Technical note ICDS300420
confidential distribution

# Sars-Cov2 hospitalisation model: Time series approach

Kurt Barbé, Susanne Blotwijk en Wilfried Cools

Please refer to this note as: K. Barbé, Susanne Blotwijk en Wilfried Cools, "Data-driven epidemiological model to monitor the sustainability of hospital care", VUB Covid19 Technical Note No. ICDS300420, 2020

**Abstract**

In this technical note, a model describing the evolution of the hospitalisation wave is described. The model is based on a time series approach. The log-transformed time series of the number of events is assumed to follow a first order autoregressive process with piecewise linear drift at arbitrary points driven by a cyclo-stationary Gaussian noise process.

In this note we introduce the model, the parameter identification as well as the model choice. The model choice can be deducted through a linearization of the classical SIR-equations around the inflexion point of the solution path of the removed compartment.

## 1 Introduction

We consider a compartmental model consisting of susceptibles $\tilde{S}$ initially set to the Belgian population of 11.46 million. The model only considers a subset of the population at risk w.r.t becoming hospitalised which is denoted by $S$. With a primary focus on hospitalisations, we consider the compartments of hospitalised patients on a regular ward $I_H$ and critical care $I_{ICU}$ patients such that the union of both consist of the total hospitalised compartment. Finally, we consider the Removed compartment $R$ consisting of both discharged as well as recovered patients.

Compartmental models are the most frequently used type of dynamical system to mathematically model a disease spread [1]-[2]. Within the Belgian modelling consortium different types of these models are used with different characteristics. One can vary in these models the level of detail to increase the number of compartments as well as to allow age stratification see the UNamur model [3] or regional clustering of the UGent model [4] and UHasselt model [5]. Another alternative approach is through a meta-population model as developed at UHasselt [6]. These groups together with VUB form the Belgian modelling consortium and are partners of the FWO RESTORE[1] project.

The aphorism of George Box "All models are wrong but some are useful" implies that models are an abstraction of reality which describes a part of the measured reality such that it is interesting how consistent the models are to explore and understand possible differences.

## 2 Model description

Consider $\log(I_x(t))$ the natural logarithm of the number of patients at time instant $t$ where $x \in \{H, ICU\}$ denoting either ICU patients or non-ICU hospitalised patients. In this section, we consider the

[1] RESTORE: REalistic forecaSTing, cOntrol and. pREparedness for coming COVID-19 waves. UAntwerpen, UGent, UHasselt, UZ. Brussel

complete compartment of hospitalised patients $I(t)$. In the absence of noise, the mathematical model used is given by:

$$\log(I(t)) = a\log\big(I(t-1)\big) + f(t|\theta)$$

where $a$ indicates the autoregressive parameter while $f(t|\theta)$ is a parametrised drift term. We consider two drift models: polynomial versus a piecewise linear function. The choice between both together with the degree of the polynomial as well as the number of knot points is selected by the Akaike's Information Criterion (AIC) [7].

To allow stochastic fluctuations due to the measured number of patients, the model equation is extended to a first order autoregressive process with drift driven by a cyclo-stationary Gaussian noise process [8,9] leading to:

$$\log(I(t)) = a\log\big(I(t-1)\big) + f(t|\theta) + \eta(t)$$

with $\eta(t) = z(t) + A\ \cos(\omega t + \phi)$ such that $\phi \sim U([0,2\pi[)$ a random phase drawn from a uniform distribution over the phases of the unit circle, $\omega = \frac{2\pi}{7}\text{days}^{-1}$ the angular frequency inducing a week(end) effect and $z(t) \sim N(0,\sigma^2)$ an identically and independently distributed Gaussian noise process with zero-mean and variance $\sigma^2$. The model is assumed discrete time daily sampled $t \in \mathbb{N}$ where $t = 0$ is identified to March 1st.

This eventually leads to the following discrete time dynamical system:

$$\begin{cases} S(t) = \exp\left(-\beta_t \int_0^t I(s)ds\right) \\ \log(I(t)) = a\log\big(I(t-1)\big) + f(t|\theta) + \eta(t) \\ R(t) = \gamma_t \int_0^t I(s)ds \end{cases}$$

Note that the parameters of the linear drift term and the parameters $\beta_t, \gamma_t$ are nonlinearly coupled. The parameters $\beta_t, \gamma_t$ are constant within the duration intervals of the linear drift component.

