A variational model for computing the effective reproduction number of SARS-CoV-2

Luis Alvarez[1], Miguel Colom[2] and Jean-Michel Morel[2]
[1]University of Las Palmas of Gran Canaria, Spain
[2] Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, France
https://www.ipol.im/ern

Seminaire Infectious Disease Outbreaks
2020/12/09

Content

Content

The renewal equation

Nishiura 2007:

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds.$$

where

- $i(t)$: incidence curve, the number of daily tested positive registered.
- $R(t)$: Effective Reproduction Number defined as the expected number of secondary cases produced by a primary case at each time t.
- $\Phi(s)$: The serial interval which gives the probability distribution of the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases.

The renewal equation

Nishiura 2007:

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds. \approx R(t)\int_0^t i(t-s)\Phi(s)ds$$

where

- $i(t)$: incidence curve, the number of daily tested positive registered.
- $R(t)$: Effective Reproduction Number defined as the expected number of secondary cases produced by a primary case at each time t.
- $\Phi(s)$: The serial interval which gives the probability distribution of the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases.

The renewal equation

Nishiura 2007:

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds. \approx R(t)\int_0^t i(t-s)\Phi(s)ds$$

$$R(t) \approx \frac{i(t)}{\int_0^t i(t-s)\Phi(s)ds}$$

where

- $i(t)$: incidence curve, the number of daily tested positive registered.
- $R(t)$: Effective Reproduction Number defined as the expected number of secondary cases produced by a primary case at each time t.
- $\Phi(s)$: The serial interval which gives the probability distribution of the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases.

The renewal equation

Nishiura 2007:

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds. \approx R(t)\int_0^t i(t-s)\Phi(s)ds$$

$$R(t) \approx \frac{i(t)}{\int_0^t i(t-s)\Phi(s)ds} \quad \rightarrow \quad R_t \approx \frac{i_t}{\sum_{s=1}^t i_{t-s}\Phi_s}$$

where

- $i(t)$: incidence curve, the number of daily tested positive registered.
- $R(t)$: Effective Reproduction Number defined as the expected number of secondary cases produced by a primary case at each time t.
- $\Phi(s)$: The serial interval which gives the probability distribution of the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases.

Content

EpiEstim estimate of $R(t)$

EpiEstim (Cori et al. 2013): If i_t is Poisson distributed with mean $R_t \sum_{s=1}^{t} i_{t-s}\Phi_s$, and $R_t \equiv R_{t,\tau}$ is constant in the time interval $[t-\tau+1, t]$ and Gamma distributed prior, $\Gamma(a,b)$, then the posterior distribution of $R_{t,\tau}$ is also Gamma distributed with a posterior mean given by

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

default value of the parameters: $\tau = 7$, $a = 1$, $b = 5$.

EpiEstim estimate of $R(t)$

EpiEstim (Cori et al. 2013): If i_t is Poisson distributed with mean $R_t \sum_{s=1}^{t} i_{t-s}\Phi_s$, and $R_t \equiv R_{t,\tau}$ is constant in the time interval $[t-\tau+1, t]$ and Gamma distributed prior, $\Gamma(a,b)$, then the posterior distribution of $R_{t,\tau}$ is also Gamma distributed with a posterior mean given by

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

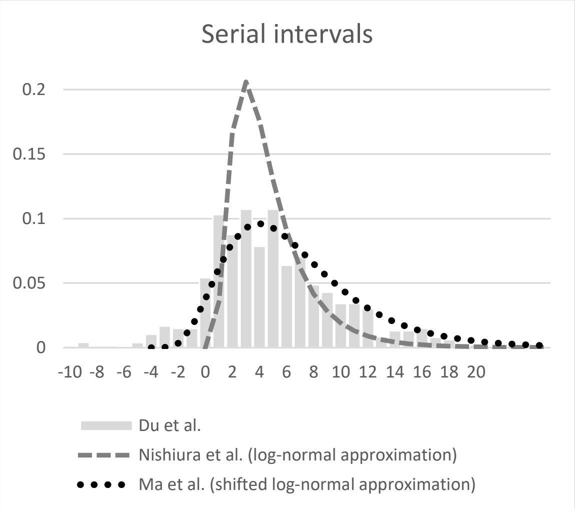
default value of the parameters: $\tau = 7,\ a = 1,\ b = 5$.

