Technical note ICDS043020
confidential distribution

Corona hospitalisation SIR model

Kurt Barbé, Susanne Blotwijk en Wilfried Cools

Please refer to this note as: K. Barbé, Data-driven epidemiological model to monitor the sustainability of hospital care, VUB Covid19 Technical Note No. ICDS043020, 2020

1. Analytical approximate solution for the SIR-equations

The goal is to find an approximate analytical solution for the SIR-equations which allows a stable identification of its parameters with a limited number of observations. The approximation should be accurate enough around the inflexion point of the removed compartment in order to describe the peak value of the number of infected appropriately.

The analysis is based on the SIR equations:

$$\begin{aligned}\frac{d}{dt}S(t) &= -\beta I(t)S(t)\\ \frac{d}{dt}I(t) &= \beta I(t)S(t) - \gamma I(t)\\ \frac{d}{dt}R(t) &= \gamma I(t)\end{aligned}$$

We start by replacing the first equation by the ratio of equation 3 and equation 1 leading to:

$$\begin{aligned}\frac{dS(t)}{dR(t)} &= -\frac{\beta}{\gamma}S(t)\\ \frac{d}{dt}I(t) &= \beta I(t)S(t) - \gamma I(t)\\ \frac{d}{dt}R(t) &= \gamma I(t)\end{aligned}$$

This allows an implicit solution for the first equation

$$\begin{aligned}S(t) &= \exp\left(-\frac{\beta}{\gamma}R(t)\right)\\ \frac{d}{dt}I(t) &= \beta I(t)S(t) - \gamma I(t)\\ \frac{d}{dt}R(t) &= \gamma I(t)\end{aligned}$$

where we used that $R(0) = 0$ and $S(0) = 1$. Next we replace the second equation by the ratio of equation two and three:

$$\begin{aligned}S(t) &= \exp\left(-\frac{\beta}{\gamma}R(t)\right)\\ \frac{dI(t)}{dR(t)} &= \frac{\beta}{\gamma}S(t) - 1 = \frac{\beta}{\gamma}\exp\left(-\frac{\beta}{\gamma}R(t)\right) - 1\end{aligned}$$

$$\frac{d}{dt}R(t) = \gamma I(t)$$

Also the second equation can be integrated as a function of $R(t)$ leading to the system of equations:

$$S(t) = \exp\left(-\frac{\beta}{\gamma}R(t)\right)$$
$$I(t) = 1 + I(0) - \exp\left(-\frac{\beta}{\gamma}R(t)\right) - R(t)$$
$$\frac{d}{dt}R(t) = \gamma I(t)$$

As a result all compartments are written implicitly through $R(t)$ as

$$S(t) = \exp\left(-\frac{\beta}{\gamma}R(t)\right)$$
$$I(t) = 1 + I(0) - \exp\left(-\frac{\beta}{\gamma}R(t)\right) - R(t)$$
$$R(t) = \gamma(t + tI(0)) - \gamma\int_0^t \exp\left(-\frac{\beta}{\gamma}R(s)\right) + R(s)ds$$

We look at the final equation only which we can study as a second order differential equation:

$$\frac{d^2}{dt^2}R(t) = \gamma\frac{d}{dt}I(t) = \beta\exp\left(-\frac{\beta}{\gamma}R(t)\right)\frac{d}{dt}R(t) - \gamma\frac{d}{dt}R(t)$$

Now, we linearize the differential equation around its point of inflexion t_0 – Note that in the original work of Kermack-McKendrick an approximation of a similar type was performed but around the start of the pandemic which reduces the solutions to logisttic growth models -which implies

$$0 = \beta\exp\left(-\frac{\beta}{\gamma}R(t_0)\right) - \gamma \Leftrightarrow R(t_0) = \frac{\gamma}{\beta}\log\left(\frac{\beta}{\gamma}\right)$$

