medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Modeling Covid-19 incidence by the renewal equation after removal of administrative bias and noise

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

1 Departamento de Informática y Sistemas, Universidad de Las Palmas de G.C., Las Palmas de G.C., 35017, Spain; lalvarez@ulpgc.es
2 Laboratory of Integrative Systems Physiology, Ecole Polytechnique Fédérale de Lausanne, EPFL/IBI/LISP - Station 15, Lausanne, CH-1015, Switzerland ; jean-david.morel@epfl.ch
3 ENS Paris-Saclay, CNRS, Centre Borelli, Université Paris-Saclay, Gif-sur-Yvette, F-91190, France ; jean-michel.morel@ens-paris-saclay.fr
* Correspondence: lalvarez@ulpgc.es;

Simple Summary: In the past two years, the COVID-19 incidence curves and reproduction number R_t have been the main metrics used by policy makers and journalists to monitor the spread of this global pandemic. However, these metrics are not always reliable in the short term, because of a combination of delay in detection, administrative delays and random noise. In this article, we present a complete model of COVID-19 incidence, faithfully reconstructing the incidence curve and reproduction number from the renewal equation of the disease and precisely estimating the biases associated with periodic weekly bias, festive day bias and residual noise.

Abstract: The sanitary crisis of the past two years has focused the public's attention on quantitative indicators of the spread of the COVID-19 pandemic. The daily reproduction number R_t, defined by the average number of new infections caused by a single infected individual at time t, is one of the best metrics for estimating the epidemic trend. In this paper, we give a complete observation model for sampled epidemiological incidence signals obtained through periodic administrative measurements. The model is governed by the classic renewal equation using an empirical reproduction kernel, and subject to two perturbations: a time-varying gain with a weekly period and a white observation noise. We estimate this noise model and its parameters by extending a variational inversion of the model recovering its main driving variable R_t. Using R_t, a restored incidence curve, corrected of the weekly and festive day bias, can be deduced through the renewal equation. We verify experimentally on many countries that, once the weekly and festive days bias have been corrected, the difference between the incidence curve and its expected value is well approximated by an exponential distributed white noise multiplied by a power of the magnitude of the restored incidence curve.

Keywords: Incidence curve ; pandemic ; COVID-19 ; reproduction kernel ; time dependent reproduction number ; administrative noise ; exponential distribution ; renewal equation ; variational inversion method.

MSC: 92C60 ; 92C55 ; 45Q05 ; 65K10

---

1. Introduction

The renewal equation, first formulated for birth-death processes in a 1907 note of Alfred Lotka [1], establishes a model for epidemic propagation based on the individual infectiousness. The infectiousness of individuals at time t is characterized by the reproduction number R_t, defined as the average number of cases generated by an infected person at time t, and by the generation time, [2,3],

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

defined as the probability distribution of the time between infection of a primary case and infections in secondary cases. This probability distribution depends on the incubation time (a permanent biological factor) and on the detection time (which we assume stationary). For these reasons, the distribution of the generation time is supposed to be independent of t. In practice, the generation time is replaced by the observable serial interval Φ_s which represents the time distribution of the delay of the onset of symptoms between primary and secondary cases. In Fig. 1, we show the serial interval obtained in [4] using 689 observed pairs of primary and secondary cases.

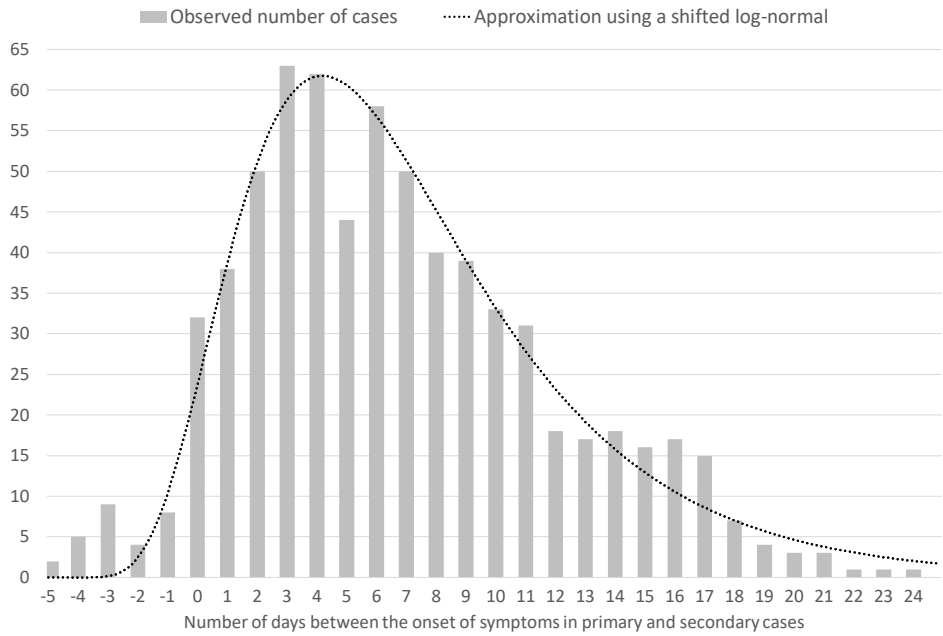

Figure 1. The serial interval Φ_s obtained by [4]. The bars represent the observed number of cases in function of the number of days between the onset of symptoms in primary and secondary cases. The dotted line is its approximation by a scaled and shifted log-normal distribution.

The case renewal equation [5,6] is a classic equation linking R_t, Φ and the incidence i_t of new daily cases,

$$i_t = \sum_s i_{t-s} R_{t-s} \Phi_s \quad \text{for} \;\; t = 0, .., t_c, \tag{1}$$

where t_c is the current time. This equation does not account for several strong perturbations of i_t. Government statistics of the observed incidence curve are indeed affected by changes in testing and polling policies and by weekend reporting delays. These recording delays and subsequent rash corrections result in impulse noise, and in a strong weekly periodic bias observable on the observed incidence curve i_t^0. In [3] this bias is corrected by a seven days sliding average and in [7] it is corrected by multiplying i_t^0 by a 7-day periodic factor q_t. These *bias correcting coefficients* q_t are learned by a variational method that we describe below. Our first purpose in this note is to resolve the festive day problem. We denote by $\mathbf{F}$ the set of festive days t, at which the i_t^0 curve is strongly affected by the reduction in the number of registered cases. This reduction is compensated by an increase in the number of registered cases the following days. No model has been proposed so far to address this problem, which creates strong impulse noise in any estimation of i_t and R_t. We tackle this problem by a variational method computing R_t, where both i_t and R_t are considered unknown on festive days and in the next few days. To that purpose, we shall denote by $\mathbf{F}_+$ the union of festive days and the ones following them affected by the festive day (typically 2 or 3 days after the festive day).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Our second purpose is to give a noise model for the difference $\hat{i}_t - i_t^r$ between the signal $\hat{i}_t$ corrected of the week-end and festive effects, and its restored version i_t^r using the renewal equation, defined by

$$i_t^r = \sum_s \hat{i}_{t-s} R_{t-s} \Phi_s. \tag{2}$$

We give strong experimental evidence that the relation between $\hat{i}_t$ and i_t^r, can be empirically modeled by

$$\hat{i}_t = i_t^r + \varepsilon_t (i_t^r)^a, \tag{3}$$

where $a > 0$ and ε_t is a white noise.

This leads us to propose a signal processing version of the renewal equation model taking into account noise and bias and justifying *a posteriori* the variational method. The proposed observation model linking the observed signal i_t^0 to the ground truth incidence i_t is

$$q_t i_t^0 = i_t + \varepsilon_t (i_t)^a \quad \text{for } t \in [0, t_c] \setminus \mathbf{F}_+, \tag{4}$$

where q_t is a quasi-periodic gain with period 7, ε_t is a white noise. The exponent a can be estimated for each country and varies between 0.6 and 0.9. The exceptional set $\mathbf{F}_+$ is introduced because festive days provoke perturbations of the observation model (4). Specifically, the 7 days period of q_t is broken for these groups of days.

We shall verify experimentally on 38 countries (and detail the results on USA, France and Germany) that the normalized error ε_t is indeed a white noise with a distribution that is well described by an exponential distribution. This *a posteriori* noise model contradicts the classic *a priori* stochastic formulation of the renewal equation where the first member i_t of equation (1) is assumed to be a Poisson variable, and the second member of this equation is interpreted as the expectation of this Poisson variable. Using this Poisson model leads to maximum likelihood estimation strategies to compute R_t [3,8–10]. As we shall see, the Poisson model is not verified. Indeed, as we mentioned, the empirically observed standard deviation of the noise follows a power law with exponent a significantly larger than 0.5, which is incompatible with the Poisson model.

The proposed observation model (4) of the pandemic's incidence curve gives a simple framework enabling:

1. a computation of the reproduction number R_t;
2. a correction of the weekend and festive days bias on i_t;
3. a verification that the difference between the observed incidence curve after bias correction and its expected value using the renewal equation is a white noise, the parameters of which can be estimated.

Plan

In section 2 we describe an anterior variational method [7] and point out its main three limitations: its weekly bias correction is strongly periodic, which does not work on long periods; the festive days cause strong perturbations in the inversion, finally no residual noise model is proposed. We therefore modify its variational formulation. In section 3 we present the results of the statistical analysis of the residual noise on many countries. These examples lead to specify the noise model and to validate *a posteriori* the proposed inversion model. In section 4 we organize and present all previous R_t and i_t analysis methods, particularly those using an *a priori* noise model. Section 5 is a final discussion.

