ESTIMATION AND INTERACTIVE VISUALIZATION OF THE TIME-VARYING REPRODUCTION NUMBER R_t AND THE TIME-DELAY FROM INFECTION TO ESTIMATION

A PREPRINT

Fabian Valka
vektorraum
Vienna, Austria
fvalka@vektorraum.com

Carla Schuler
Vienna, Austria
carla.schuler@aon.at

May 2, 2020

ABSTRACT

R_t plays a key role in the development of the COVID-19 pandemic. The methods used for building an interactive website for the visualization of the time-varying reproduction number R_t and a novel way to visualize the time delay from infection to estimation overlayed with the R_t estimate are described and analyzed with regards to the influence of parameters chosen and compared to published estimates for Austria. Visualizing the time delays enables the interactive exploration of suspected changes in transmission and their effect on the R_t estimate.

Keywords COVID-19 · Time-Varying Reproduction Number · R · Austria

1 Introduction

An interactive web-based tool was developed for visualizing and exploring the time-varying reproduction number and time delays from infection to onset, to reporting and inclusion in the estimation of $R_{t,\tau}$ on a regional level for Austria. Users can choose the Austrian state, sliding time window length and overlay the AGES $R_{t,13}$ estimate [1].

The tool is available as free open-source software on GitHub:
https://github.com/fvalka/r_estimate.

1.1 Importance of Time Delays from a Control Theory Perspective

The introduction, discontinuation, reintroduction, and change of non-pharmaceutical interventions can also be viewed from a control theory perspective.[2] Some models suggest a necessity for on-off style control of lockdown measures throughout the post-pandemic period.[3]

Introducing time-delays into otherwise stable closed control-loops can make them unstable.[4, 5] Careful treatment of the time-delays involved in the available observables is therefore essential.

2 Estimation of the Time-Varying Reproduction Number R_t

In analogy to the method published by Richter, et al.[6] from AGES, the time-varying reproduction number R_t is estimated according to Cori, et al.[7] Sliding time windows of length τ days are used with the assumption that within the time window τ the reproduction number R_t is constant, defining $R_{t,\tau}$.

The R-package EpiEstim[8] provided by Cori, et al. is used for the implementation.

Estimation and Interactive Visualization of the Time-Varying Reproduction Number R_t and the Time-Delay from Infection to Estimation

A PREPRINT

| | | 95% ci | | |
Parameter	Mean	Lower	Upper	Std. Dev.
Serial interval mean μ_s	4.46	4.160	4.760	0.153
Serial interval standard deviation σ_s	2.63	2.369	2.891	0.133
		parameters for uncertain_si		

Table 1: Serial interval Gamma distribution Γ_s parameters based upon estimates by Richter, et al.[9]

2.1 Serial Interval

Accuracy of the serial interval estimation plays a key role in the accuracy of the estimation of $R_{t,\tau}$.

Representing the uncertainties contained in the current serial interval estimates[9] was realized using the uncertain_si method in estimate_R. This samples the parameters for the serial intervals Gamma distribution Γ_s from a truncated normal distribution. The median of the $R_{t,\tau}$ estimate is only affected to a limited extent by this, but the credible interval will be increased around those median values[7], giving a more accurate representation based on our current beliefs about the serial interval.

Parameters used in the implementation are documented in Table 1.

Standard deviations of the mean serial interval and the standard deviation of the standard deviation were obtained based upon the assumption of normality from the 95% credible interval, see Equation 1.

$$\sigma_{\mu_x} = \frac{\mu_x - \text{CI}_{lower}}{1.96} \tag{1}$$

Based upon these parameters multiple serial intervals are explored by estimate_R.[8] Explored serial interval distributions are shown in Figure 3. During each estimation 500 discretized Gamma distributions of the serial interval are explored.
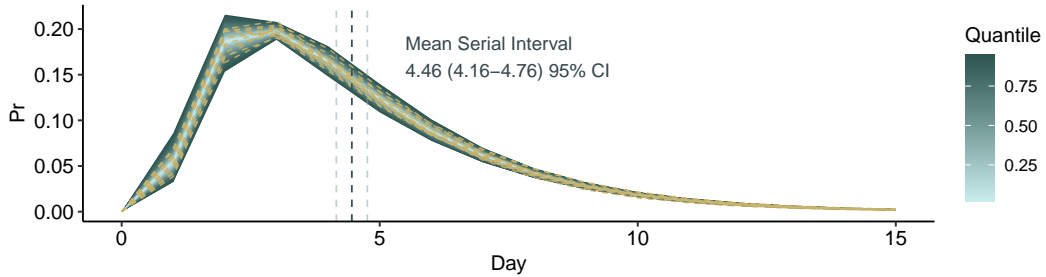

Figure 1: Serial interval distributions explored during $R_{t,\tau}$ estimation. Quantiles shown in color scaling (blue), with the outer border marking the 0.95 quantile overlayed with samples from the explored discrete serial interval distributions (yellow, dashed).