Let us consider a first order Taylor approximation for $\exp\left(-\frac{\beta}{\gamma}R(t)\right)$ around t_0:

$$\exp\left(-\frac{\beta}{\gamma}R(t)\right) \approx \exp\left(-\frac{\beta}{\gamma}R(t_0)\right) - \frac{\beta}{\gamma}\exp\left(-\frac{\beta}{\gamma}R(t_0)\right)\dot{R}(t_0)(t - t_0)$$
$$= \frac{\gamma}{\beta} - \dot{R}(t_0)(t - t_0)$$

This simplifies the equation to:

$$\frac{d^2}{dt^2}R(t) = -\beta\dot{R}(t_0)(t - t_0)\frac{d}{dt}R(t)$$

Using the substitution $u = \frac{d}{dt}R(t)$ we find:

$$\frac{d}{dt}u = -\beta\dot{R}(t_0)(t - t_0)u(t)$$

leading to

$$u(t) = u(t_0)\exp\left(-\frac{\beta}{2}\dot{R}(t_0)(t - t_0)^2\right)$$

As a result we obtain an approximate solution for $I(t)$ which immediately solves the remaining signals:

$$S(t) = \exp\left(-\beta\int_0^t I(s)ds\right)$$
$$I(t) = I(t_0)\exp\left(-\frac{\beta\gamma}{2}I(t_0)(t - t_0)^2\right)$$
$$R(t) = \gamma\int_0^t I(s)ds$$

A linearization around the point of inflexion for $R(t)$ makes a Gaussian approximation of the number of infected. Interpreting as a scaled Gaussian density leads to:

$$\mu = t_0$$
$$s^2 = \frac{1}{\beta\gamma I(t_0)}$$
$$M = I(t_0)$$
$$\mathrm{A} = \sqrt{\frac{2\pi I(t_0)}{\beta\gamma}}$$

which can be solved to:

$$t_0 = \mu$$
$$\beta\gamma = \frac{2\pi M}{A^2}$$

Thus we obtain that:

$$R_0 = \frac{2\pi MN}{R_\infty^2}$$

To compare the analytical approximation to the numerical solutions of the SIR model, we apply a simulation example which realistically exhibits the current hospitalisation flow of the Covid19-pandemic. We use the following parameters $R_0 = 3.2617$, $\gamma = 0.1194$, $\beta = 0.3894$, $N = S(0) = 17\,496$, $R(0) = 0$ and $I(0) = 2$. This gives the peak moment $t_0 = 38$ days.

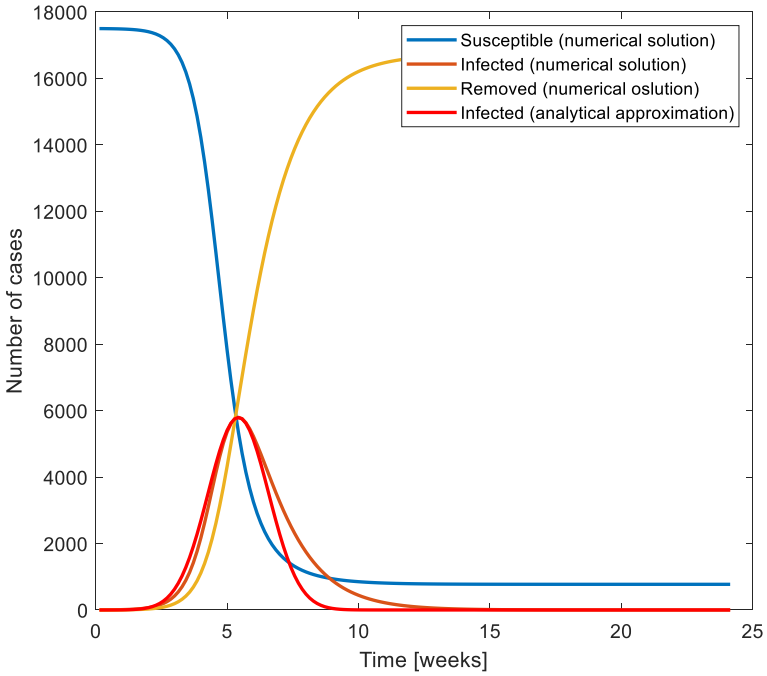

Figure 1: Numerical solution of the SIR equations confronted with the approximate solution