Timely estimates of restored versions of i_t and R_t are extremely useful to tame a pandemic. The proposed restoration and inversion algorithm can be run through an online demo [11] for every day in every country and U.S. state. The demo plots the objects of this paper, namely the incidence curve i_t^0, its bias corrected version $\hat{i}_t$, its fully restored version i_t^r, finally the main pandemic index, the time-dependent reproduction number R_t. Fig. 2 illustrates the application of the variational method of section 2 to USA on February 1st, 2022, as displayed by the online demo. Fig. 3 compares the results of

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

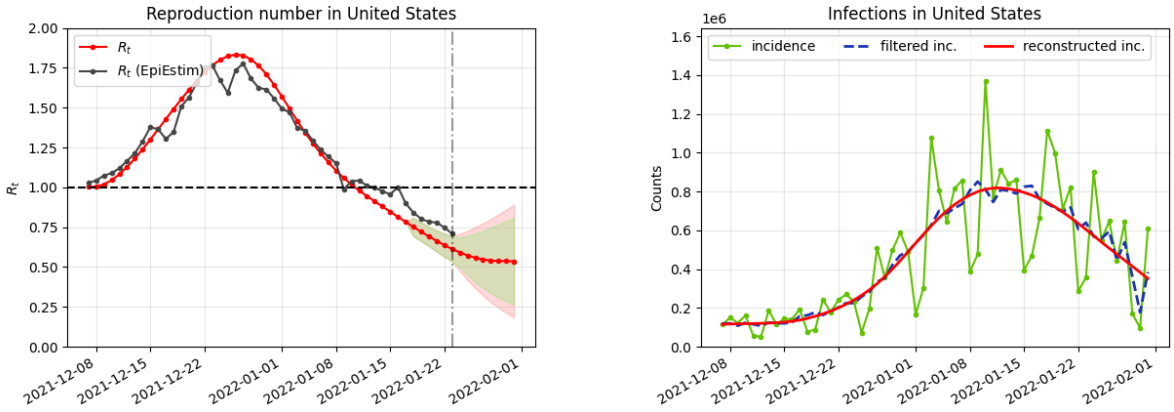

Figure 2. Illustration of the online inversion method [11]. On the left in red, the obtained reproduction number R_t and in black its estimate obtained by the classic EpiEstim method. On the right in green, the original incidence curve i_t of new cases, in blue the incidence curve $\hat{i}_t$ corrected of the weekend and festive biases, and in red the final reconstructed incidence curve i_t^r obtained from R_t by the application of the renewal equation. Estimate obtained for USA on February 1st, 2022.

this inversion method, applied with and without festive day bias correction, obtained for France on January 6, 2022.

2. The proposed variational model

The EpiInvert method proposed in [7] is a deconvolution + denoising procedure to solve the functional equation (1) using the Tikhonov-Arsenin [12], [13] variational approach. EpiInvert estimates both R_t and a restored i_t corrected for the weekend bias. To remove the weekend effect, it computes a 7-day periodic multiplicative factor $q = (q_0, q_1, q_2, q_3, q_4, q_5, q_6)$. From the observed incidence curve and the serial interval, R_t and q are jointly estimated by minimizing

$$E(R,q) = \sum_{t=0}^{t_c} \left(\frac{q_{t\%7} i_t^0 - \sum_s R_{t-s} i_{t-s}^0 q_{(t-s)\%7} \Phi_s}{median_{(t-\tau,t]}(i^0)} \right)^2 + w \sum_{t=1}^{t_c} (R_t - R_{t-1})^2 \tag{5}$$

where $t\%7$ denotes the remainder of the Euclidean division of t by 7 and $median_{(t-\tau,t]}(i^0)$ is the median of i_t^0 in the interval $(t-\tau, t]$ used to normalize the energy with respect to the size of i_t. The total number of cases is preserved by adding to (5) the constraint on q_t :

$$\sum_{t=t_c-T+1}^{t_c} i_t^0 = \sum_{t=t_c-T+1}^{t_c} q_{t\%7} i_t^0, \tag{6}$$

where T is a period of analysis empirically fixed to $T = 56$ days. The minimization of the above energy yields estimates of R_t, q and a restored incidence curve.

One limitation of using a 7-day periodic formulation to model the weekend effect is that it does not take into account the variation over time of the seasonal profile. To deal with this issue, we consider q_t for $t = 0, .., t_c$ allowing different correction factors q_t for every day but keeping the values $q_t - q_{t-7}$ small which forces q_t to be quasi-periodic. A regularity assumption for the seasonality is commonly used in the study of time series as it is the case of the standard Holt-Winters' seasonal method [14].

In addition to the weekend bias, festive days can introduce a strong bias in the incidence values. On a festive day $t \in \mathbf{F}$, a sharp decrease in the number of registered incident cases is generally observed. This is compensated by increased incidence numbers in the next few days. Assuming that each festive day, $t \in \mathbf{F}$, mainly affects the value of the incidence curve in the festive day and in the next M_t days (where M_t is an algorithm parameter (by default we fix $M_t = \min\{2, t_c - t\}$)), we consider the values

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

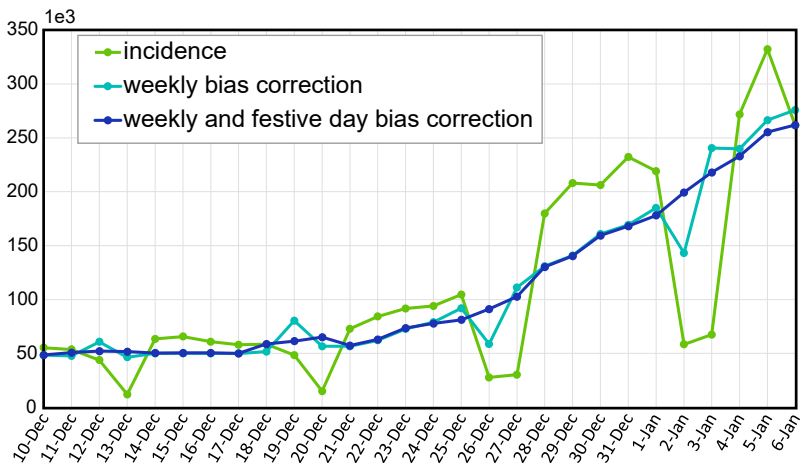

Figure 3. Incidence curve (in green) of France up to January 6, 2022. In cyan, the incidence corrected of the weekly bias, in blue the incidence corrected of the weekly and festive day. The Christmas holidays introduce a distortion in the weekly bias corrected incidence that is corrected by the festive day bias correction.

of $i^0_t, i^0_{t+1}, .., i^0_{t+M_t}$ as unknown. We denote by $\mathbf{F}_+$ the union of the festive days $t \in \mathbf{F}$ and the M_t days following them. We set $i^f_t = i^0_t$ for $t \notin \mathbf{F}_+$ and consider the values $(i^f_t)_{t\in\mathbf{F}_+}$ as unknowns. Then the new proposed inversion functional is

$$E(R,q,(i^f_t)_{t\in\mathbf{F}_+}) = \sum_{t=0}^{t_c}\left(\frac{q_t i^f_t - \sum_s R_{t-s} i^f_{t-s} q_{t-s}\Phi_s}{median_{(t-\tau,t]}(i^0)}\right)^2 + w_R\sum_{t=1}^{t_c}(R_t - R_{t-1})^2 +$$
$$\sum_{t\in\mathbf{F}}\lambda_t\left(\frac{\sum_{k=0}^{M_t} i^f_{t+k} - \sum_{k=0}^{M_t} i^0_{t+k}}{median_{(t-\tau,t]}(i^0)}\right)^2 + w_q\sum_{t=7}^{t_c}(q_t - q_{t-7})^2, \tag{7}$$

The values i^f_t for $t \in \mathbf{F}_+$ are set free in the minimization. Yet the third term in the functional ensures that the overall number of cases in the affected days remains unchanged. For each $t \in \mathbf{F}$, $\lambda_t \geq 0$ represents the weight we assign to this constraint for each festive day. We fix, experimentally, $\lambda_t = 2^{t_c - t - 2}$ if $t_c > t$ and $\lambda_t = 0$ if $t_c = t$. In other terms, the value of λ_t is adjusted according to the number of days that have passed since the festive day. To keep a smooth seasonality we add to the energy a regularization term where we penalize high values of $q_t - q_{t-7}$. The parameters w_R and w_q are regularization weights with default values $w_R = w_q = 5$. Their values are proven in [7] to be nearly optimal for Covid-19 incidence curves.

By minimizing this energy we obtain the reproduction number R_t, the seasonality q_t and i^f_t, which corresponds to the original incidence i^0_t but with the optimized values in the festive days. The bias corrected incidence $\hat{i}_t$ defined in model (3) is given by $\hat{i}_t = q_t i^f_t$.

The estimated incidence curve must preserve the number of cases. In the original EpiInvert formulation this constraint is enforced by (6) on its analysis interval $(t_c - T, T]$. In the new formulation, the interval time of analysis is the whole time interval $[0, t_c]$. Extra conditions are required to keep i^0_t close to $\hat{i}_t$ and i^r_t. Therefore, to preserve the number of cases we add to the energy (7) the constraints on q_t:

$$\sum_{t=0}^{t_c} i^f_t = \sum_{t=0}^{t_c} q_t i^f_t \;\; ; \qquad \sum_{t_c-14(k+1)}^{t_c-14k} i^f_t = \sum_{t_c-14(k+1)}^{t_c-14k} q_t i^f_t \qquad \text{for } k = 0,1,2,... \tag{8}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The first constraint corresponds to a global preservation of the number of cases in the whole period and the second one corresponds to a local preservation of the number of cases every 2 weeks. In particular, the second constraint ensures a good agreement between the epidemiological indicator given by the accumulated number of cases in the last 14 days of the original incidence curve and the estimated ones using the proposed method. This indicator is currently widely used to evaluate the current epidemic transmission.

The minimization of the energy (7) is obtained by alternating steps computing in turn R_t, q_t, and then i_t^f (for $t \in \mathbf{F}_+$) until convergence. The above constraints are added to the minimization by the Lagrange multiplier technique.

3. Results

We used the incidence data published in [15] for France, [16] for Germany, [17] for Spain and [18] for the rest of countries. We checked the observation model and its inversion on the 626 daily incidence data from March 24, 2020 to December 9, 2021 for 38 countries and will detail the results for France, Germany, and the USA. In general, for the festive days we fixed $M_t = 2$, so the method estimated the incidence value of the festive day and of the next 2 days. However, not all festive days disturb the incidence in the same way. Parameter M_t allows us to adapt the number of days affected. To illustrate this option we set $M_t = 5$ for Thanksgiving in the USA in 2021 because this festive day causes in 2021 a longer perturbation in the number of registered cases. Figs. 5,A1,A2 show the minimization results for the energy (7). They display for each country (i) the original incidence curve i_t^0, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve i_t^r using the renewal equation (1), (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error defined by

$$\varepsilon_t = \frac{\hat{i}_t - i_t^r}{(i_t^r)^a}. \tag{9}$$

The power a was obtained through log-log linear regression. Indeed, if $|\hat{i}_t - i_t^r|$ is proportional to $(i_t^r)^a$, then $\log(|\hat{i}_t - i_t^r|) \approx a \cdot \log(i_t^r) + b$, and a and b can be estimated by a linear regression between $\log(|\hat{i}_t - i_t^r|)$ and $\log(i_t^r)$. Its results are illustrated for 38 countries in fig. 4 and table A2. The Pearson correlation p-values in this table confirm the linear relation. The estimated exponent a varies between 0.7 and 0.9, and the constant coefficient b varies between -0.11 and -2.6. For the world we have $a = 0.76$ and $b = -1.16$.