2.2 Sliding Time Window τ

Choosing τ is a trade-off between less noise and less delay. Smaller time windows τ lead to a more rapid detection of changes in $R_{t,\tau}$, but also to more statistical noise.[10]

Cori, et al. describe a method for choosing the time window size τ based upon the minimum number of cases included in each time window [10]. Based upon a chosen target CV of at least 0.3 a minium of 12 cases in each time window is required. With this in consideration and the usage of a time window of 7 days in Cori, et al.[7] a time window size τ of 7 days for the UI was chosen. AGES uses a 13 days time window for their estimation, no explanation is given for this choice and this τ was therefore only used for comparisons but can be chosen by the user on the interactive front-end.

The effect a change in τ has on the estimated $R_{t,\tau}$ is illustrated in Figure 2. Comparing the first date where the median of τ has fallen below 1 illustrates the trade-off discussed above. For $\tau = 7$ the condition $R_{t,7} < 1$ is first met on the 31. of March 2020, for $\tau = 13$ the condition $R_{t,13} < 1$ is only met 4 days later on the 4th of April 2020.

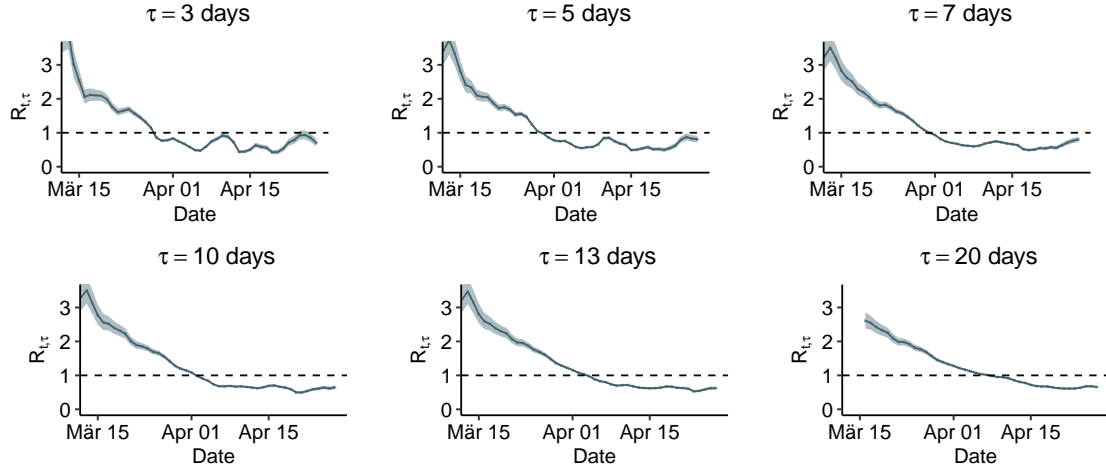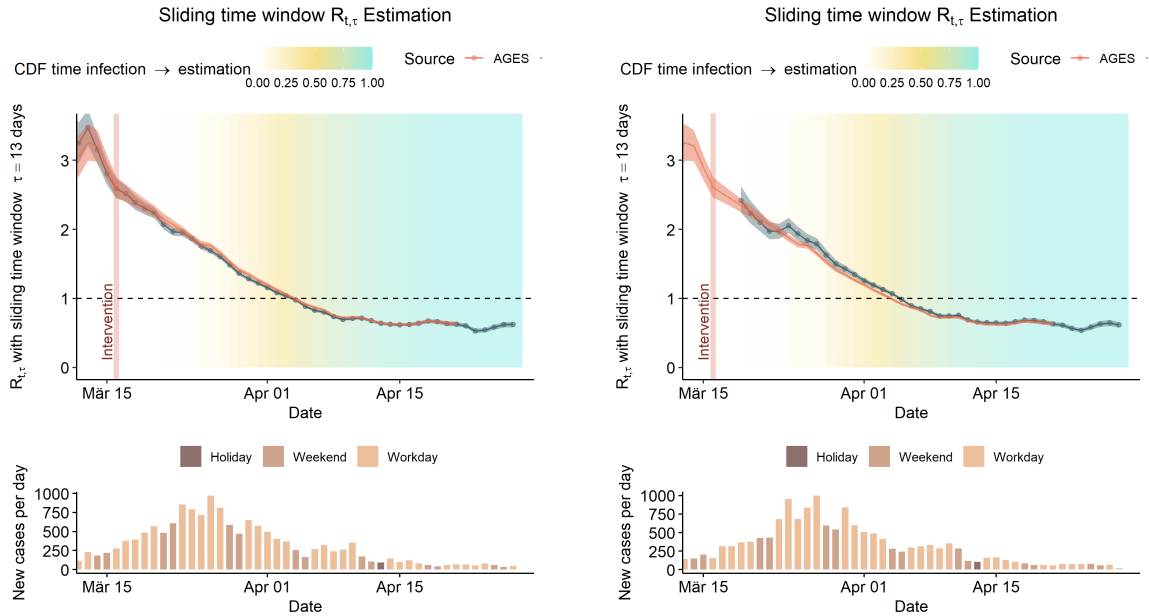A 7-day window has the added advantage that it always contains exactly 2 weekend days.