## 3 Model deduction from the SIR-equations

The goal is to transform the classical SIR-equations into a description identifiable from a time series perspective. The representation 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:

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

This allows an implicit solution for the first equation

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

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:

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

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 logistic 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 $t_0$:

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

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:

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

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. The right skew behaviour is not supported by the symmetric solution but is corrected for by the piecewise linear drift.

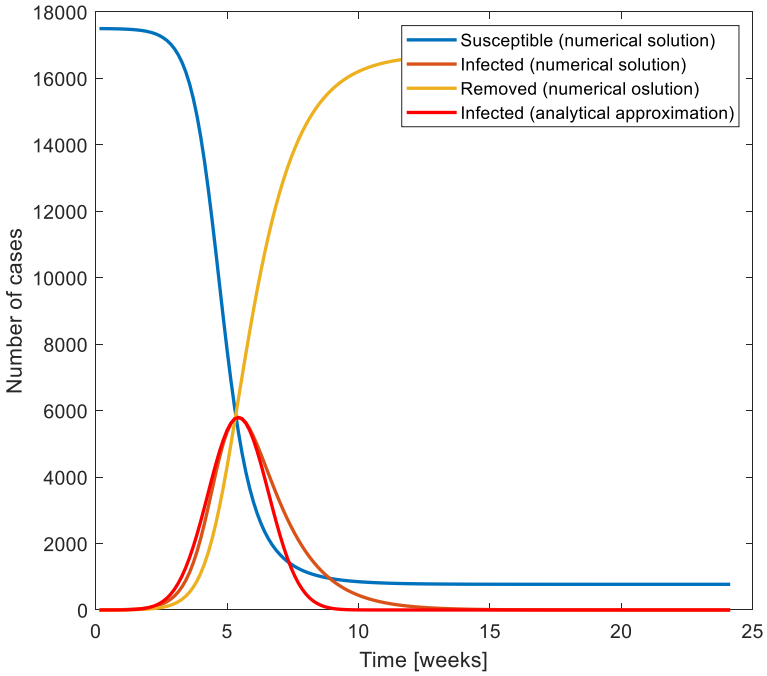


*Figure 1: Numerical solution of the SIR equations confronted with the approximate solution*

## 4 Numerical methods and parameter identification

### *4.1 Parameter identification of the time series model*

The time series model

$$\log\big(I(t)\big) = a\log\big(I(t-1)\big) + f(t|\theta) + \eta(t)$$

is linear in its parameters and can be identified by an iterated Weighted Least Squares estimation as a two-step method. We start by the Least Squares estimate followed by characterizing the residuals. The Fourier transform of the residuals at frequency $\omega = \frac{1}{7} days^{-1}$ serves as an estimate for

$$A\cos(\omega t + \phi) = \left(\frac{2}{\sqrt{N}} Re\big(P(\omega)\big)\cos(\omega t) - \frac{2}{\sqrt{N}} Im\big(P(\omega)\big)\sin(\omega t)\right)$$

with $P(\omega)$ the Fourier coefficient of the residuals at frequency $\omega$. As a result, the variable $\frac{2}{\sqrt{N}} Re\big(P(\omega)\big)\cos(\omega t) - \frac{2}{\sqrt{N}} Im\big(P(\omega)\big)\sin(\omega t)$ is used as a non-parametric noise model [10,11] to correct the least squares estimate for the weekend effect.

### *4.2 Identification of the knot points*

The knot points are the time instants when presumable a change in behaviour is expected for the drift term. The drift term consists of a slope parameter which exhibits how fast the pandemic is expanding or shrinking. Indeed $\exp(f(t|\theta))$ is a measure for the deterministic trend of relative daily growth of the observed cases. The user can introduce knot points manually corresponding to intervention points due to imposed measures. The algorithm performs a grid search around the knot point introduced in

order to make a compromise between data-fit and piecewise continuity of $f(t|\theta)$. This allows the data to take into account that measures imposed do not have an instantaneous effect in the observed data. On top of that, the algorithm will only use the knot point introduced if it lowers the AIC-statistic w.r.t. the situation without the knot point. As a result, the knot point which increases the number of parameters of the model must significantly improve the data fit to avoid overfitting.