Alvarez el al 2020: If $a,\ b$ are time-variant and satisfy:

$$\mathbf{E}(\Gamma(a,b)) = ab = \frac{\sum_{s=t-\tau+1}^{t} i_s}{\sum_{s=t-\tau+1}^{t} \sum_{k=1}^{f} i_{s-k}\Phi_k} \ \rightarrow\ \mathbf{E}(R_{t,\tau}) = \frac{\sum_{s=t-\tau+1}^{t} i_s}{\sum_{s=t-\tau+1}^{t} \sum_{k=1}^{f} i_{s-k}\Phi_k}$$

that is, the means of the prior and posterior distributions are the same.

Serial interval



Significance of the non-positive values in the serial interval

Du et al.

$$\sum_{s\leq 0} \Phi_s = 1.2425 \cdot 10^{-1}$$

Nishiura et al.

$$\sum_{s\leq 0} \Phi_s = 1.2558 \cdot 10^{-4}$$

Ma et al.

$$\sum_{s\leq 0} \Phi_s = 5.7317 \cdot 10^{-2}$$

Content

Extending the Nishiura formula

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t) = \int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

A variational model to compute $R(t)$

Extending the Nishiura formula

$$i(t)=\int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t)=\int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

Variational model

$$E(\{R_t\})=\sum_{t=0}^{t_c}\left(\frac{i_t-\sum_s i_{t-s}R_{t-s}\Phi_s}{p_{90}(i)}\right)^2+\sum_{t=1}^{t_c} w_t(R_t-R_{t-1})^2+\sum_{m=0}^{M}\beta_m(R_{t_m}-\bar{R}_{t_m})^2,$$

where t_c is the current time

A variational model to compute $R(t)$

Extending the Nishiura formula

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t) = \int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

Variational model

$$E(\{R_t\}) = \sum_{t=0}^{t_c} \left(\frac{i_t - \sum_s i_{t-s}R_{t-s}\Phi_s}{p_{90}(i)} \right)^2 + \sum_{t=1}^{t_c} w_t (R_t - R_{t-1})^2 + \sum_{m=0}^{M} \beta_m (R_{t_m} - \bar{R}_{t_m})^2,$$

$p_{90}(i)$ is the 90th percentile of $\{i_t\}_{t=0,..,t_c}$ used to normalize the energy with respect to the size of i_t

boundary conditions: $R_t = R_0$ for $t < 0$ and $R_t = R_{t_c}$ for $t > t_c$

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

where $I_t \equiv \sum_{k=0}^{t} i_k \approx I_0 e^{at}$.

A variational model to compute $R(t)$

Extending the Nishiura formula

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t) = \int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

Variational model

$$E(\{R_t\}) = \sum_{t=0}^{t_c} \left(\frac{i_t - \sum_s i_{t-s} R_{t-s} \Phi_s}{p_{90}(i)} \right)^2 + \sum_{t=1}^{t_c} w_t (R_t - R_{t-1})^2 + \sum_{m=0}^{M} \beta_m (R_{t_m} - \bar{R}_{t_m})^2,$$

$$w_t = \hat{w}_0 \frac{G_{\sigma_w} * i(t)}{p_{90}(i)}$$

where $\hat{w}_0 > 0$ and $G_{\sigma_w} * i(t)$ represents the convolution of i_t with a Gaussian kernel of standard deviation σ_w.

Extending the Nishiura formula

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t) = \int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

Variational model

$$E(\{R_t\}) = \sum_{t=0}^{t_c} \left(\frac{i_t - \sum_s i_{t-s}R_{t-s}\Phi_s}{p_{90}(i)} \right)^2 + \sum_{t=1}^{t_c} w_t(R_t - R_{t-1})^2 + \sum_{m=0}^{M} \beta_m (R_{t_m} - \bar{R}_{t_m})^2,$$

we can use any "a priori" estimate of R_0 as the prescribed value $\bar{R}_{t_0}$ (with $t_0 = 0$). For instance if we assume that $I_t \equiv \sum_{k=0}^{t} i_k \approx I_0 e^{at}$ then:

$$R_0 \approx \frac{1 - e^{-a}}{\sum_s (e^{-sa} - e^{-(s+1)a})\Phi_s}$$

A variational model to compute $R(t)$

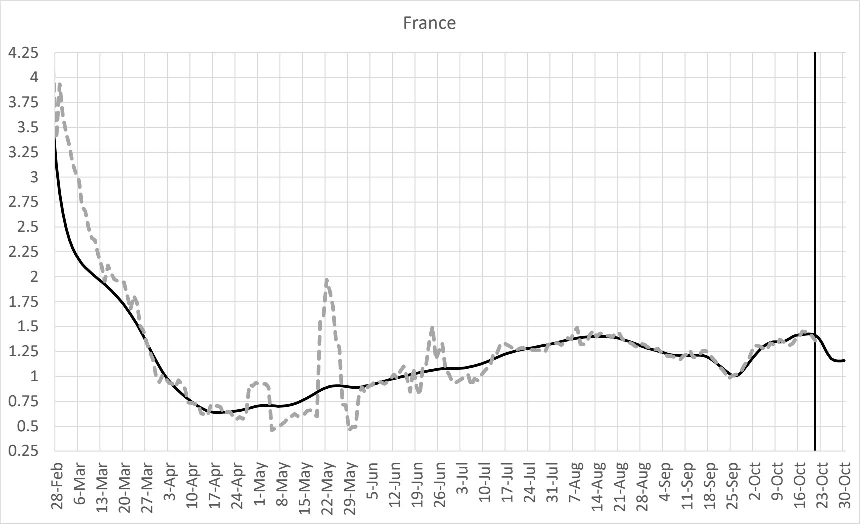
Extending the Nishiura formula

$$i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds \quad \rightarrow \quad i(t) = \int_{-\infty}^{\infty} i(t-s)R(t-s)\Phi(s)ds$$

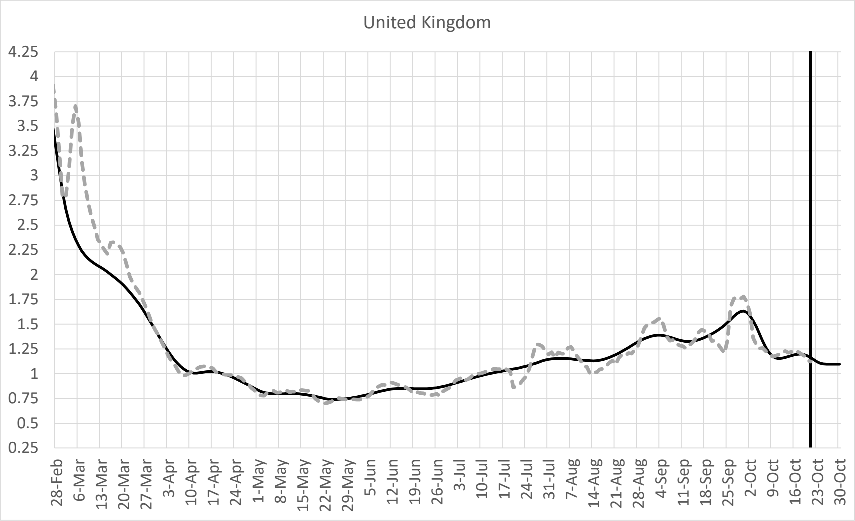
Variational model

$$E(\{R_t\}) = \sum_{t=0}^{t_c} \left(\frac{i_t - \sum_s i_{t-s} R_{t-s} \Phi_s}{p_{90}(i)} \right)^2 + \sum_{t=1}^{t_c} w_t (R_t - R_{t-1})^2 + \sum_{m=0}^{M} \beta_m (R_{t_m} - \bar{R}_{t_m})^2,$$

