Tracking R of COVID-19 using Kalman Filter

Navami Kairanda, Pankhuri Vanjani

Universität Des Saarlandes

1   Introduction and Problem Statement

R or the effective reproductive number is defined as the average number of secondary cases produced by a primary case. Estimation of R is important for detecting changes in disease transmission over time. During the COVID-19 pandemic, public health officials are using R to assess the effectiveness of non-pharmaceutical interventions and to frame public polices. However, estimation of R from available data presents several challenges. In this work, we employ Kalman filter to accurately estimate and track R using SIR model, [4] where the key idea is that R is linearly related to the growth rate of the number of infected individuals.

2   Background Information

SIR Model An SIR model [4] is an epidemiological model that computes the number of people infected with a contagious disease in a closed population over time. The model mathematically relates the number of susceptible people S_t, number of people infected I_t and the number of recovered. The growth rate of infected and effective reproduction number R_t can both be estimated using the model as described in [1]

Median Filter Assuming that the noise in time series is additive and uncorrelated to the signal, median filtering on chosen window size achieves smoothing of signal noise. In [5], noise reduction coefficient is determined for Gaussian distributed noise by computing the variances of original signal and signal after median filtering

Kalman Filter Kalman Filter [3,2] is one of the most common object tracking algorithm. The Kalman Filter produces estimates of hidden variables based on inaccurate and uncertain measurements. The Kalman Filter can as well provide a prediction of the future system state, based on the past estimations.

Linear Regression Linear regression is a basic machine learning model that fits a line to the input data. A simple linear regression can model linear relationship between single input variable and a response.
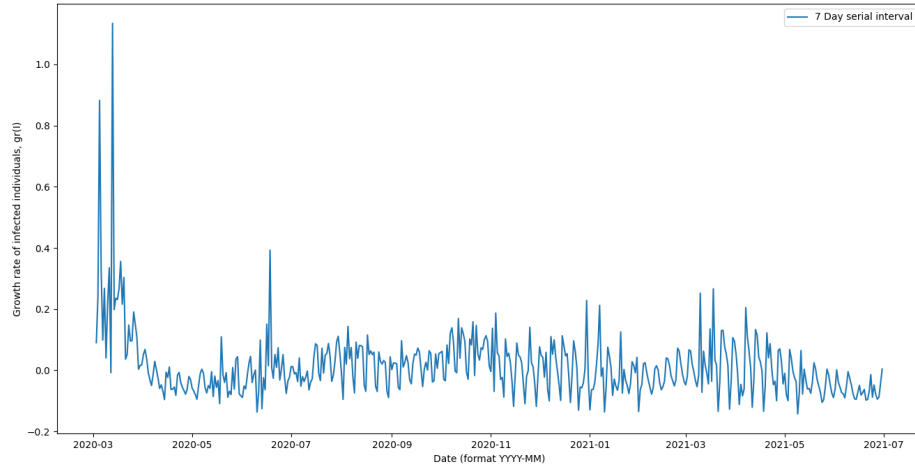
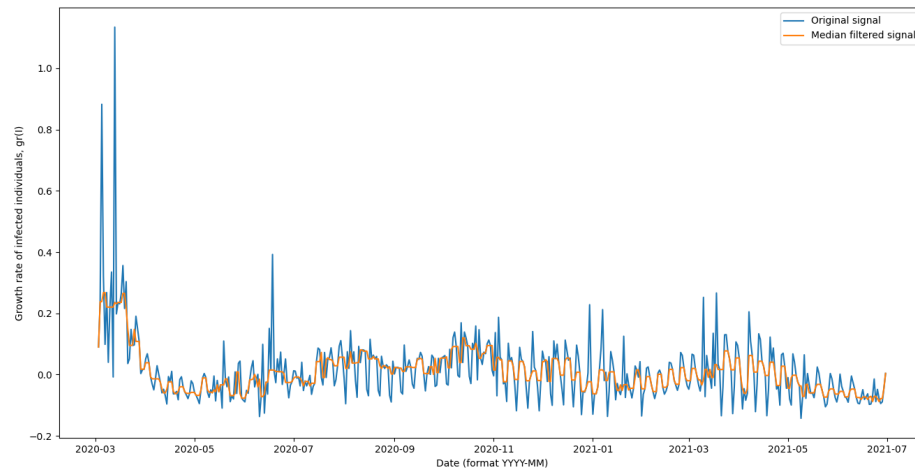**Fig. 1.** Time-series growth rate of the infected individuals from the provided time-series input data