### *4.3 Forecast*

The drift fit gives an instantaneous idea of the deterministic relative increase of the number of patients of interest through $\exp(f(t|\theta))$ leading to its immediate doubling period computed as $\log(2)/f(t|\theta)$. This leads to an effective reproduction number $R_e = \exp(\tau f(t|\theta))$ with $\tau$ the mean generation infection time [12] which assumes that the population is homogeneously mixing.

The forecast is computed by solving the SIR equations numerically with a standard 4/5 Runge-Kutta numerical scheme w.r.t the immediate population at risk consisting of $S(0) = N \times p_e \times p_{pr} \times (1 - p_i)$ where $N$ is the total population size, $p_e$ is the probability of the event (for instance hospitalisation) given a positive test, $p_{pr}$ is the probability of infection (measured by for instance the positivity rate) and $p_i$ the fraction of immunity. Since $R_e$ is estimated previously the effective $\beta_e, \gamma_e$ is estimated by considering a validation window $T$ (for instance 7 days). The forecast is tuned such that $\beta_e$ with $\gamma_e = \beta_e/R_e$ can be estimated to minimize the prediction error in the validation interval. Note that the forecast does not assume future interventions and describes a fading pandemic under the currently observed situation.

### *4.4 Uncertainty analysis*

The confidence interval of the forecast is measured by bootstrapping. We bootstrap the data as follows:

$$\log(I_b(t)) = \hat{a}\log\big(I_b(t-1)\big) + f(t|\hat{\theta}) + \eta_b(t)$$

where $\eta_b(t) = \left(\frac{2}{\sqrt{N}} Re\big(P(\omega)\big)\cos(\omega t) - \frac{2}{\sqrt{N}} Im\big(P(\omega)\big)\sin(\omega t)\right) + \epsilon(t)$ where $\epsilon(t)$ is Gaussian white noise with zero-mean and variance estimated from the residuals. As a result, each bootstrapped time series allows re-estimation of the parameters including $R_e, \beta_e, \gamma_e$. As a result for each forecasted time instant, the various solution paths are considered as the bootstrapped uncertainty assessment leading to the bootstrapped confidence interval of the forecast.

## 5 Model realisations

### *5.1 Early pandemic*

On March 14 the hospitalisations in Belgium started although the infections were in an exponential growth phase. For the university hospital, a projection was made for a potential peak moment as well as a peak load to assess hospital sustainability. The time series model was used of the measured infections since March 1st and a proxy statistic of the Italian hospitalisation rates leading to $p_h = 0.35$ and $p_{icu} = 0.07$ for the probability of hospital and ICU-admission given a positive test. The positivity rate was around 11%.

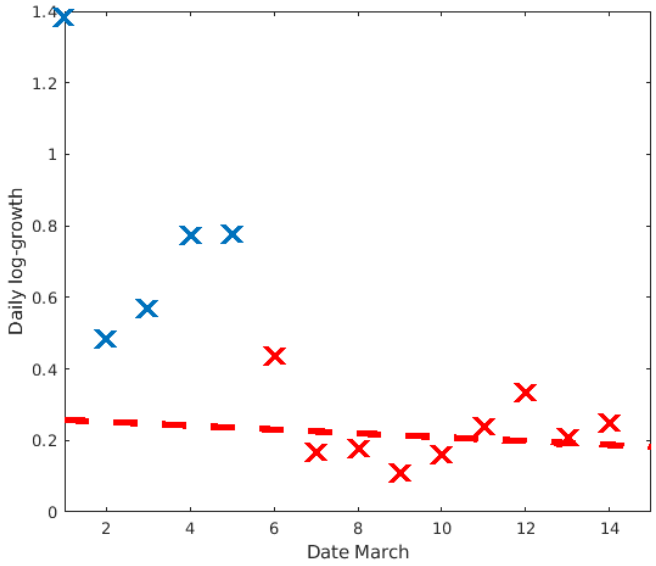


*Figure 2: Linear drift estimation of the early pandemic in Belgium for the number of positive tests (trend line – red dashed line, blue crosses – unused unstable observations, red crosses – used stable observations)*

In Figure 2 the estimate of the log-linear drift term applied to the observations in the time window March 6$^{th}$ up to March 14$^{th}$. The first observations are in transient regime converging to a steady state pattern due to the initial conditions of the dynamical system. The model estimated the doubling period to 2.33 days which is nearly half shorter than the incubation time leading to a lock down to slowing the virus spread down.