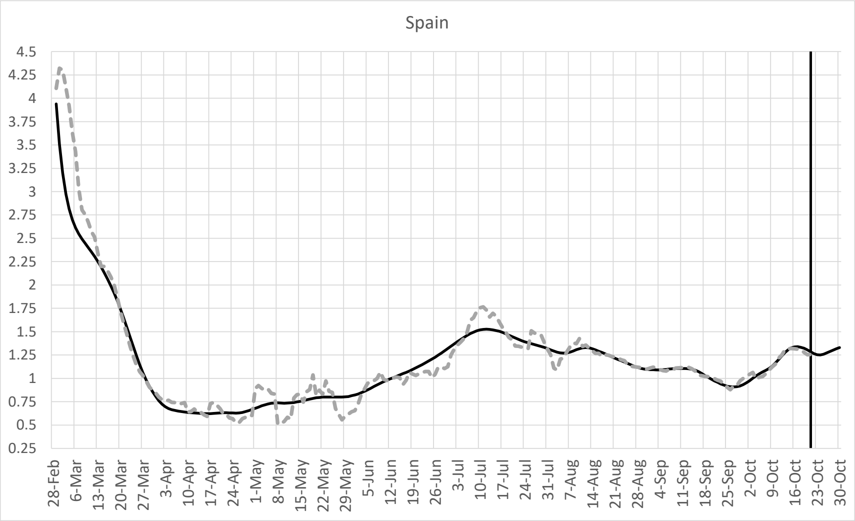
to obtain a more robust estimate of R_{t_c} we can combine the values of R_{t_c} computed the last 3 days to obtain an initial estimate of R_{t_c} that we use as $\bar{R}_{t_1}$. Notice that the parameter β_m determines the confidence we assing to such estimate. Moreover comparing this last estimate of R_t and the three ones obtained in the last 3 days provides a measure of the variability of R_t in the last 3 days.



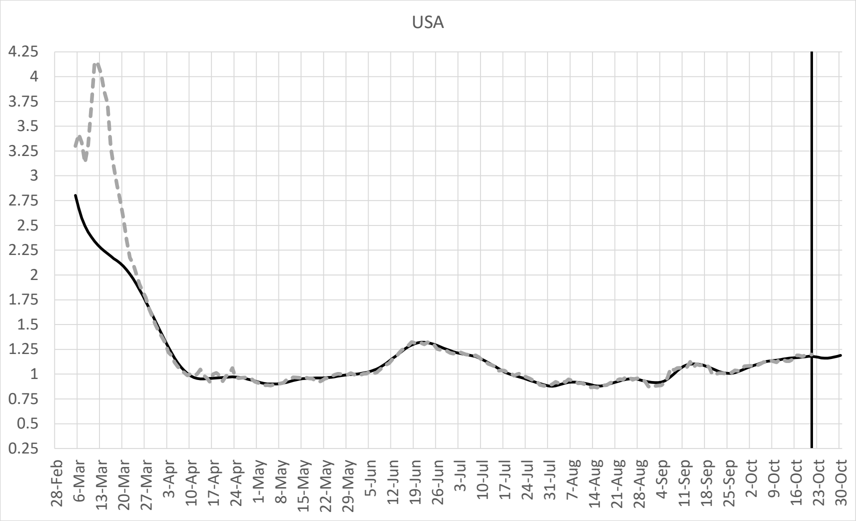
we show the R_t estimate obtained by **EpiEstim** (dotted line) and the **variational model** (solid line). The vertical line represents the 9-day shift applied to the **EspiEstim** estimate to fit the results of the variational technique.



we show the R_t estimate obtained by **EpiEstim** (dotted line) and the **variational model** (solid line). The vertical line represents the 9-day shift applied to the **EspiEstim** estimate to fit the results of the variational technique.



we show the R_t estimate obtained by **EpiEstim** (dotted line) and the **variational model** (solid line). The vertical line represents the 9-day shift applied to the **EspiEstim** estimate to fit the results of the variational technique.



we show the R_t estimate obtained by **EpiEstim** (dotted line) and the **variational model** (solid line). The vertical line represents the 9-day shift applied to the **EspiEstim** estimate to fit the results of the variational technique.

Why do we have a shift between the EpiEstim and the variational technique estimates of $R(t)$?

$$\underbrace{i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds}_{\textbf{Nishiura}} \quad \rightarrow \quad \underbrace{i(t) = R(t)\int_0^t i(t-s)\Phi(s)ds}_{\textbf{EpiEstim}}$$

We can expect approximately a shift between both estimates of $R(t)$ close to the mean of the serial interval $\Phi(s)$ (6.7 in the case of Ma et al.).

Why do we have a shift between the EpiEstim and the variational technique estimates of $R(t)$?