We performed a control test on a Brownian motion simulated by starting from 10000 and sampling $i_{t+1} - i_t \simeq \mathcal{N}(0, 100)$. The obtained exponent a is negative ($a = -1.01$) and we have $b = 13.4$. Both values are far away from the group of coefficients of real incidence curves. The p-value for the control is anyway non significant (0.0844), compared to the extremely small p-values for the real incidence curves. Fig. A3 shows the results of the variational inversion method on the Brownian control. For this control, both R_t and the weekly seasonality correction coefficients stay very close to 1 as should be expected, with means 1.001 and 1.00002, and standard deviations 1.7% and 0.3% respectively.

Next, we looked for a stochastic model of the normalized error ε_t defined by (9). Figs. 5-A2 visually support a stationarity assumption for ε_t in France, Germany and USA. In fig. 6 we show the autocorrelation function for these three countries. For most non-zero shifts, its value stays inside the 95% confidence interval for the stationarity assumption. (This interval is indicated by horizontal blue lines in the plot.) Similar results were obtained on 33 more countries, as illustrated in fig. A5. These results support a white noise assumption for ε_t.

We finally estimated the parameters of the distribution of ε_t assuming an exponential power distribution with density

$$\frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^\beta}, \tag{10}$$

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

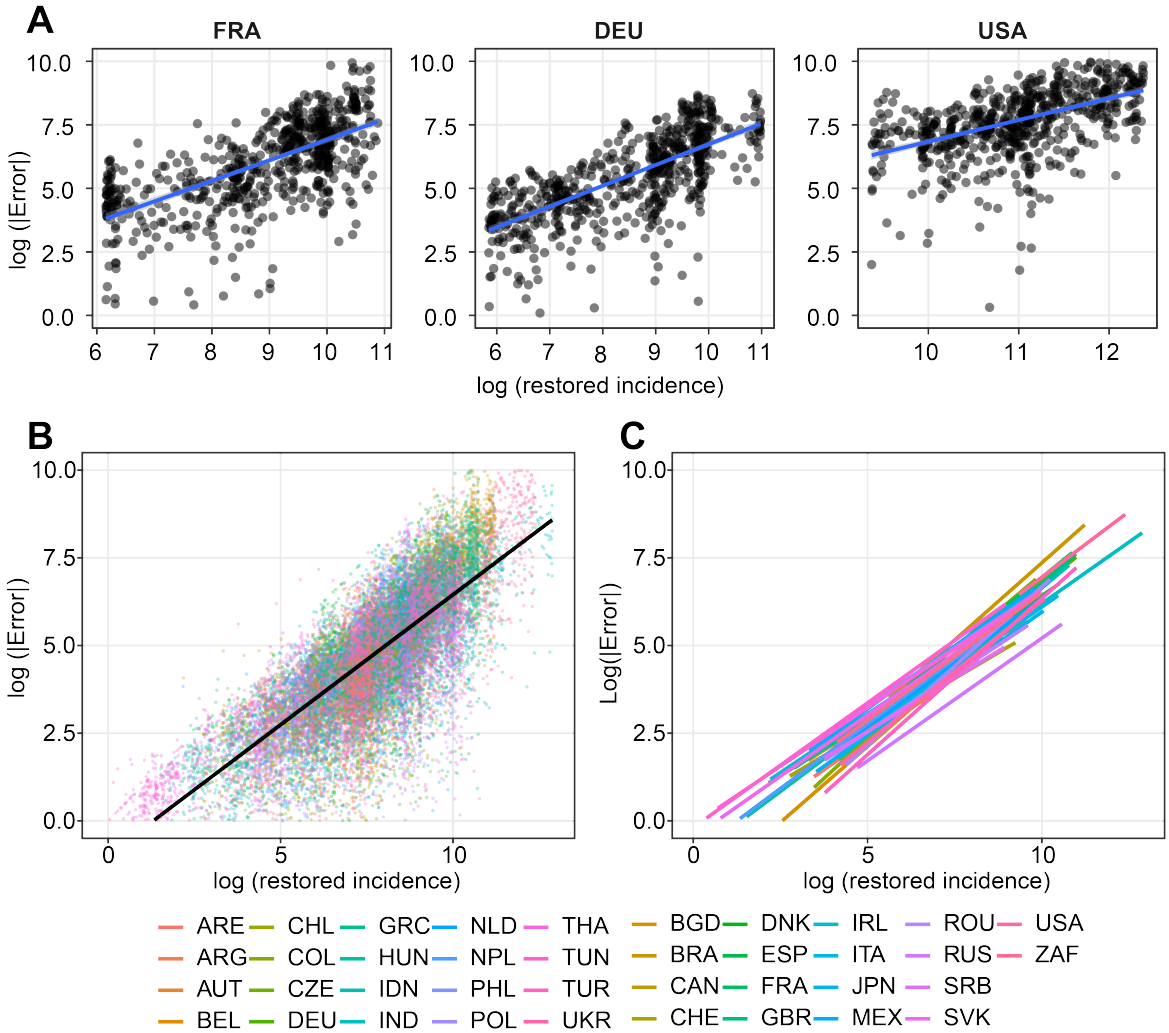

Figure 4. Worldwide log-log correlations between restored incidence i_t^r and the residual $|\hat{i}_t - i_t^r|$ (defined as restored incidence - bias-corrected incidence). The plot presents the log(error) as a function of the log(incidence). The regression parameters were computed through robust linear regression by the R package MASS. A: Correlation in France, Germany, and USA, with festive day correction. B: Spread of the values for 38 countries, without festive corrections. C: Robust linear regression curves for all countries. The linear regression coefficients a and b can be found in table A2. The worldwide coefficients are $a = 0.76$ and $b = -1.16$.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

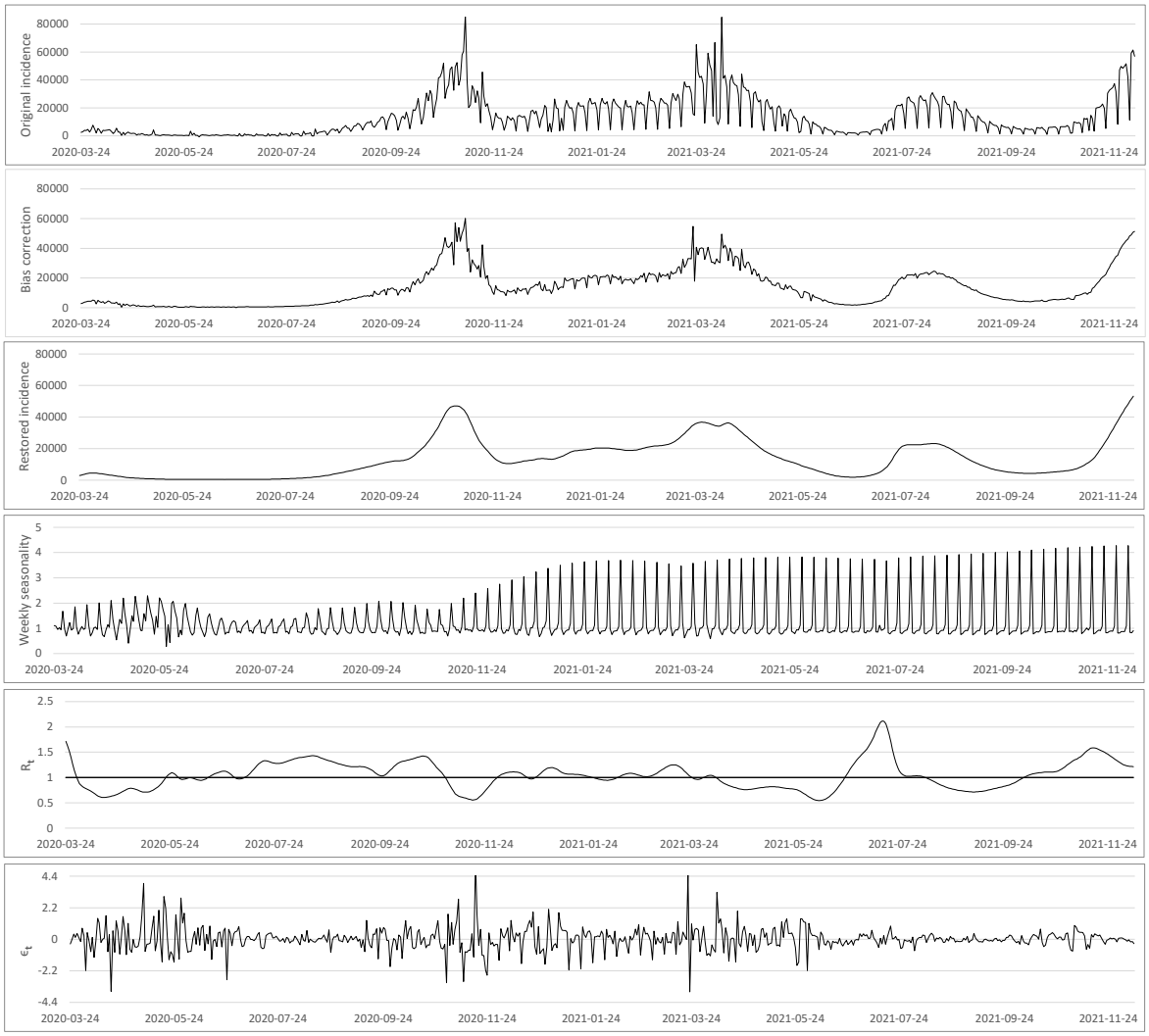

Figure 5. From top to bottom : (i) the original incidence curve i_t^0 of France, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table A2).