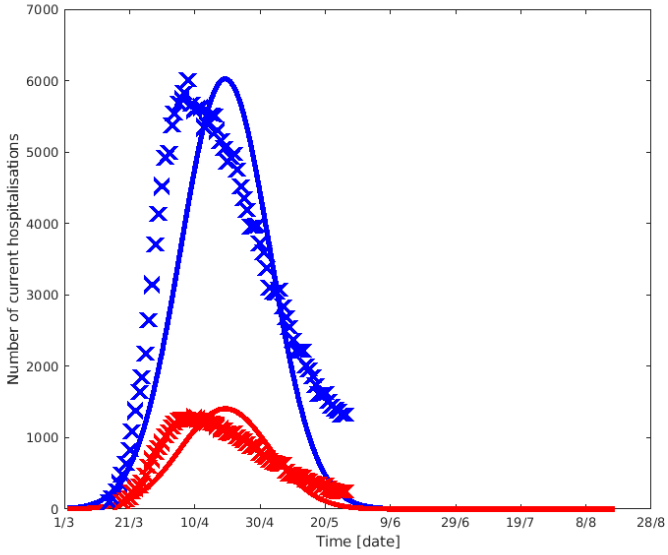


*Figure 3: Forecast of the hospitalisation wave modelled on March 14$^{th}$ and confronted with the observations up to May 25$^{th}$ – crosses (observations), solid line (model) where all hospitalisations (blue) and ICU (red)*

The model estimated the peak moment approximately 8 days later than observed. Moreover the model gives a symmetrical projection due to the model description used. The model assumes the daily

log-growth to proceed along the linear trajectory without changes. The maximum hospitalisation load was fairly well estimated which implies that the proxy probabilities used from the Italian statistics compares to the Belgian situation. As a result this suggests that the virus strikes in a similar way than in Italy. We did not see a hospitalisation wave which was too high to exceed sustainability due to the earlier lock-down and the resulting linear drift.

### *5.2 Shrinking pandemic*

While the time series become longer the model could re-calibrate where needed to obtain a better model-fit. The model's trajectory was given a first knot point on April 8th. This is normal because the model description is a linearization of the system of nonlinear differential equations governing the pandemic such that a knot point is needed at the pandemic's stationary point of the hospitalisation wave to compensate for the right skewed behaviour.