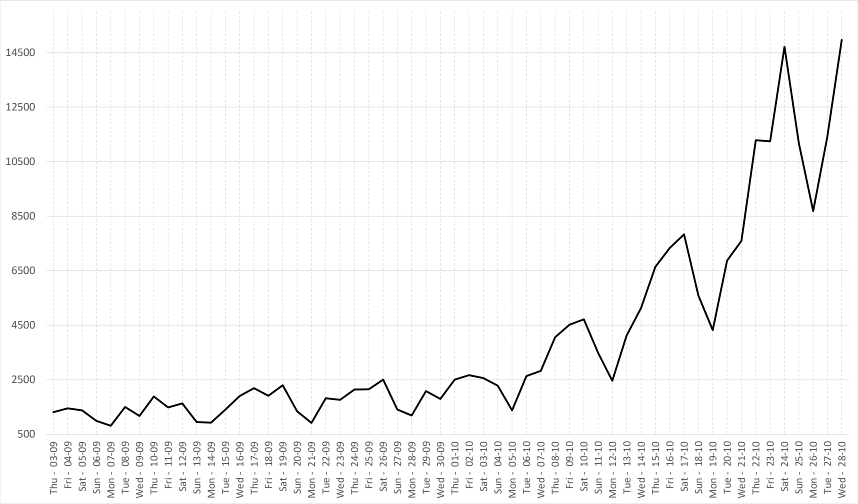
$$\underbrace{i(t) = \int_0^t i(t-s)R(t-s)\Phi(s)ds}_{\textbf{Nishiura}} \quad \rightarrow \quad \underbrace{i(t) = R(t)\int_0^t i(t-s)\Phi(s)ds}_{\textbf{EpiEstim}}$$

We can expect approximately a shift between both estimates of $R(t)$ close to the mean of the serial interval $\Phi(s)$ (6.7 in the case of Ma et al.).

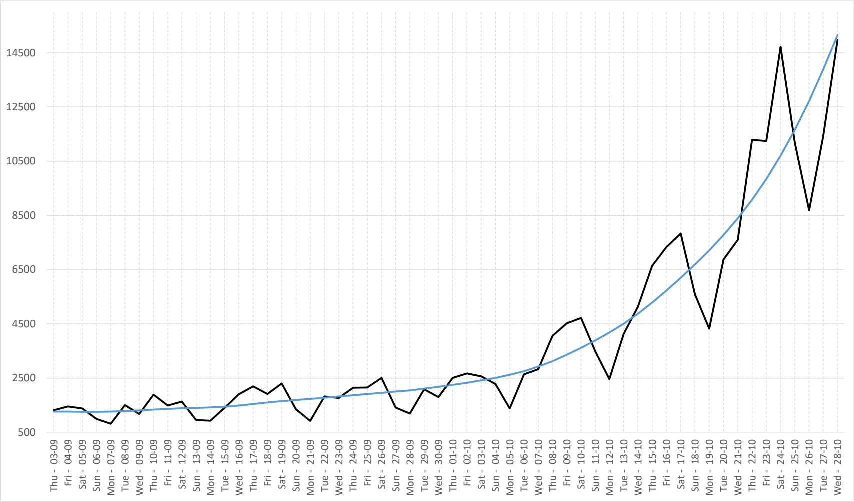
Moreover in **EpiEstim** $R(t)$ is assumed to be constant in the last 7 days (from $t-7$ to t). That introduces an extra shift.

	France	United Kingdom	Spain	USA
shift	8.60	8.46	9.18	8.33
RMSE	0.073807	0.091506	0.081461	0.019639

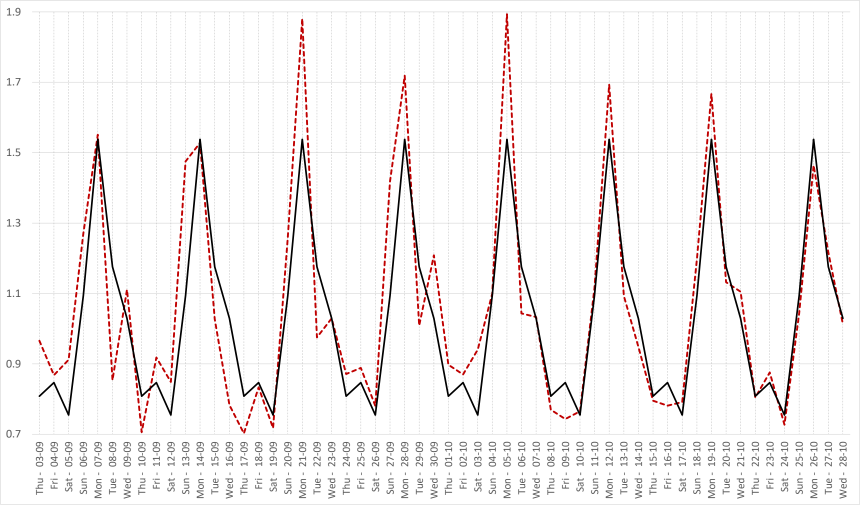
Content



In the case of Germany, we plot i_t between September 9 and October 28. We observe a clear weekly periodic pattern. For instance, on Monday, the number of cases is systematically underestimated and on Saturday the opposite.