2. Numerical methods and parameter identification

Although the approximate solution is a Gaussian density for the number of infected, an immediate identification of its parameters does not lead to high quality estimates. A standard curve fitting is prone to errors outside of its observation window which will introduce a bias on the parameters due to the extrapolation. On top of that, such an approach would treat the observation noise as independent which is not an accurate assumption. It is important that the differential equation itself is identified such that the solution is not parametrized but the differential equation. This approach is standard in system identification as it is well known for dynamical systems that the solution if parametrized is typically badly conditioned while the differential equation is well conditioned. We use the numerical schemes developed in the reference section although this problem is far simpler than what the references deal with.

We identify the parameters based on the time series of the infected compartment (i.e. this can be for instance the number of beds occupied in a hospital among other choices)

Since the analytical solution is a Gaussian density, we obtain:

$$I(t) = I(t_0)\exp\left(-\frac{\beta\gamma}{2}I(t_0)(t-t_0)^2\right) = \theta_0\exp(\theta_1 t - \theta_2 t^2)$$

where a reparametrization is used to obtain a standard canonical polynomial. The initial estimates are obtained by using the time series of ratios

$$\log\left(\frac{I(t+1)}{I(t)}\right) = \theta_1 - \theta_2(2t+1)$$

This can be easily estimated through a linear regression.

Inspecting the residuals of the fit in Figure 2, reveals reasonably Gaussian residuals. On top of that the Fourier coefficients of the residuals show a spectral density (expressed in log-scale) which passes the visual whiteness test. The whiteness test uses that Fourier coefficients of the residuals should follow the distribution $X(\omega) \sim \frac{\sigma^2}{2}\chi_2^2$ for $\omega \neq 0$. It shows the mean value and its 95% confidence bounds of the resulting scaled chi-squared distribution.

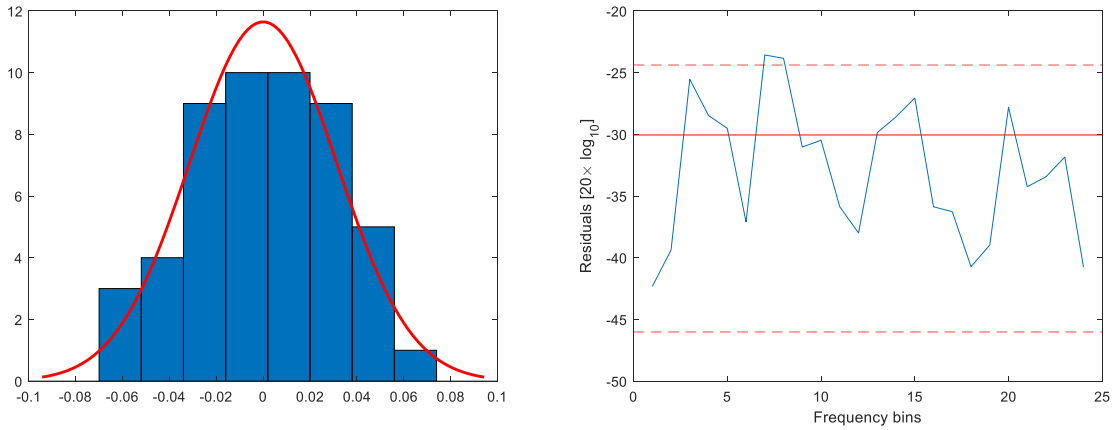

Figure 2: Residual diagnostics – empirical distribution (left), Fourier coefficients with confidence bounds (right)