Figure 2: Effect of different time window, τ, choices on the $R_{t,\tau}$ estimate for Austria, as of 2020-04-27. Left censored at first entry in data-set $t_0 + \tau$ Data source: data.gv.at.[11]

(a) Closest reproduction using the data.gv.at dataset[11] and parametric_si .

(b) Dataset used in actual visualization and using uncertain_si in estimate_R.

Figure 3: Comparison between own estimates and $R_{t,\tau}$ estimates published by AGES.[12]

2.3   Crossvalidation with AGES estimates for Austria

The output of the estimation described in this section was compared to the $R_{t,\tau}$ estimates obtained by AGES on the 24th of April 2020[1].

Two different methods were chosen for the comparison, one where the dataset which allows the closest reproduction of the estimates obtained by AGES was used as shown in figure Figure 3a. In the actual interactive visualization a different dataset is used. The comparison against this dataset is shown in Figure 3b. For a further discussion on the datasets see section 4.

3 Estimation of the time delay from infection to inclusion in the $R_{t,\tau}$ estimation

The total time delay from infection to inclusion in the estimate $t_{infection,estimation}$ will be split up into three different time delays. The time delay from infection to onset $t_{infection,onset}$, the incubtion period. The time delay from onset to reporting $t_{onset,reporting}$ and the time delay from reporting to inclusion in the estimation $t_{reporting,estimation}$, which is caused by the $R_{t,\tau}$ estimation method of assuming a constant R_t within the time window τ.

3.1 Time delay from infection to reporting $t_{infection,reporting}$

Both $t_{infection,onset}$ and $t_{onset,reporting}$ are assumed to be independent random variables and a Gamma distribution was used for the estimation. Estimates of the Gamma distribution parameters for $t_{infection,onset}$ were taken from the supplementary appendix of the study by Zhang, et al.[13] based upon case data from Mainland China, excluding Hubei province. The parameters used are $\alpha_{infection,onset} = 4.23(\text{SD } 1.28)$ and $\beta_{infection,onset} = 0.81(\text{SD } 0.24)$ with a mean of 5.2days and a 95% CI of the mean of (95% CI: $1.5 - 11.3$) days.

Onset to reporting Gamma distribution parameters were also taken from the estimates from Mainland China, excluding Hubei[13] from the second period (Jan 28 – Feb 17) were used. Those parameters were obtained from 2079 observations and are $\alpha_{onset,reporting} = 3.18$ and $\beta_{onset,reporting} = 0.59$ giving a mean of 5.3 (95% CI: $1.2 - 13.1$) days.