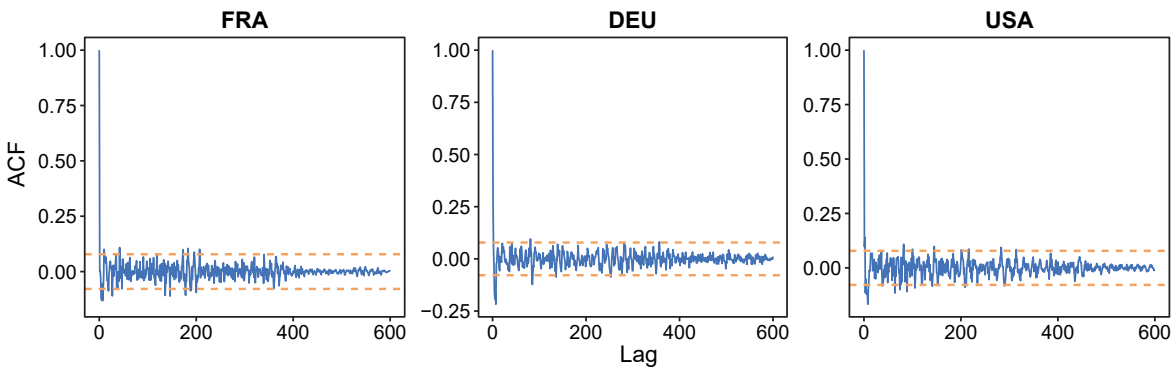

Figure 6. For France, Germany and USA, autocorrelation of the normalized error ε_t, using the festive day correction, obtained with the R-software functionalities (acf() function). The orange dotted line gives the 95% confidence interval for non-correlation. Similar plots for the same countries and 33 more countries, without using the festive day correction, are displayed in fig. A5.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

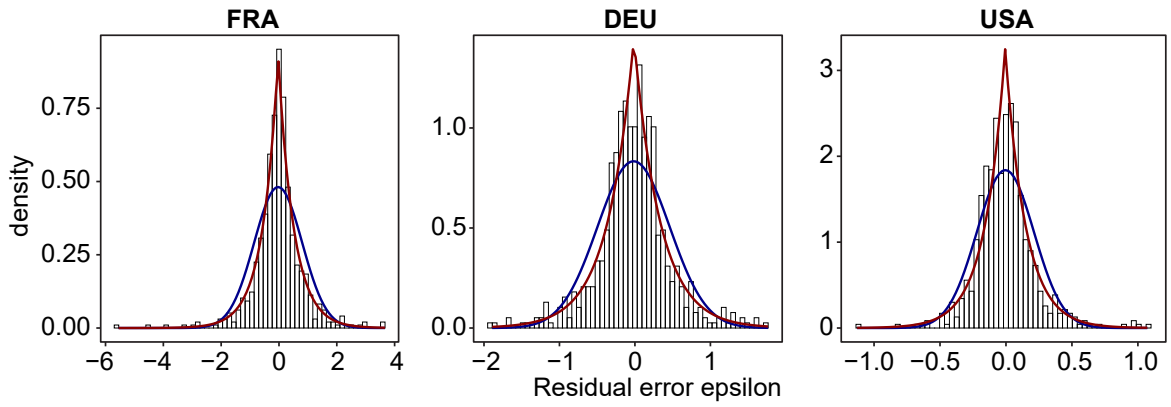

Figure 7. For France, Germany and USA, histogram of the normalized error ε_t, using the festive day correction, its normal approximation (blue line) and its optimal approximation using an exponential distribution (red line) (we use the R-package *normalp* to approximate ε_t by an exponential distribution). See fig. A4 for the results for the same countries and 33 more countries without using the festive day correction.

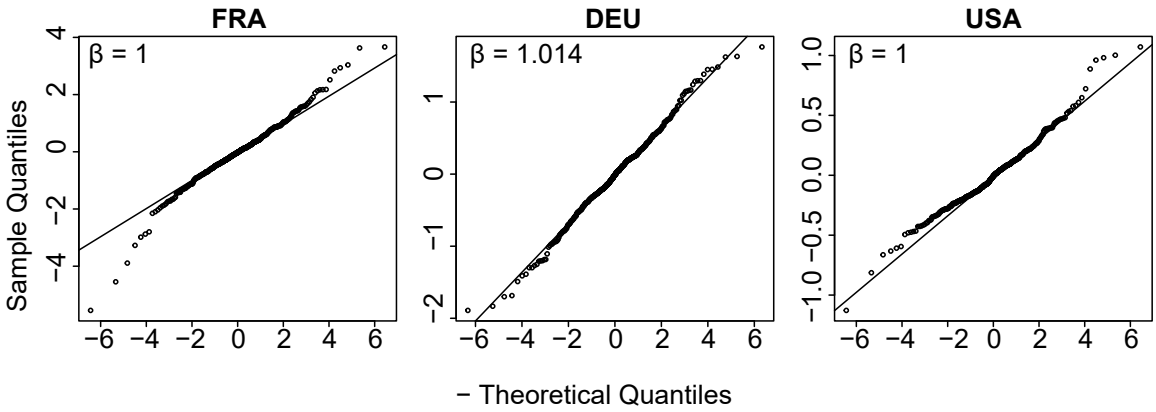

Figure 8. Quantile-quantile plot with France, Germany and the USA comparing ε_t, using the festive day correction, with the optimal exponential distribution using the R-package *normalp*.

where μ is the location, α the scale and β the shape. These parameters to approximate ε_t by an exponential power distribution were estimated by the R-package *normalp* [19].

In Fig. 7, we plot for these three countries the histogram of the distribution of ε_t and its approximation by a normal ($\beta = 2$) and by the obtained optimal exponential distribution. We display the same result for 33 more countries in fig. A4.

Table A1 gives the results for all countries. Columns 5 to 8 in the table give the parameters of the optimal exponential law: location, scale, shape. In all cases the exponent remains close to 1. Fig. 8 displays a quantile-quantile plot comparing ε_t with the estimated exponential distribution for three countries: France, Germany, USA. The linear fit is excellent, and this goodness of fit is confirmed for 33 more countries in Figure A6.

4. Review of previous models

4.1. The Fraser renewal equation

In our proposed incidence model, we used the general integral equation (1), which is a functional equation in R_t. Integral equations have been previously used to estimate R_t: in [20], the authors

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

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. A simpler functional equation than (1) was proposed in Fraser [21] (equation (9)),

$$i_t = R_t \sum_s i_{t-s} \Phi_s. \tag{11}$$

This equation is derived from the general case renewal equation (1) by assuming that R_t is constant in the serial interval support. It computes the "instantaneous reproduction number" and represents the number of secondary cases arising from an individual showing symptoms at a particular time, assuming that conditions remain identical after that time, in contrast with the case renewal equation (1). This last equation applied to the incidence curve is coherent if Φ_s denotes the serial interval between two cases, which can have negative dates, because an infectious may be detected after the infection cases she caused. Using (11) requires that Φ_s only has positive dates. This explains why [22] proposed to estimate the generation time, namely the (always positive) time between two infections, before using it in (11). The advantage of equation (11) is that R_t is estimated at time t from the past incidence values i_{t-s} by a simple division, provided that $\Phi_s = 0$ for $s < 0$:

$$R_t = \frac{i_t}{\sum_s i_{t-s} \Phi_s}. \tag{12}$$

4.2. Deterministic implementations using Fraser's renewal equation (11) and other models

Many papers estimating R_t use the deterministic causal renewal equation (11). This is the case of [23], [24] [25]. This last paper also involves the Wallinga-Teunis formulation [2], also based on the renewal equation but only allowing a backward estimate of R_t (see the discussion in [7]). Some papers like [26] propose a simplified version of (11). See also [27], who use this equation but estimate the probability distribution Φ_s by a maximum entropy method. A few papers use another deterministic model, the Wallinga-Teunis formulation, to compute R_t [28], or a SIR model, like in [29], where the time variable parameter $\beta(t)$ of the three ODE's of a SIR model is estimated from incidence data in a seven days sliding window.

4.3. Stochastic observation models for i_t and R_t

The renewal equation (11) is often endowed with an *a priori* stochastic Poisson model as

$$i_t = \mathcal{P}\left(R_t \sum_s i_{t-s} \Phi_s \right). \tag{13}$$

In this stochastic formulation, the first member i_t of Equation (11) 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 [3,8–10,30]). This form of the renewal equation is proposed and used in [3] and in the EpiEstim software. It is highly recommended in a recent review [31] signed by representatives from ten different epidemiological labs from several continents. Many papers dedicated to the computation of R_t use this model, for example [32], [33] and [34], who also assume that R_t is a Poisson variable, and [35] who also assume that R_t also is a random variable following a Gamma distribution. In [36], the authors use the stochastic form of the renewal equation (13) where they call Φ_s *causal serial interval*. Then R_t is estimated jointly on all regions of a country by a variational model containing a spatial total variation regularization to ensure that R_t is piecewise constant, and the L^1 norm of its time Laplacian to ensure time regularity. The functional also penalizes outliers, typically Sundays and holidays by assuming a sparse structure of such events. See also [37] for an exposition of the application of this method.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In [38] the method Epifilter is introduced as an extension of EpiEstim and of the Wallinga-Teunis formulation. Epifilter has been applied in practical studies like [39]. The core of Epifilter is again the causal renewal equation in Poisson form (13). Yet, the author proposes a doubly stochastic model, as R_t is assumed to follow a recursive discrete Brownian motion of the sort

$$R_t = R_{t-1} + \eta\sqrt{R_{t-1}}\epsilon_{t-1}, \tag{14}$$

where $\epsilon \simeq \mathcal{N}(0,1)$ and η is a user parameter, that we can interpret as a regularity control on R_t. Then $(R_s)_{s\leq t}$ is computed from the incidence data $(i_t)_{s\leq t}$ by recursive filtering. The method is complemented by Bayesian (backward) recursive smoothing that brings a better estimate on low incidence periods.

Similarly in [40], a parametric model with a stochastic multiplicative term is proposed for R_t where the stochastic term is a Gamma law with prescribed standard deviation. The parameters are estimated in several prefectures in interaction to give the best fit to incidence data linked to R_t through the causal renewal equation (11).

A few papers assume a negative binomial *a priori* for the incidence [41]. Nevertheless the equations given in the paper indicate the adoption of the renewal equation (11) and put the stochastic process on R_t by assuming $R_t \simeq R_{t-1}GP$ where GP is a squared exponential kernel. The very same model is used in [42], and is based on the authors' software EpiNow2. Similarly in [43], incidence i_t and reproduction number R_t are linked through the classic SIR model; a parametric piecewise linear model for R_t is estimated by fitting the parameters to real incidence data. Here the daily incidence data are modeled as a negative binomial, with mean given by the deterministic solution of the SIR equations and unknown dispersion.

In [44], a direct stochastic model is proposed for R_t, assuming that its log derivative is Brownian, namely

$$d(log(\beta(t)) = \nu dB(t)$$

where ν is the volatility of the Brownian process to be estimated. Then we have

$$R_t = C\beta(t)s(t),$$

where C is a constant depending on steady transmission characteristics and $s(t)$ is the proportion of the population that is susceptible. The case incidence is then estimated through an SEIR model. We refer to [45] for a still more complex stochastic model for R_t, depending on three stochastic parameters.

5. Discussion

In [7] we have proven extensively by simulations and experiments on live worldwide Covid-19 incidence data that using the simplified causal renewal equation (11) incurs in a five days delay in the estimation of R_t, compared to the Nishiura renewal equation (1). This is why we used here this second model.