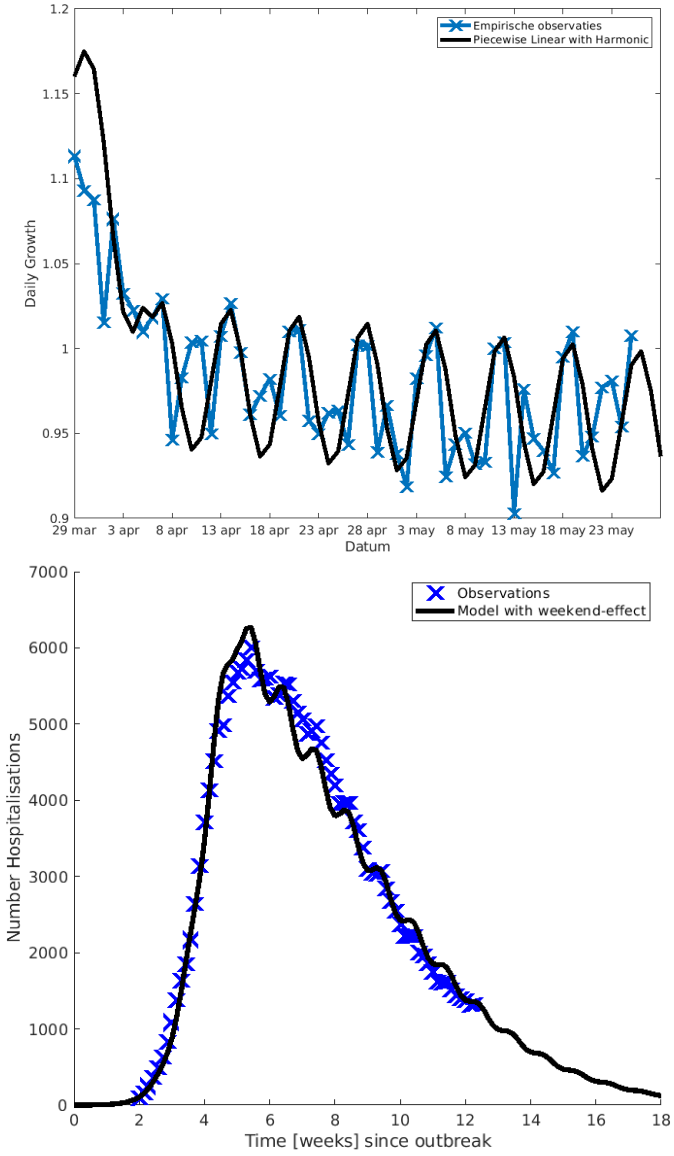


*Figure 4: Top: Linear drift with oscillating weekend effect with a knot point around April 8th – Bottom hospitalisation wave*

### *5.3 Early warnings of a potential summer increase*

In the period of June various exit strategies where introduced up to July 1st. On June 24th the model gave the first warning of a potential hospitalisation increase. The model anticipated the growing number of hospitalisations for July 18th. The model was targeting an increase of the last two weeks preceding June 24th. Particularly the minimum and maximum daily growth in these weeks were rising in this period which was a violation of the time series stationarity.

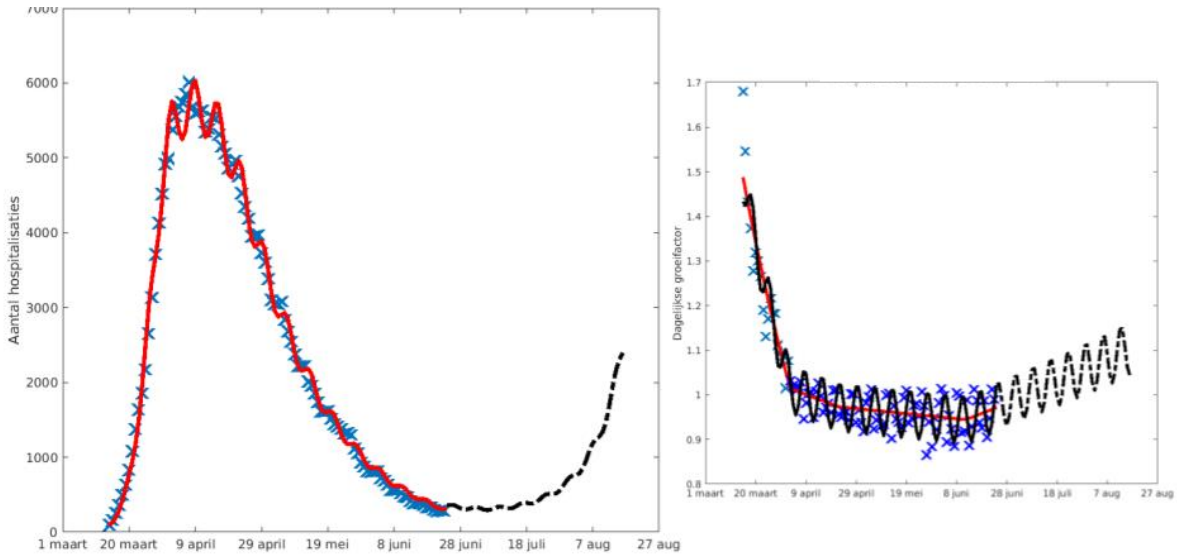


*Figure 5: Model of June 24th revealing a potential re-growing epidemic by July 18th.*

The model stayed on this course up to the July 16th when the first increase was reported. The model was targeting the relative growth which was increasing although it remained under the critical value of 1. The increasing trend of the growth was correctly followed although the absolute number kept going down.