In blue we plot the result of applying the **Nishiura** renewal equation $\bar{i}_t = \sum_s i_{t-s} R_{t-s} \Phi_s$. Our main assumption is that the ratio $\bar{i}_t / i_t$ follows a 7-day periodic dynamic.



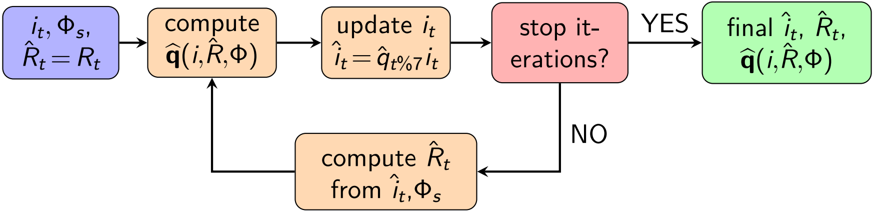
In red we plot the ratio $\frac{\sum_s i_{t-s} R_{t-s} \Phi_s}{i_t}$ and in black its approximation using a 7-day periodic function. In the case of Germany we can observe that this ratio follows quite well a 7-day periodic dynamic.

Capturing the 7-day periodic dynamic

$$\widehat{\mathbf{q}}(i,R,\Phi) = \underset{\mathbf{q}=(q_0,q_1,q_2,q_3,q_4,q_5,q_6)}{\arg\min} \quad E(i,R,\Phi,\mathbf{q}) \equiv \sum_{t=t_c-T+1}^{t_c} \left(\sum_s i_{t-s} R_{t-s} \Phi_s - q_{t\%7} i_t \right)^2$$

where T represents the number of days used in the estimation (in our experiments we use $T = 56$, that is 8 weeks). We point out that the value $\hat{i}_t = \hat{q}_{t\%7} i_t$ can be considered as an update of i_t where we have removed the weekly administrative noise. To preserve the number of accumulated cases in the period of estimation, we add to the minimization problem the constraint

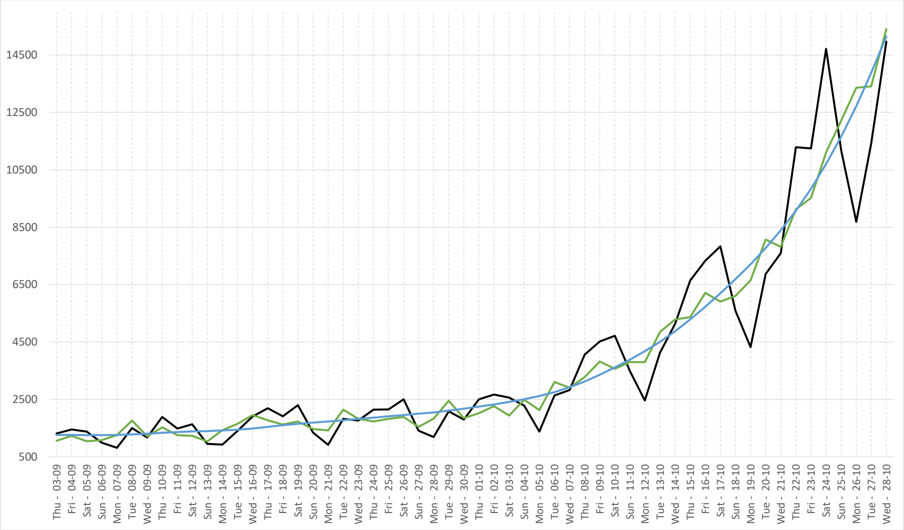
$$\sum_{t=t_c-T+1}^{t_c} i_t = \sum_{t=t_c-T+1}^{t_c} \hat{i}_t = \sum_{t=t_c-T+1}^{t_c} \hat{q}_{t\%7} i_t.$$



Flowchart of the iterative update of $\hat{i}_t$, $\hat{R}_t$ and $\widehat{\mathbf{q}}(i, \hat{R}, \Phi)$. First R_t is computed from i_t and Φ_s using the variational method and $\hat{R}_t = R_t$ is initialized. Next, $\widehat{\mathbf{q}}(i,\hat{R},\Phi)$ is obtained using the explained strategy. Then i_t is updated as $\hat{i}_t = \hat{q}_{t\%7} i_t$. The iteration is stopped when the efficiency measure $\mathcal{I}$ defined below does not improve in the current iteration. Otherwise, $\hat{R}_t$ is updated by the variational method from $\hat{i}_t$ and Φ_s and the iteration goes on.