All of the stochastic models mentioned in section 4.3 are formulated *a priori*. To the best of our knowledge, no there has been no *a posteriori* verification of their noise models on i_t or R_t. In contrast, we have proposed to learn the noise model from data and to verify *a posteriori* that the noise model is correct. Our experiments show that the weekly and festive administrative perturbations are more important than the noise. Hence they must be corrected first to enable a proper noise analysis.

These experiments seem to confirm the validity of the observation model (4). As we saw, this model can be inverted by minimizing the energy (7). This minimization yields three signals: a restored incidence on the festive days, the administrative bias correcting coefficients q_t that are quasi-periodic with period 7, and the time varying reproduction number R_t, arguably the pandemic's most useful control parameter. Last but not least, the renewal equation deduces a restored incidence i_t^r by (2) from the bias compensated incidence $\hat{i}_t$. The modeling loop was closed by verifying that the normalized error defined by (9) is a white noise. We also found that this noise follows an exponential distribution.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

This analysis discards the Poisson model for the pandemic's case count i_t. A pure case count *should* be a Poisson noise, but we saw that the main perturbation was an administrative bias which, once compensated, leaves behind a noise with standard deviation proportional to a power larger than 0.5 of the case count i_t. Under the Poisson model this standard deviation would have been equal to the square root of i_t.

In summary, based on the renewal equation inversion, this work contributes to a better understanding of the dynamic of the registered administrative observation of the incidence curve, its weekly seasonality, the influence of the festive days and the expected noise model in the observation of the incidence curve.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1. Lotka, A.J. Relation between birth rates and death rates. *Science* **1907**, *26*, 21–22.
2. Wallinga, J.; Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology* **2004**, *160*, 509–516.
3. Cori, A.; Ferguson, N.M.; Fraser, C.; Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* **2013**, *178*, 1505–1512.
4. Ma, S.; Zhang, J.; Zeng, M.; Yun, Q.; Guo, W.; Zheng, Y.; Zhao, S.; Wang, M.H.; Yang, Z. Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *Medrxiv* **2020**.
5. Nishiura, H. Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918–19. *Theoretical Biology and Medical Modelling* **2007**, *4*, 20.
6. Nishiura, H.; Chowell, G., The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. In *Mathematical and Statistical Estimation Approaches in Epidemiology*; Chowell, G.; Hyman, J.M.; Bettencourt, L.M.A.; Castillo-Chavez, C., Eds.; Springer Netherlands: Dordrecht, 2009; pp. 103–121.
7. Alvarez, L.; Colom, M.; Morel, J.D.; Morel, J.M. Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique. *Proceedings of the National Academy of Sciences* **2021**, *118*, 1–10.
8. Thompson, R.; Stockwin, J.; van Gaalen, R.D.; Polonsky, J.; Kamvar, Z.; Demarsh, P.; Dahlqwist, E.; Li, S.; Miguel, E.; Jombart, T.; others. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **2019**, *29*, 100356.
9. Liu, Q.H.; Ajelli, M.; Aleta, A.; Merler, S.; Moreno, Y.; Vespignani, A. Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences* **2018**, *115*, 12680–12685.
10. Obadia, T.; Haneef, R.; Boëlle, P.Y. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC medical informatics and decision making* **2012**, *12*, 147.
11. L. Alvarez, M. Colom, J. D. Morel, J. M. Morel. EpiInvert Online Interface, IPOL : Image Processing On Line. , Accessed December 9, 2021.
12. Tikhonov, A.N.; Arsenin, V.Y. Solutions of ill-posed problems. *New York* **1977**, *1*, 30.
13. Benning, M.; Burger, M. Modern regularization methods for inverse problems. *Acta Numerica* **2018**, *27*, 1–111. doi:10.1017/S0962492918000016.
14. Hyndman, R.; Athanasopoulos, G. *Forecasting: principles and practice, 2nd edition*; OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on December, 9, 2021, 2018. ISBN 1886529043.
15. Government of France. Informations Covid-19, Carte et données. , Accessed December 9, 2021.
16. Robert Koch-Institut. COVID-19-Dashboard. , Accessed December 9, 2021.
17. Spanish Goverment. Situación actual COVID-19 . , Accessed December 9, 2021.
18. H. Ritchie et al.. Coronavirus Pandemic (COVID-19), OurWorldInData.org, . Accessed December 9, 2021.
19. Mineo, A.M. On the estimation of the structure parameter of a normal distribution of order p. *Statistica* **2003**, *63*, 109–122. doi:10.6092/issn.1973-2201/342.
20. Demongeot, J.; Oshinubi, K.; Seligmann, H.; Thuderoz, F. Estimation of Daily Reproduction rates in COVID-19 Outbreak. *medRxiv* **2021**, [https://www.medrxiv.org/content/early/2021/01/04/2020.12.30.20249010.full.pdf]. doi:10.1101/2020.12.30.20249010.
21. Fraser, C. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. *PLOS ONE* **2007**, *2*, 1–12. doi:10.1371/journal.pone.0000758.
22. Knight, J.; Mishra, S. Estimating effective reproduction number using generation time versus serial interval, with application to COVID-19 in the Greater Toronto Area, Canada. *Infectious Disease Modelling* **2020**, *5*, 889–896.
23. Bonifazi, G.; Lista, L.; Menasce, D.; Mezzetto, M.; Pedrini, D.; Spighi, R.; Zoccoli, A. A simplified estimate of the effective reproduction number

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

R t using its relation with the doubling time and application to Italian COVID-19 data. *The European Physical Journal Plus* **2021**, *136*, 1–14.

24. et al., S.F. Estimating the number of infections and the impact of nonpharmaceutical interventions on COVID-19 in 11 European countries. Imperial College COVID-19 Response Team, https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-Europe-estimates-and-NPI-impact-30-03-2020.pdf.
25. Koyama, S.; Horie, T.; Shinomoto, S. Estimating the time-varying reproduction number of COVID-19 with a state-space method. *PLoS computational biology* **2021**, *17*, e1008679.
26. Drewes, H.; Flaeschner, G.; Moeller, P. Improving the reproduction number calculation by treating for daily variations of SARS-CoV-2 cases. *medRxiv* **2021**.
27. Tao, Y. Maximum entropy method for estimating the reproduction number: An investigation for COVID-19 in China and the United States. *Physical Review E* **2020**, *102*, 032136.
28. Wang, K.; Zhao, S.; Li, H.; Song, Y.; Wang, L.; Wang, M.H.; Peng, Z.; Li, H.; He, D. Real-time estimation of the reproduction number of the novel coronavirus disease (COVID-19) in China in 2020 based on incidence data. *Annals of translational medicine* **2020**, *8*.
29. Shapiro, M.B.; Karim, F.; Muscioni, G.; Augustine, A.S. Adaptive Susceptible-Infectious-Removed Model for Continuous Estimation of the COVID-19 Infection Rate and Reproduction Number in the United States: Modeling Study. *Journal of Medical Internet Research* **2021**, *23*, e24389.
30. Boulmezaoud, T.Z.; Alvarez, L.; Colom, M.; Morel, J.M. A Daily Measure of the SARS-CoV-2 Effective Reproduction Number for all Countries. *Image Processing On Line* **2020**, *10*, 191–210. https://doi.org/10.5201/ipol.2020.304.
31. Gostic, K.M.; McGough, L.; Baskerville, E.B.; Abbott, S.; Joshi, K.; Tedijanto, C.; Kahn, R.; Niehus, R.; Hay, J.A.; De Salazar, P.M.; others. Practical considerations for measuring the effective reproductive number, R t. *PLoS computational biology* **2020**, *16*, e1008409.
32. You, C.; Deng, Y.; Hu, W.; Sun, J.; Lin, Q.; Zhou, F.; Pang, C.H.; Zhang, Y.; Chen, Z.; Zhou, X.H. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *International Journal of Hygiene and Environmental Health* **2020**, *228*, 113555.
33. Chintalapudi, N.; Battineni, G.; Sagaro, G.G.; Amenta, F. COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy. *International Journal of Infectious Diseases* **2020**, *96*, 327–333.
34. Hong, H.G.; Li, Y. Estimation of time-varying reproduction numbers underlying epidemiological processes: A new statistical tool for the COVID-19 pandemic. *PloS one* **2020**, *15*, e0236464.
35. Salas, J. Improving the estimation of the COVID-19 effective reproduction number using nowcasting. *Statistical Methods in Medical Research* **2021**, p. 09622802211008939.
36. Pascal, B.; Abry, P.; Pustelnik, N.; Roux, S.G.; Gribonval, R.; Flandrin, P. Nonsmooth convex optimization to estimate the Covid-19 reproduction number space-time evolution with robustness against low quality data. *arXiv preprint arXiv:2109.09595* **2021**.
37. Abry, P.; Pustelnik, N.; Roux, S.; Jensen, P.; Flandrin, P.; Gribonval, R.; Lucas, C.G.; Guichard, É.; Borgnat, P.; Garnier, N. Spatial and temporal regularization to estimate COVID-19 reproduction number R (t): Promoting piecewise smoothness via convex optimization. *Plos one* **2020**, *15*, e0237901.
38. Parag, K.V. Improved estimation of time-varying reproduction numbers at low case incidence and between epidemic waves. *PLoS Computational Biology* **2021**, *17*, e1009347.
39. Mee, P.; Alexander, N.; Mayaud, P.; Gonzalez, F.d.J.C.; Abbott, S.; de Souza Santos, A.A.; Acosta, A.L.; Parag, K.V.; Pereira, R.H.; Prete Jr, C.A.; others. Tracking the emergence of disparities in the subnational spread of COVID-19 in Brazil using an online application for real-time data visualisation: a longitudinal analysis. *The Lancet Regional Health-Americas* **2021**, p. 100119.
40. Jung, S.m.; Endo, A.; Akhmetzhanov, A.R.; Nishiura, H. Predicting the effective reproduction number of COVID-19: inference using human mobility, temperature, and risk awareness. *International Journal of Infectious Diseases* **2021**, *113*, 47–54.
41. Abbott, S.; Hellewell, J.; Thompson, R.N.; Sherratt, K.; Gibbs, H.P.; Bosse, N.I.; Munday, J.D.; Meakin, S.; Doughty, E.L.; Chun, J.Y.; others. Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research* **2020**, *5*, 112.
42. Sherratt, K.; Abbott, S.; Meakin, S.R.; Hellewell, J.; Munday, J.D.; Bosse, N.; working group, C.C..; Jit, M.; Funk, S. Exploring surveillance data biases when estimating the reproduction number: with insights into

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

subpopulation transmission of Covid-19 in England. *Philosophical Transactions of the Royal Society B* **2021**, *376*, 20200283.