Fig. 2. Median filter on the time-series growth rate of the infections smooths the signal by filtering additive noise

3   Methodology and Experiments

Daily growth rate in the number of infections Daily growth rate (gr_t) is expressed in terms of evolution of Infected operations (I_t) as:

$$gr\left(I_t\right) = \frac{I_t - I_{t-1}}{I_{t-1}} \tag{1}$$

From the Standard SIR model, I_t is described by the equation:

$$I_t = I_{t-1} + \beta_t I_{t-1}\frac{S_{t-1}}{N} - \gamma I_{t-1} \tag{2}$$

which can be simplified to

$$I_t = I_{t-1}\left(1 + \beta_t\frac{S_{t-1}}{N} - \gamma\right) - (1) \tag{3}$$

Using above form in $gr\left(I_t\right) = \frac{I_t - I_{t-1}}{I_{t-1}}$, we get

$$gr\left(I_t\right) = 1 + \beta_t\frac{S_{t-1}}{N} - \gamma - 1 \tag{4}$$

Simplifying the above expression gives

$$gr\left(I_t\right) = \frac{I_{t-1}\left(1 + \beta_t\frac{S_{t-1}}{N} - \gamma\right) - I_{t-1}}{I_{t-1}} \tag{5}$$

So, the final expression for growth rate in terms of infections comes out to be:

$$gr(I_t) = \beta_t\frac{S_{t-1}}{N} - \gamma) \tag{6}$$

Estimation of the noise variance using median filter We assume the noise in growth rate measurement is Gaussian distributed and uncorrelated to the signal. For this, the median filtering method [5] is used to filter and estimate the noise variance as $\sigma_\beta^2 = \frac{2N}{2N-\pi}\left(V_0 - V_{med}\right)$ where V_0 is the variance of measured growth rate of infections and V_{med} being the variance after median filtering the signal. The corresponding noise in R can then be computed as $\sigma_\epsilon^2 = \frac{1}{\gamma}\sigma_\beta^2$ as $R = 1 + \frac{1}{\gamma}gr\left(I_t\right)$.

Estimation of R using Kalman Filter We specify the following state-space model for the effective reproduction number R as given in [1]

$$R_t^{mdl} = R_{t-1}^{mdl} + \eta_t \quad \eta_t \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$$

Measurements of the system can also be performed using

$$R_t^{mes} = R_{t-1}^{mes} + \epsilon_t \quad \epsilon_t \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$$

where $R_t^{mes} = 1 + \frac{1}{\gamma}\hat{gr}\left(I_t\right)$ and $\hat{gr}\left(I_t\right) = \frac{I_t - I_{t-1}}{I_{t-1}}$ is computed using JSSE data
The Kalman filter equations for the prediction stage are

$$\hat{R}_{t|t-1} = \hat{R}_{t-1|t-1}$$
$$P_{t|t-1} = P_{t-1|t-1} + \sigma_\eta^2$$

The Kalman measurement update equations are given by

$$K_t = P_{t|t-1}\left(P_{t|t-1} + \sigma_\epsilon^2\right)^{-1}$$
$$\hat{R}_{t|t} = \hat{R}_{t|t-1} + K_t\left(R_t^{mes} - \hat{R}_{t|t-1}\right)$$
$$P_{t|t} = P_{t|t-1} - K_t P_{t|t-1}$$

Here, the system is simplified as the state transition matrix and measurement are both one-dimensional identity matrices. We also initialize the system with following values $\hat{R}_{0|0} \sim \mathcal{N}\left(0.25, 0.15^2\right)$, $\quad P_{0|0} = 0.15^2 \quad$ and $\quad \sigma_\eta = \frac{\sigma_\epsilon}{2}$
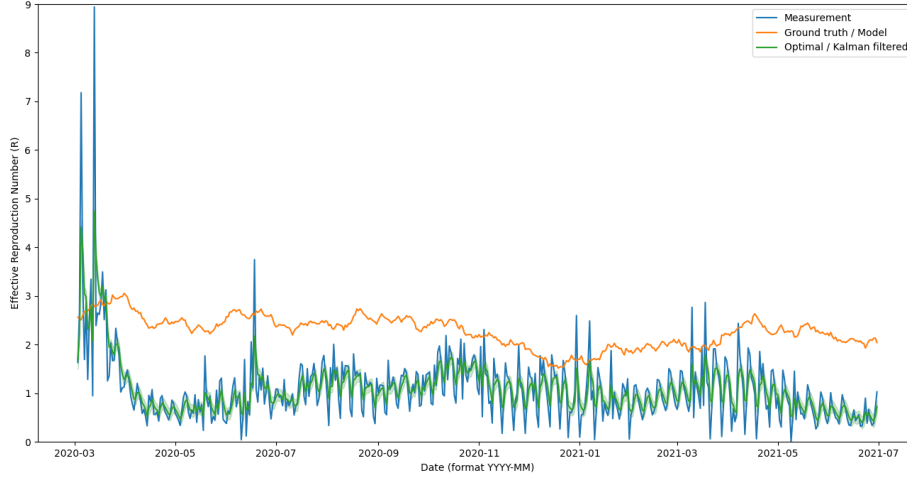