In order to obtain the empirical CDF of $t_{infection,reporting}$ random samples were drawn from the Gamma distributions of $t_{infection,onset}$ and $t_{onset,reporting}$ using the R-function **rgamma**[14]. Random samples were added up pair-wise as shown in Equation 2.

$$t^i_{infection,reporting} = t^i_{infection,onset} + t^i_{onset,reporting} \tag{2}$$

A total of 10^6 samples were drawn from each distribution and the resulting emperical distribution function $\widehat{F}_{infection,reporting}$ was obtained by applying the ecdf function[14] to the list of $t^i_{infection,reporting}$.

3.2 Time delay from reporting to inclusion in the $R_{t,\tau}$ estimation $t_{reporting,estimation}$

To derive the time-delay from reporting to estimation $t_{reporting,estimation}$ we need to consider the method used for the $R_{t,\tau}$ estimation in detail. The mean of $R_{t,\tau}$ given the the parameters a and b of the Gamma distribution prior is given by Equations 3 and 4.[10]

$$\Lambda_t = \sum_{s=1}^{t} I_{t-s} w_s \tag{3}$$

$$\bar{R}_{t,\tau} = \left(a + \sum_{s=t-\tau+1}^{t} I_s \right) \left(\frac{1}{b} + \sum_{s=t-\tau+1}^{t} \Lambda_s \right)^{-1} \tag{4}$$

Modeling of the time delay was performed by investigating the propagation of a time pulse through the estimation function defined in Equation 4. This time pulse was modeled as a dirac δ function at $t = 0$. The estimation process is discretized such that $t \in \mathbb{N}$ therefore the discrete number of cases for the dirac δ is given in Equation 5.

$$I_\delta(t) \colon \mathbb{N} \to \mathbb{R} \text{ such that } \begin{cases} I_\delta(t) = 1 & \text{for } t = 0 \\ I_\delta(t) = 0 & \text{otherwise} \end{cases} \tag{5}$$

3.2.1 Validation using a Compartmental Model

Quality control and validation of the derived time delay from reporting to estimation $t_{reporting,estimation}$ was performed using a simple compartmental SEIR model[15] with parameters chosen not related to COVID-19 and without vital dynamics.
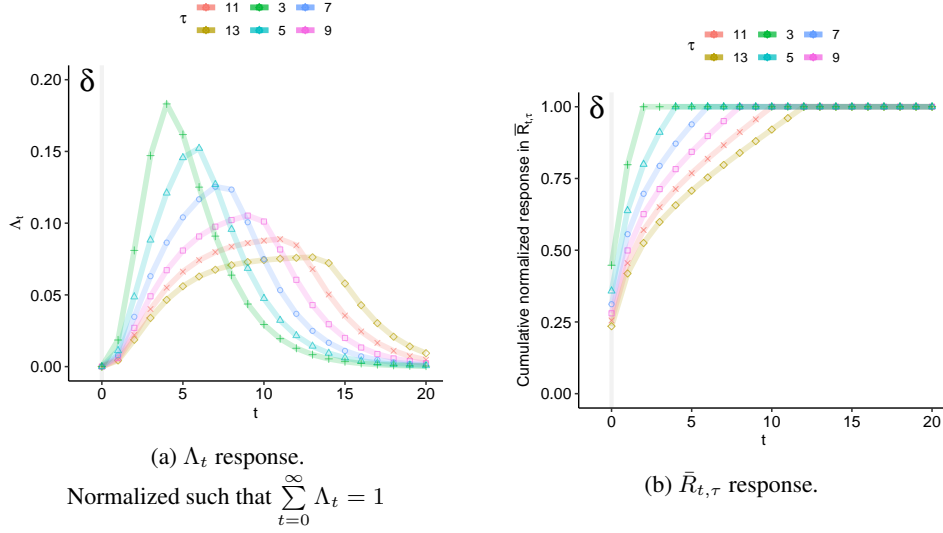
$$\frac{dS}{dt} = -\beta \frac{IS}{N} \tag{6}$$

Estimation and Interactive Visualization of the Time-Varying Reproduction Number R_t and the Time-Delay from Infection to Estimation

A PREPRINT

(a) Λ_t response.
Normalized such that $\sum\limits_{t=0}^{\infty} \Lambda_t = 1$

(b) $\bar{R}_{t,\tau}$ response.