43. Karnakov, P.; Arampatzis, G.; Kičić, I.; Wermelinger, F.; Wälchli, D.; Papadimitriou, C.; Koumoutsakos, P. Data-driven inference of the reproduction number for COVID-19 before and after interventions for 51 European countries. *Swiss medical weekly* **2020**, *150*, w20313.
44. Cazelles, B.; Champagne, C.; Nguyen-Van-Yen, B.; Comiskey, C.; Vergu, E.; Roche, B. A mechanistic and data-driven reconstruction of the time-varying reproduction number: Application to the COVID-19 epidemic. *PLoS computational biology* **2021**, *17*, e1009211.
45. Mellan, T.A.; Hoeltgebaum, H.H.; Mishra, S.; Whittaker, C.; Schnekenberg, R.P.; Gandy, A.; Unwin, H.J.T.; Vollmer, M.A.; Coupland, H.; Hawryluk, I.; others. Subnational analysis of the COVID-19 epidemic in Brazil. *MedRxiv* **2020**.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

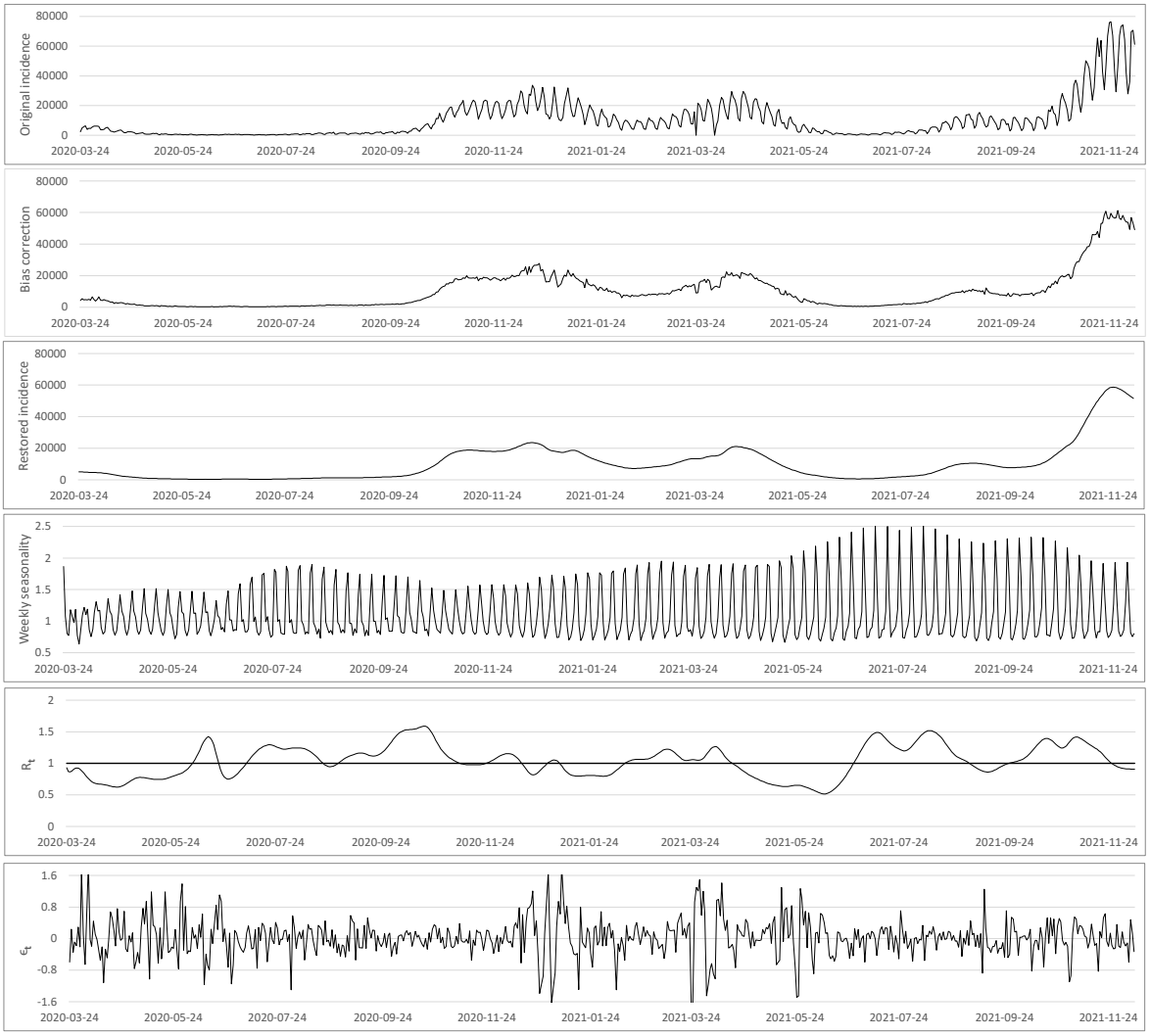

Figure A1. From top to bottom : (i) the original incidence curve of Germany i_t^0, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table A2).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

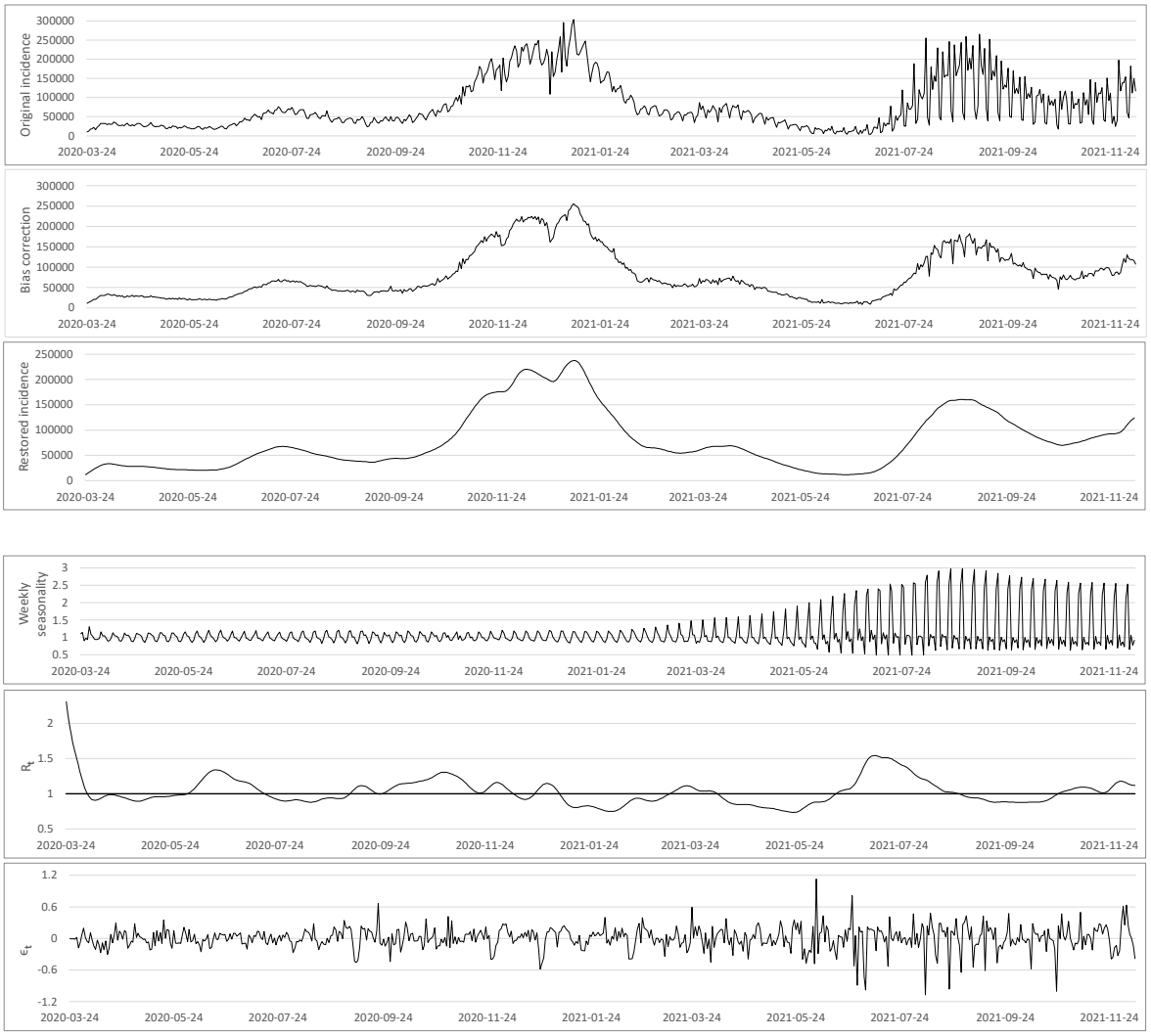

Figure A2. From top to bottom : (i) the original incidence curve i_t^0 of USA, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table A2).