As a result, in order to identify $\theta_1, \theta_2, \gamma$ simultaneously the following regression is computed:

$$\arg\min_{\theta_1,\theta_2}(1-\lambda)\sum_{i=0}^{N-1}\left(\log\left(\frac{I(t+1)}{I(t)}\right)-\theta_1-\theta_2(2t+1)\right)^2+\lambda(R(t+1)-R(t)-\gamma\, I(0)\exp(\theta_1 t-\theta_2 t^2))^2$$

This can be initialised by solving the minimization for $\lambda = 0$ or the linear regression of $I(t)$ only versus the one where $\lambda = 1$ is chosen to initialise γ. Any other choice of $\lambda \in]0,1[$ forces the parameters to match the reported data of both the infections as well as the removed time series. The weight λ can be used as weighting reflecting the relative importance of both time series with $\lambda = 0.5$ as a default.

3. Adding model flexibility

In order to model the right skewed behaviour which is not supported by the simplified model parametrisation, one may increase the degree of the polynomial. This will also improve the model fit in case the SIR parameter vary over time due to lock down measures.

As a result, the model can be any polynomial $P(t)$ of degree $k \in \mathbb{N}_0$ such that:

$$\frac{d}{dt}I(t) = P(t|\underline{\theta})I(t)$$

Additionally piecewise polynomials can be used which introduces knot points to represent changing parameters due to changes in lock down measures. To avoid risk of over-parametrisation, an increase of model complexity is currently penalized by the AIC-criterion. An increase of model complexity is chosen in case it lowers the AIC-criterion.

In Figure 3 this approach is shown on the time series of reported active cases by the federal government. For this time series, the algorithm favours a cubic polynomial over a linear polynomial which adds skewness to the time evolution.

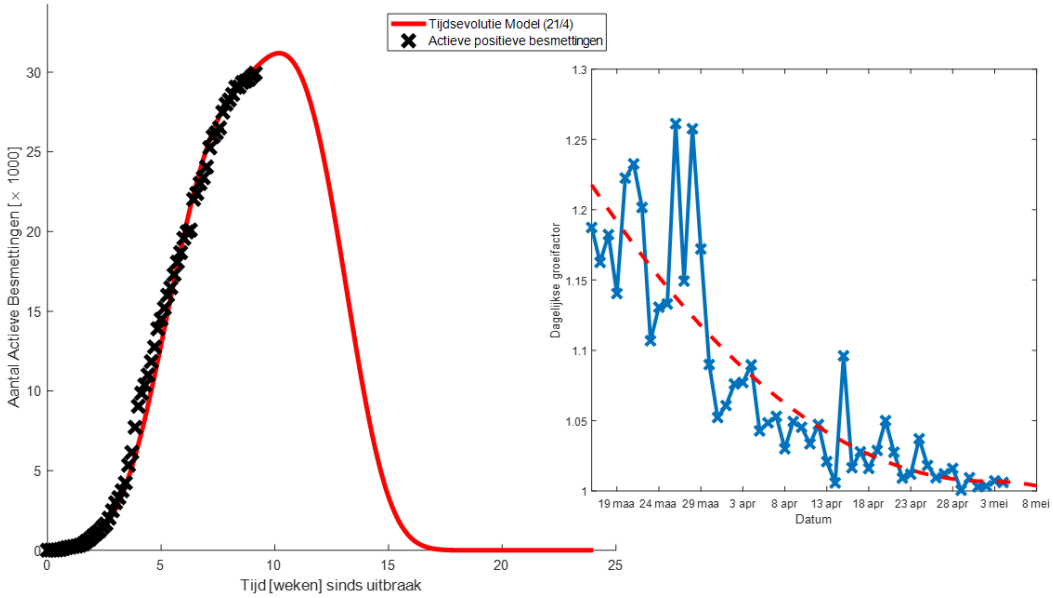

Figure 3: Number of active cases (black crosses) and the estimated time evolution by the approximate SIR equations with a cubic polynomial as a function of time – effective active cases (left) as a function of weeks since outbreak and relative growth $\frac{I(t+1)}{I(t)}$ *(right) by date.*