Fig. 3. The effective reproduction number R (shown with 95 % CI) is estimated using Kalman filter. The Kalman method acts as a smoothing filter on the noisy observations

4   Results and Discussion

Daily growth rate in the number of infections is visualized in Fig 1. It shows a sharp increase during the onset of COVID-19 and stabilizes over time
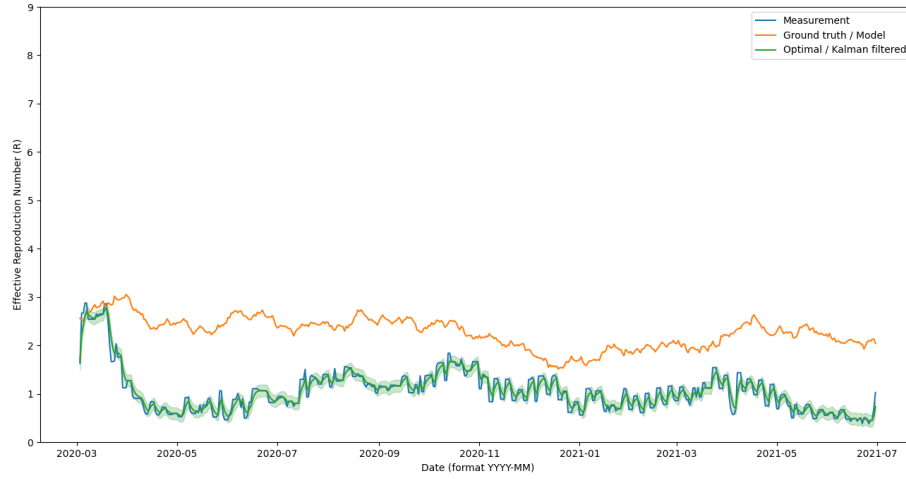
Fig. 4. A smoother estimation of R (shown with 95 % CI) can be obtained with Kalman method if median filtered signal is used as measurements. Here, the estimated R is smoother than that of Fig 3
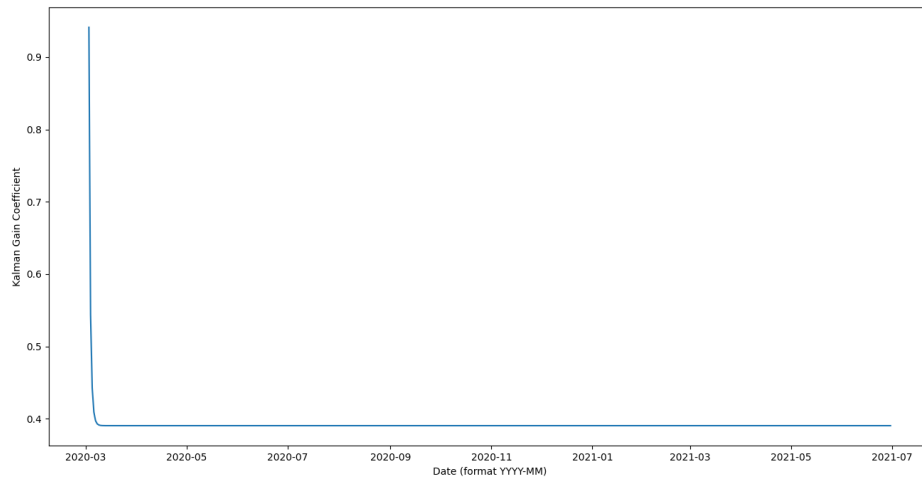