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

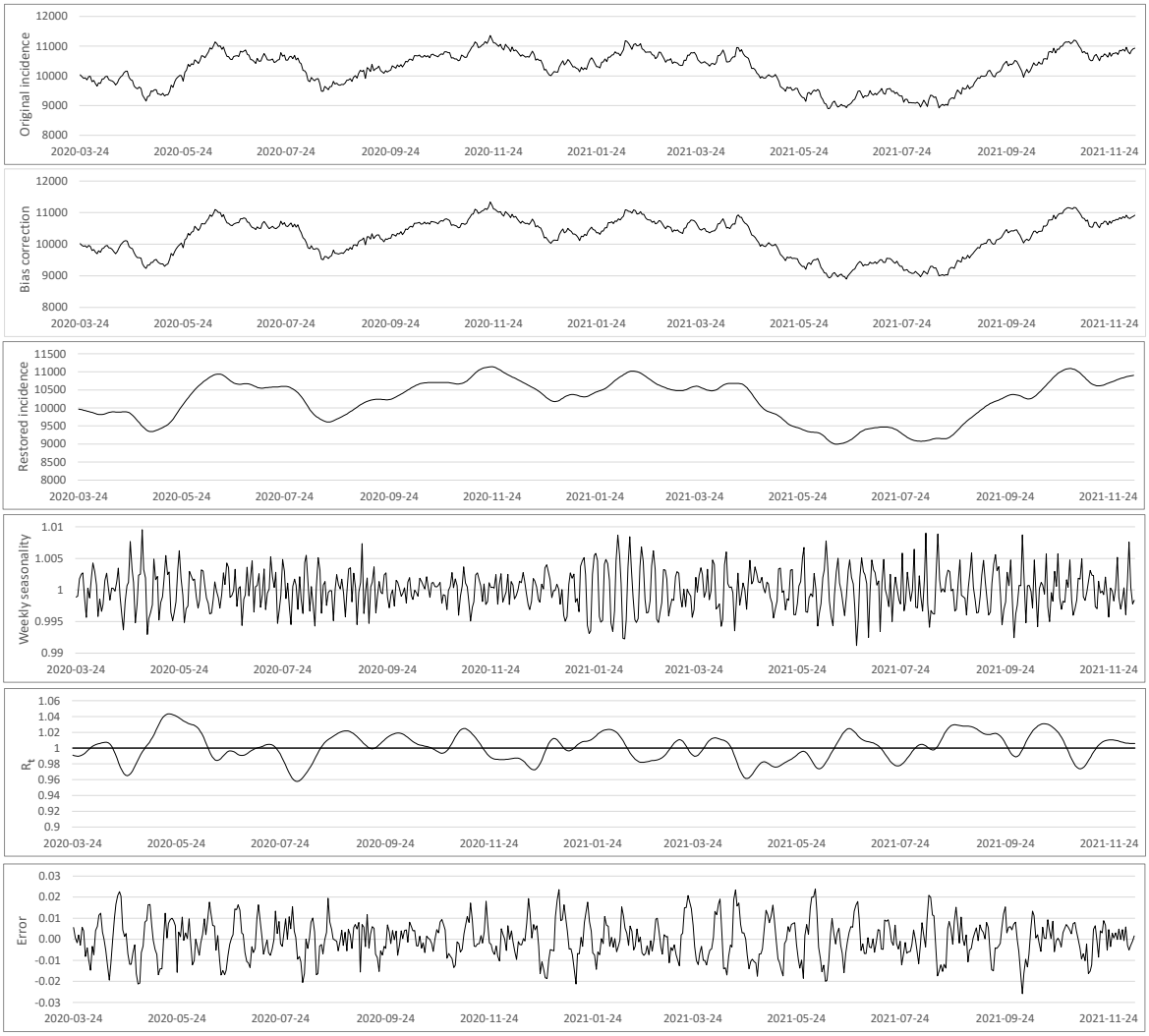

Figure A3. Control test, from top to bottom : (i) the test incidence curve i_t^0 which is a Brownian motion, (ii) the test curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the relative error $(i_t^r - \hat{i}_t)/i_t^r$. Both R_t and the weekly seasonality correction coefficients stay very close to 1, with means 1.001 and 1.00002, and standard deviations 1.7% and 0.3% respectively.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Country	Mean	Std	Location	Scale	Shape (β)
			Exponential	*Exponential*	*Exponential*
FRA*	-0.0283	0.8290	-0.0286	0.5394	1.0000
DEU*	-0.0178	0.4785	-0.0135	0.3433	1.0144
USA*	-0.0044	0.2169	-0.0059	0.1537	1.0000
FRA	0.0109	1.0024	-0.0316	0.6026	1.0000
DEU	0.0091	0.5143	0.0050	0.3458	1.0000
USA	0.0032	0.4779	-0.0097	0.3003	1.0000
ARG	0.0025	0.4430	-0.0286	0.3153	1.0000
AUT	0.0419	1.1030	-0.0041	0.9035	1.2701
BEL	0.0413	1.2175	-0.0366	0.8304	1.0000
BRA	-0.0018	0.4825	-0.0368	0.3312	1.0000
CAN	0.0068	1.2720	-0.0252	0.8290	1.0000
CHL	0.0019	0.2960	-0.0082	0.2138	1.0252
COL	-0.0026	0.2006	-0.0107	0.1490	1.0751
CZE	0.0116	0.5671	-0.0415	0.3755	1.0000
DNK	0.0278	1.2446	-0.0298	0.8126	1.0000
GRC	0.0218	1.2764	-0.0410	0.8847	1.0000
HUN	0.0069	0.6600	-0.0267	0.4410	1.0000
IND	0.0419	0.9891	-0.0084	0.6786	1.0000
IDN	-0.0015	0.3374	-0.0140	0.2607	1.1466
IRL	0.0030	1.1778	-0.0748	0.8252	1.0000
ITA	0.0368	1.1441	0.0141	0.7130	1.0000
JPN	0.0243	0.6647	-0.0254	0.4515	1.0000
MEX	-0.0318	1.7329	-0.0955	1.1091	1.0000
NPL	0.0035	0.8994	0.0005	0.5652	1.0000
NLD	0.0437	0.7185	-0.0404	0.4910	1.0000
PHL	-0.0196	2.0401	-0.0930	1.4011	1.0000
POL	-0.0017	0.1911	-0.0043	0.1268	1.0000
ROU	0.0063	0.9465	-0.0011	0.5798	1.0000
RUS	0.0107	0.3383	0.0066	0.2270	1.0000
SRB	0.0675	1.0140	0.0758	0.7932	1.1728
SVK	0.0024	1.3671	-0.0778	0.8194	1.0000
ZAF	0.0139	0.9110	-0.0320	0.7059	1.1497
ESP	0.0637	1.6068	-0.0047	1.0840	1.0000
CHE	0.0528	1.2228	0.0017	0.8667	1.0000
THA	0.0299	1.3738	-0.0312	0.9374	1.0000
TUN	0.0123	1.3033	-0.0845	0.9224	1.0000
UKR	0.0034	0.4117	-0.0215	0.2586	1.0000
ARE	0.0108	0.4192	-0.0127	0.3265	1.1588
GBR	0.0085	0.3304	-0.0171	0.2163	1.0000

Table A1. Table with the mean and standard deviation of ε_t and the parameters of the best fit to the exponential distributions for 36 countries. The data of starred countries in the first three rows have undergone the festive bias correction.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Country	a	b	p-value	Country	a	b	p-value
FRA*	0.8074272	-1.164141	2.01E-75	FRA	0.8136197	-1.1710322	2.76E-71
DEU*	0.8233846	-1.496739	5.99E-95	DEU	0.8235076	-1.5057318	3.01E-92
USA*	0.9076139	-2.264255	3.16E-42	USA	0.8638492	-1.7287377	6.37E-37
ARG	0.8340299	-1.5574878	1.71E-101	AUT	0.6628437	-0.5661912	3.45E-86
BGD	0.9104934	-2.5672893	6.14E-56	BEL	0.7184413	-0.6589731	3.65E-61
BRA	0.8906214	-1.536314	1.03E-58	CAN	0.7240632	-0.6726824	2.96E-44
CHL	0.8349688	-1.9543089	2.64E-40	COL	0.9175985	-2.2638884	3.03E-112
CZE	0.8520268	-1.4708978	2.88E-133	DNK	0.6900743	-0.6284769	2.78E-68
GRC	0.6555842	-0.5683038	2.58E-102	HUN	0.7838904	-1.3618843	4.47E-142
IND	0.7042499	-0.8457334	8.20E-68	IDN	0.8406915	-1.7674138	5.30E-97
IRL	0.7043354	-0.5484242	1.35E-89	ITA	0.6964125	-0.8659193	2.53E-71
JPN	0.7222903	-1.2445353	5.65E-85	MEX	0.725394	-0.4661005	1.76E-32
NPL	0.7548857	-1.0559482	1.42E-55	NLD	0.7494921	-1.1280471	3.35E-96
PHL	0.6715338	-0.1103984	1.90E-47	POL	0.9306078	-2.6041615	3.02E-133
ROU	0.6920366	-1.0282145	4.11E-77	RUS	0.7212814	-2.0048746	4.05E-26
SRB	0.628712	-0.65103	6.76E-92	SVK	0.7381511	-0.7853881	8.53E-164
ZAF	0.7275793	-0.7811203	9.48E-69	ESP	0.6806819	-0.3916179	2.03E-42
CHE	0.6138378	-0.5491828	1.38E-75	THA	0.7110685	-0.4672682	1.63E-222
TUN	0.7539949	-0.503523	3.03E-163	TUR	0.8998264	-2.658924	1.32E-68
UKR	0.8172308	-1.8996555	1.75E-70	ARE	0.7511088	-1.5460453	3.80E-52
GBR	0.8705096	-1.9395546	1.70E-96	**World**	0.7631129	-1.1389749	0.0000000
Brownian	-1.0155743	13.3969412	0.0844				

Table A2. Coefficients a and b for 38 countries of the log-log linear regression $ax + b$ between restored incidence i_t^r and the residual $|\hat{i}_t - i_t^r|$ as displayed in Fig. 4. The Pearson correlation p-values given by the *stats* R package confirm a linear relation. The exponent a varies between 0.7 and 0.9. Stars* indicate countries with festive correction. The pvalues are slightly better with festive correction than without. The last row shows the results on the control curve, simulated as a Brownian process. Its large p-value discards a linear log-log relation, and the estimated values of a and b also stand far away from the estimated values for real incidence curves.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