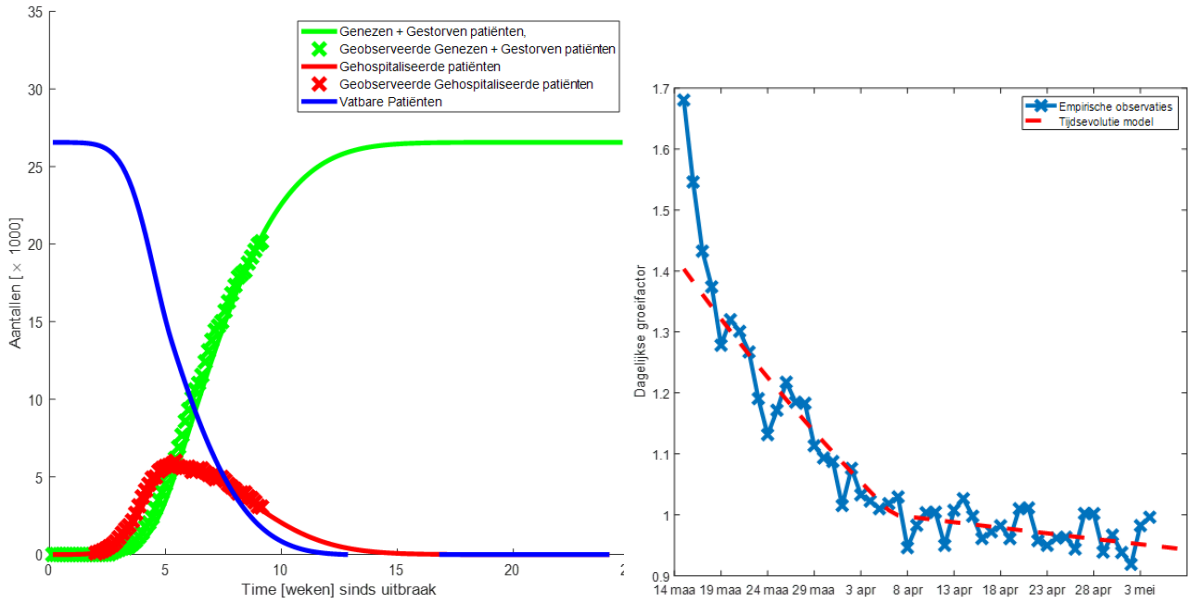

Figure 4: left plot -Number of hospitalised patients (red), modelled number of removed patients (green) and the resulting number of susceptibles (arbitrary shift). Right plot – growth factor of the number of hospitalisations.

In Figure 4 the approach is illustrated where the algorithm favours the introduction of a knot point by using two linear polynomials. The knot point was estimated on April 5th. The integration of the

hospitalisation curve with the current estimate of the parameter γ allows an estimate of the removed group. Note that the parameters γ is identified based on the removed group consisting of the number of deaths reported (death in a hospital as well as elderly homes) as well as the number of recovered patients in the hospital. The visualisation of the susceptible patients is arbitrary due to a lack of data to confront the model with. As a result, there is an unknown vertical shift as $S(0)$ is undetermined.

4. Extension of the SIR hospitalisation model

Besides modelling the number of hospitalisations, one can also model the number of ICU patients. These two are coupled as the ICU form a subset of the number of hospitalised patients. As a result the hospitalised patients are split among the ICU patients and the non-ICU patients. The model is not fully accurate as it currently does not describe ICU patients moving to a non-ICU ward and the other way around. However the ICU and non-ICU hospitalised patients are treated similarly such that it allows the same algorithm. The following equations are estimated:

$$\frac{d}{dt}S(t) = -\big(\beta_H I_H(t) + \beta_{ICU} I_{ICU}(t)\big)S(t)$$
$$\frac{d}{dt}I_H(t) = \beta_H I_H(t)S(t) - \gamma_H I_H(t)$$
$$\frac{d}{dt}I_{ICU}(t) = \beta_{ICU} I_{ICU}(t)S(t) - \gamma_{ICU} I_{ICU}(t)$$
$$\frac{d}{dt}R(t) = \gamma_H I_H(t) + \gamma_{ICU} I_{ICU}(t)$$

This approach is illustrated in Figure 5 on the current data of hospitalisations.