We use as efficiency the ratio of the RMSE after and before the application of the weekly noise removal given by:

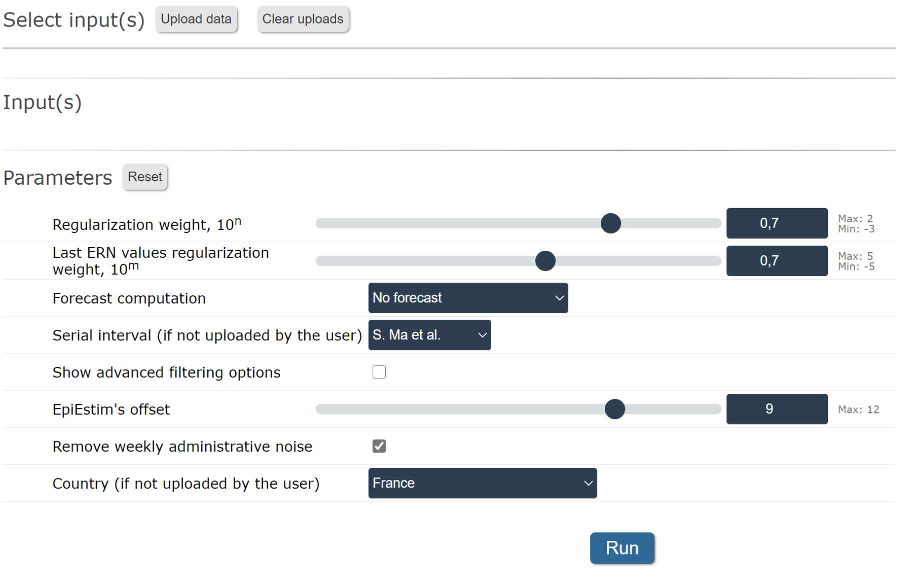
$$\mathcal{I} = \sqrt{\frac{E(\hat{i}, \hat{R}, \Phi, \widehat{\mathbf{q}}(\hat{i}, \hat{R}, \Phi))}{E(i, R, \Phi, \widehat{\mathbf{q}}(i, R, \Phi))}}$$



For Germany we plot, in black, the incidence curve i_t, in blue, the **Nishiura** formula $\sum_s i_{t-s} R_{t-s} \Phi_s$, and in green $\hat{i}_t = \hat{q}_{t\%7} i_t$. ($\mathcal{I} = 0.2717$).

Content

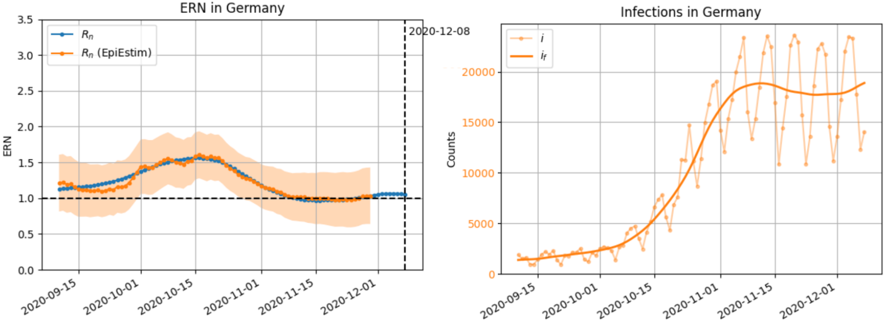
The online interface at https://www.ipol.im/ern



The online interface at https://www.ipol.im/ern

Effective reproduction number (ERN)

plots



Zoom 1x

Additional Info

```
Weekly administrative noise reduction:
RMSE Reduction factor :  0.246567
Weekly noise correction factors (the last factor corresponds to the last day of the registered number of infections):
1.007298 - 0.836259 - 0.798128 - 0.800169 - 1.123279 - 1.523517 - 1.246714 -
```