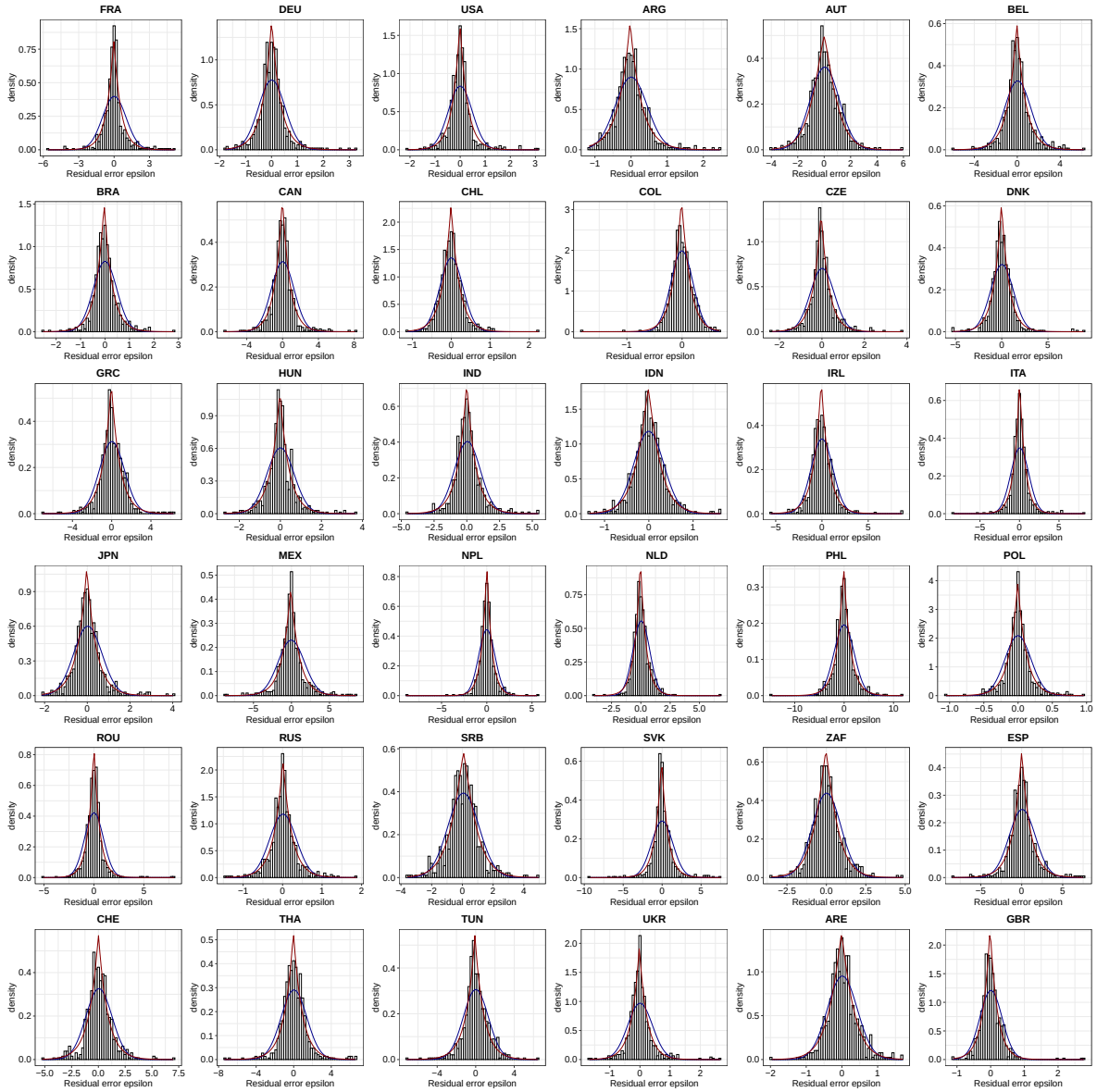

Figure A4. Fit of exponential distributions for 36 countries. In red, the best fitting exponential distribution with shape larger or equal to 1. In black, the best fitting normal law.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

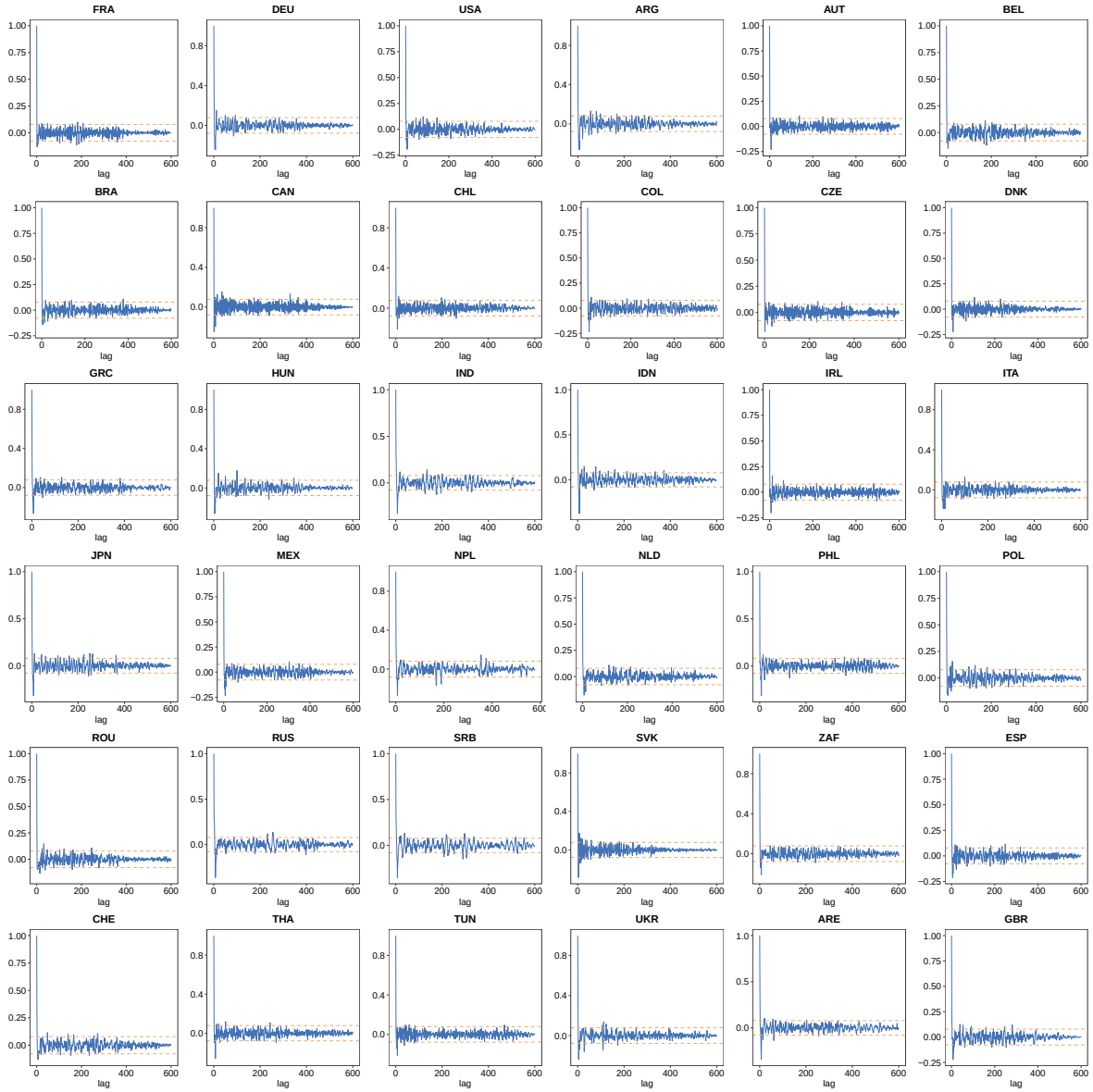

Figure A5. Autocorrelation of the normalized error ε_t using the R-software functionalities (acf() function) for 36 countries. The dotted lines give the 95% confidence interval for non-correlation.

medRxiv preprint doi: https://doi.org/10.1101/2022.02.13.22270901; this version posted March 7, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

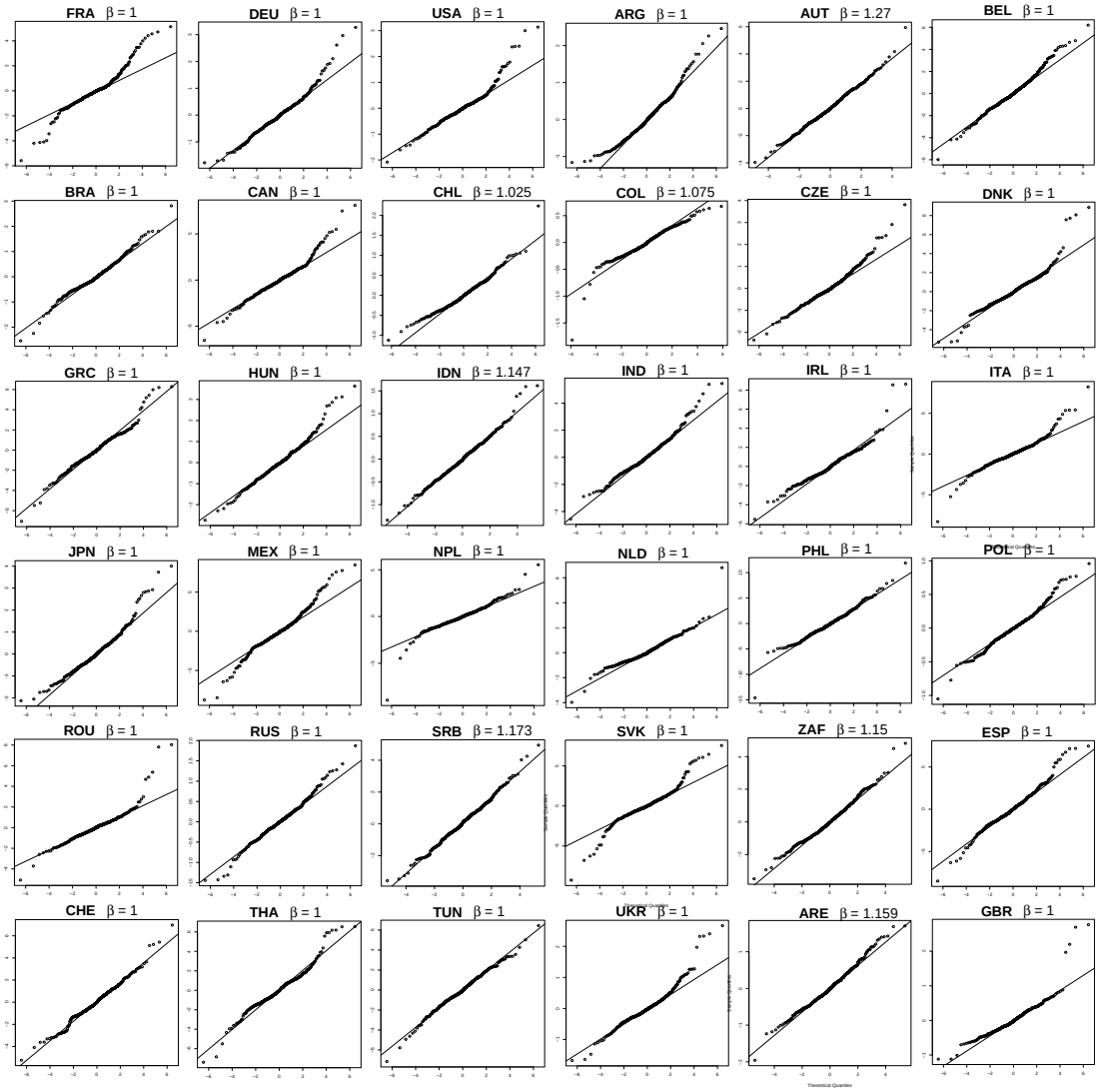

Figure A6. Quantile-quantile plot with 36 countries comparing ε_t (without using the festive day correction) with the optimal exponential distribution using the R-package *normalp*. In the horizontal axis we show the theoretical quantiles and in the vertical axis, the sample quantiles. Note that the exponential distribution shape parameter β, indicated on the graphs can have values >1.