Fig. 5. The entry of Kalman gain matrix reveals that the estimation trusts the measurement better than the state-space model
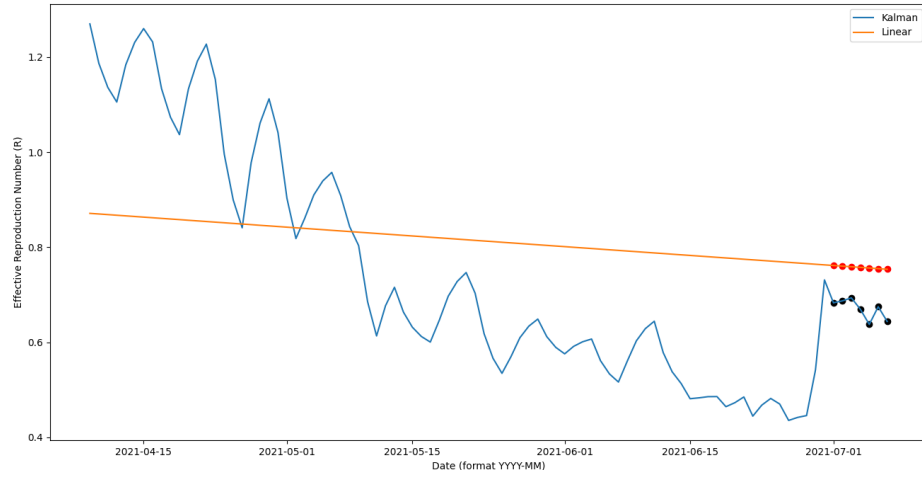
Fig. 6. The future R can be predicted for the next 7 days when the observed data is not available using Kalman filter and linear regression models

Noise variance estimated using median filter is used for modelling measurement and process variances in Kalman filter. Filtered signal is visualized in Fig 2

Estimation of R using Kalman Filter Kalman filter smooths noisy data and provides an optimal estimate of effective reproduction number R as shown in Fig 3

Smoother estimation of R can be obtained with using noise-free measurements for R. We run Kalman filter on median filtered growth rate of infections and results are shown in Fig 4. In comparison to Fig 3, where no such smoothing was performed the optimal R in Fig 4 is more stable over time.

Comments on state - space dynamics representation The state-space dynamics used in [1] does not effectively represent the reality. Such a model assumes that R does not show any trend, but rather constant over time except for random noise. This is also evident from visualization of Kalman gain coefficients in Fig 5. Kalman filter trusts measurement over the state-space model, hence acts more of a smoother of noisy measurements.

Prediction of future R can be performed using the state-space model defined with Kalman filter and is visualized for the future 7 dates in Fig 6. We also note that a linear regression model can better capture the trend in R as visualized in Fig 6

Basic reproduction number R_0 is the average number of individuals infected by a single infectious individual when the population is fully susceptible. We estimate R_0 is by the average value of R_t in the first week of the epidemic and the value determined for Germany is 2.41

5   Conclusion

In the current work, we developed a method to track R based on SIR model. We first computed the growth rate of the number of infected individuals using given data and corresponding R was computed using linear relationship. Further, median filtering and Kalman filter was applied to obtain smooth measurement. We also show ways of predicting future R or R for missing observations using the Kalman and linear regression model.

References

1. Arroyo-Marioli, F., Bullano, F., Kucinskas, S., Rondón-Moreno, C.: Tracking r of covid-19: A new real-time estimation using the kalman filter. PLOS ONE **16**(1), 1–16 (01 2021). https://doi.org/10.1371/journal.pone.0244474, https://doi.org/10.1371/journal.pone.0244474
2. Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods (08 2001)
3. Faragher, R.: Understanding the basis of the kalman filter via a simple and intuitive derivation [lecture notes]. IEEE Signal Processing Magazine **29**(5), 128–132 (2012). https://doi.org/10.1109/MSP.2012.2203621
4. Kermack, W., McKendrick, A.: Contributions to the mathematical theory of epidemics—i. Bulletin of Mathematical Biology **53**(1), 33–55 (1991). https://doi.org/https://doi.org/10.1016/S0092-8240(05)80040-0, https://www.sciencedirect.com/science/article/pii/S0092824005800400
5. Peksinski, J., Mikolajczak, G.: Estimation of the noise variance in time series using a median filter. Computer Applications in Electrical Engineering **Vol. 12**, 316–323 (2014)