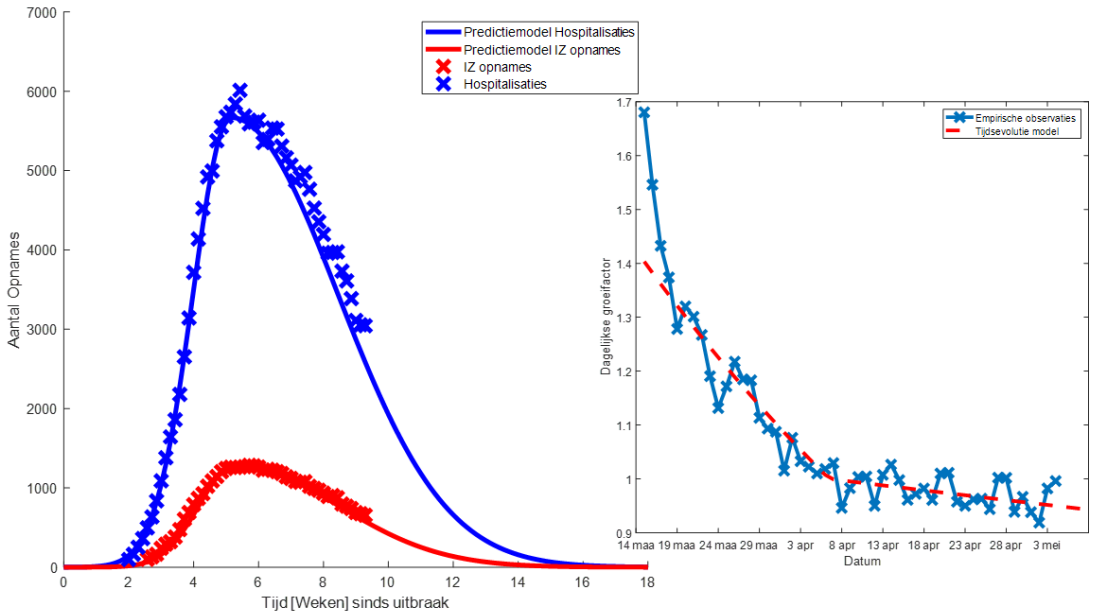

Figure 5: The approximate SIR-model estimate for $I_H(t) + I_{ICU}(t)$ in blue versus$I_{ICU}(t)$ in red modelled by a linear polynomial with a knot point on April 5th.

The approach ensures that the number of ICU patients is lower than the total number of hospitalisations while allowing differences in model fit.

Further reading

Modelling of differential equations and continuous dynamical systems

K. Barbé, J. Schoukens and R. Pintelon, "Frequency-Domain, Errors-in-Variables Estimation of Linear Dynamic Systems Using Data From Overlapping Subrecords," in *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 8, pp. 1529-1536, Aug. 2008.

K. Barbé, O. Olarte, W. Van Moer, L. Lauwers, “Fractional models for modeling complex linear systems under poor frequency resolution measurements,” in *Digital Signal Processing*, vol. 23, no. 4, pp. 1084-1093, Jul. 2013.

K. Barbe, J. Schoukens and R. Pintelon, "The Use of Nonparametric Noise Models Extracted From Overlapping Subrecords for System Identification," in *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4635-4647, Oct. 2011.

Other numerical methods of interest

G. Inghelbrecht, R. Pintelon and K. Barbé, "Large-Scale Regression: A Partition Analysis of the Least Squares Multisplitting," in *IEEE Transactions on Instrumentation and Measurement*, To appear 2020.

M. A. H. Shaikh and K. Barbé, "Wiener–Hammerstein System Identification: A Fast Approach Through Spearman Correlation," in *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 5, pp. 1628-1636, May 2019.