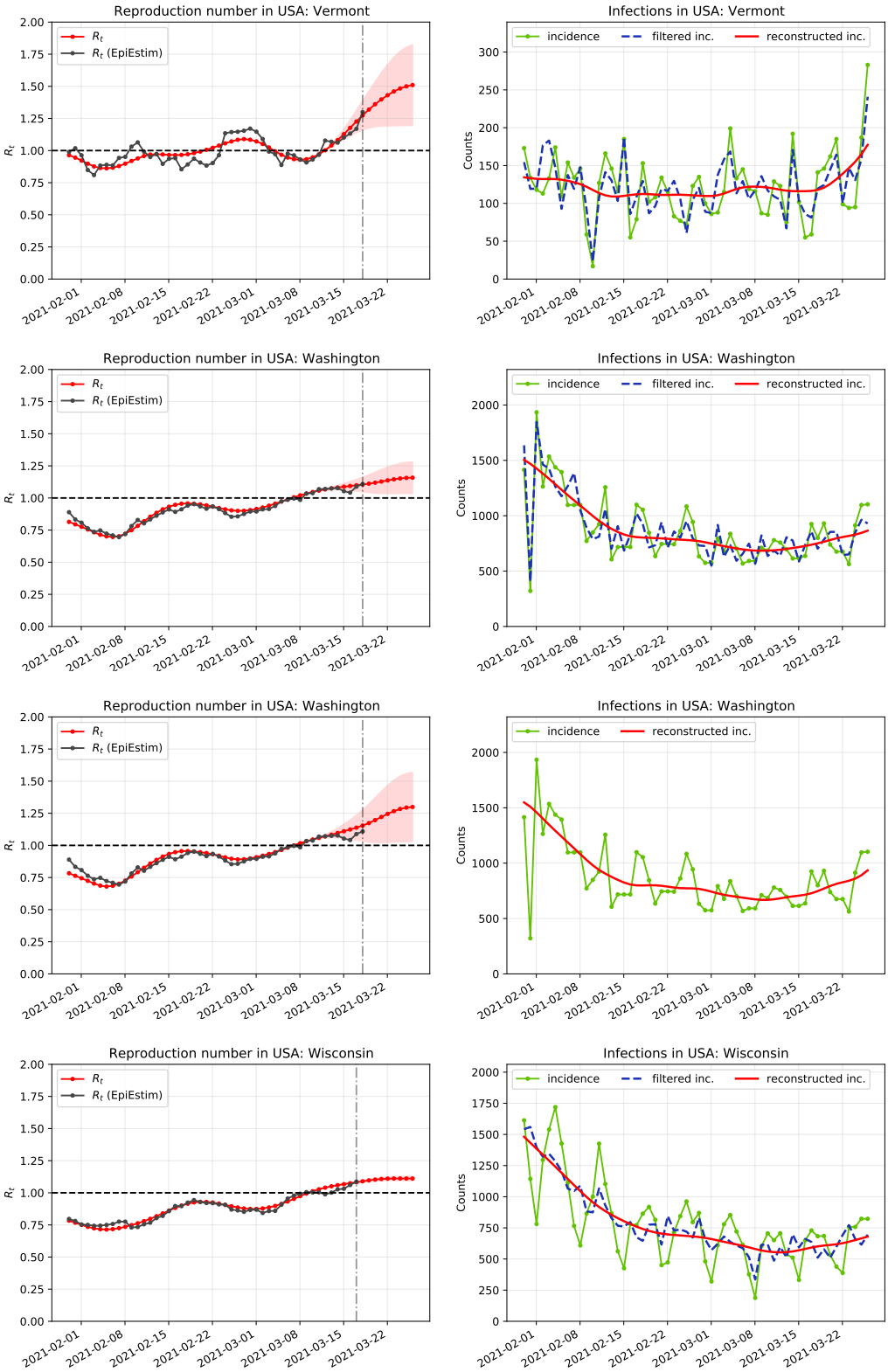

Figure S24: From top to down: Vermont ($\mathcal{I} = 0.760$, $\tilde{t} = 8.00$, $\mathcal{S}(\tilde{t}) = 0.057$, $R^i(t_c) = 1.300$, $R(t_c) = 1.510$), Washington ($\mathcal{I} = 0.816$, $\tilde{t} = 8.23$, $\mathcal{S}(\tilde{t}) = 0.027$, $R^i(t_c) = 1.109$, $R(t_c) = 1.158$), Washington ($\tilde{t} = 8.47$, $\mathcal{S}(\tilde{t}) = 0.037$, $R^i(t_c) = 1.109$, $R(t_c) = 1.299$) and Wisconsin ($\mathcal{I} = 0.344$, $\tilde{t} = 9.29$, $\mathcal{S}(\tilde{t}) = 0.027$, $R^i(t_c) = 1.086$, $R(t_c) = 1.111$).