Download the output (CSV) of the algorithm and the figures:

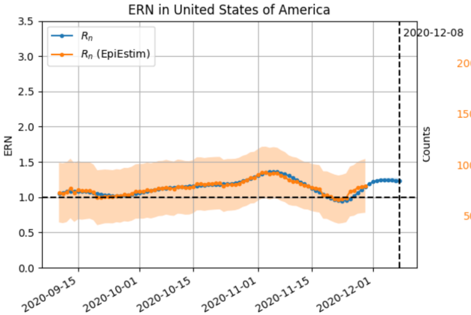
Rn (CSV) | Figure: ERN | Figure: infections

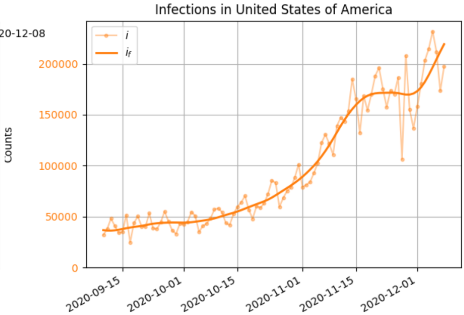
The online interface at https://www.ipol.im/ern

	A	B	C	D	E	F
1	Rt	Nishiura formula	original incidence	variability the last 3 days		
323	0.970097	18479.15233	10824	0.000359		
324	0.970964	18327.05622	14419	0.000434		
325	0.971989	18186.52475	17561	0.000518		
326	0.973119	18077.91702	22609	0.000607		
327	0.974665	18004.64462	23648	0.000692		
328	0.977064	17951.38856	22964	0.000764		
329	0.980588	17894.47728	15741	0.000822		
330	0.985231	17824.16959	10864	0.000887		
331	0.99081	17756.83395	13554	0.001013		
332	0.997283	17717.20916	18633	0.001271		
333	1.004887	17715.26023	22268	0.001735		
334	1.01381	17739.1558	22806	0.002453		
335	1.023935	17766.56375	21695	0.003444		
336	1.034627	17783.82993	14611	0.004698		
337	1.044913	17792.48938	11169	0.006186		
338	1.053719	17806.92637	13604	0.007874		
339	1.060262	17849.09012	17270	0.009706		
340	1.06416	17936.65328	22046	0.011624		
341	1.065434	18076.54668	23448	0.013578		
342	1.064605	18266.12333	23318	0.015542		
343	1.062389	18484.96068	17767	0.017531		
344	1.059378	18699.41155	12332	0.019874		
345	1.056076	18887.3261	14055	0.02257		
346						
347						

Effective reproduction number (ERN)

plots





Zoom 1x

Additional Info

```
Weekly administrative noise reduction:
RMSE Reduction factor :  0.664453
Weekly noise correction factors (the last factor corresponds to the last day of the registered number of infections):
1.019808 - 0.958767 - 0.968317 - 0.851732 - 1.002146 - 1.197887 - 1.049581 -
```

Download the output (CSV) of the algorithm and the figures:

Rn (CSV) | Figure: ERN | Figure: infections

Content

Conclusions

- Based on the **Nishiura** formulation of the renewal equation we propose a variational technique to compute Rt.
- The method can use serial intervals with negative days (as it is the case for the SARS-CoV-2). Thus, it avoids an artificial truncation of the distribution.
- The method computes a point estimate of R_t up to the current date. It seems to provide a more to date (by more than 8 days) estimate of R_t than EpiEstim.
- The method does not assume any statistical distribution for R_t. The main assumptions are that the R_t estimate should follow the renewal equation but keeping R_t regular enough. We include this regularity hypothesis in the model using standard techniques of calculus of variations.
- we propose to capture and to remove the weekly administrative noise assuming that the ratio of the evaluation of the **Nishiura** formula and the incidence curve follows a 7-day periodic dynamics.
- An implementation of the method is available online at `https://www.ipol.im/ern`.

THANKS FOR YOUR ATTENTION

Publications (MedRxiv preprints)
A variational model for computing the effective reproduction number of SARS-CoV-2
https://www.medrxiv.org/content/10.1101/2020.08.01.20165142v3

Removing weekly administrative noise in the daily count of COVID-19 new cases. Application to the computation of Rt.
https://www.medrxiv.org/content/10.1101/2020.11.16.20232405v1

Online interface: https://www.ipol.im/ern