Figure 4: Responses to a dirac δ time pulse in Λ_t and in the $R_{t,\tau}$ estimation.

$$\frac{dE}{dt} = \beta \frac{IS}{N} - \sigma E \tag{7}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{8}$$

$$\frac{dR}{dt} = \gamma I \tag{9}$$

For simulating a change in transmission the transmission rate β has been given a time dependency $\beta(t)$ with a step function, modeled using a logistic function to keep the model continuously differentiable. The model can change between two different transmission rates β_0, before t_{swap} and β_1 after t_{swap}. The steepness of the curve, k, was set to 20.

$$\beta(t) = \frac{\beta_0}{1 + e^{k(t-t_{swap})}} + \frac{\beta_1}{1 + e^{-k(t-t_{swap})}} \tag{10}$$

4   Data Sources

Case data for Austria are obtained from a GitHub repository maintained by the Complexity Science Hub Vienna based upon the official data published by the Sozialministerium. This dataset contains the case numbers for the 3 pm CEST closing time.

4.1   Data Correction

Some days in the Austrian states contain a negative number of cases on a given day. Those negative numbers are corrected by adding the negative correction to the previous day and adjusting the day containing the negative number to 0, keeping the total sum of cases constant.

4.2   Data Quality Issues

Data published by the Austrian Sozialministerium combines data into bins based upon different dates, depending on the availability of the information for each given case.[16] Currently cases are assigned to a bin by the date in the following order: Symptom onset, laboratory confirmation, reporting. This could potentially have a big impact on the estimation of $R_{t,\tau}$ and also on the time-delay involved. It is currently unknown how many of the data entries binned using which

date. Estimates of $R_{t,\tau}$ published by AGES[1] therefore differ from the reproduction of their method. Comparing the results of the estimation using this dataset and the AGES estimation suggests that there might be a limited difference between those datasets, see Figure 3a.

References

[1] Lukas Richter, Daniela Schmid, Ali Chakeri, Sabine Maritschnik, Sabine Pfeiffer, and Ernst Stadlober. Epidemiologische Parameter des COVID19 Ausbruchs - Update 24.04.2020, Österreich, 2020. page 5, 2020.

[2] Klaske van Heusden Greg Stewart and Guy A. Dumont. How Control Theory Can Help Us Control COVID-19 - IEEE Spectrum, May 2020. Library Catalog: spectrum.ieee.org.

[3] Stephen M. Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science*, April 2020. Publisher: American Association for the Advancement of Science Section: Report.

[4] Steven A. Frank. Time Delays. In Steven A. Frank, editor, *Control Theory Tutorial: Basic Concepts Illustrated by Software Examples*, SpringerBriefs in Applied Sciences and Technology, pages 95–102. Springer International Publishing, Cham, 2018.

[5] Rifat Sipahi, Silviu-Iulian Niculescu, Chaouki Abdallah, Wim Michiels, and Keqin Gu. Stability and Stabilization of Systems with Time Delay. *Control Systems, IEEE*, 31:38–65, March 2011.

[6] Lukas Richter, Daniela Schmid, and Ernst Stadlober. Methodenbeschreibung für die Schätzung von epidemiologischen Parametern des COVID19 Ausbruchs, Österreich. page 4.

[7] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013.

[8] Anne Cori. *EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves*, 2019. R package version 2.2-1.

[9] Lukas Richter, Daniela Schmid, Ali Chakeri, Sabine Maritschnik, Sabine Pfeiffer, and Ernst Stadlober. Schätzung des seriellen Intervalles von COVID19, Österreich. page 3, 2020.

[10] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics: Web material. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013.

[11] Sozialministerium. COVID-19: Alle verfügbaren Dateien (gezippt) - data.gv.at. Library Catalog: www.data.gv.at.

[12] Wissen Aktuell: Epidemiologische Parameter des COVID19 Ausbruchs, Österreich, 2020.

[13] Juanjuan Zhang, Maria Litvinova, Wei Wang, Yan Wang, Xiaowei Deng, Xinghui Chen, Mei Li, Wen Zheng, Lan Yi, Xinhua Chen, Qianhui Wu, Yuxia Liang, Xiling Wang, Juan Yang, Kaiyuan Sun, Ira M Longini, M Elizabeth Halloran, Peng Wu, Benjamin J Cowling, Stefano Merler, Cecile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases*, page S1473309920302309, April 2020.

[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[15] Julie C. Blackwood and Lauren M. Childs. An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5(1):195–221, December 2018.

[16] Erläuterungen zum Amtlichen Dashboard COVID19, April 2020.