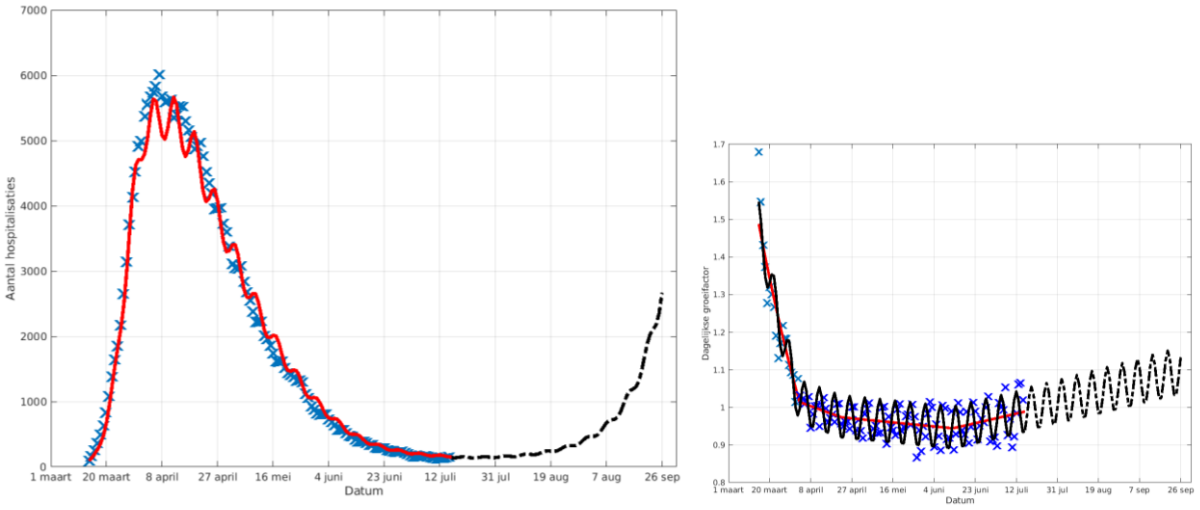


*Figure 6: Model of June 24th revealing a potential re-growing epidemic by July 18th.*

The increase in hospitalisations occurred on July 14th. Although the model could identify an increase in the relative daily growth, its forecast was quite negative. It projected a doubling period of 19 days. To reduce the risk, measures were installed end of July which finally lead to an outcome where the hospitalisations doubled during one cycle from 12th July with 130 hospitalised patients to 17th august 341 patients. This gives a rough estimate of a doubling period of 25.87 days as observed.

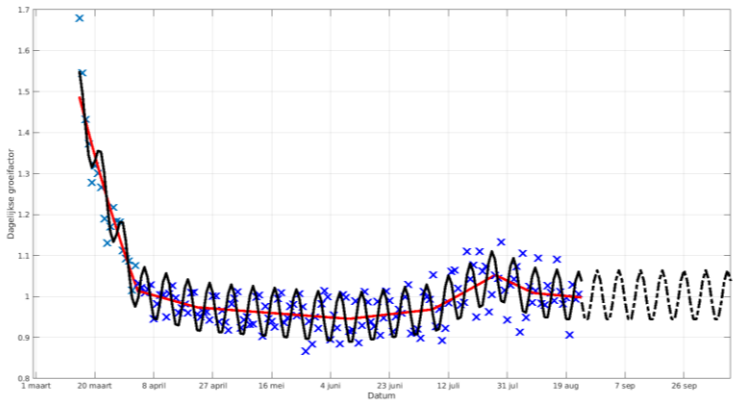


*Figure 7: Growth plot on August 24th.*

Finally the model re-calibrated when the measures imposed kicked in and gave a stable solution path for the immediate future. The model identified the change in stationarity altering the increasing trend to a decreasing trend. It is difficult to assess the validity of the negative projection but the model's negative forecast was shared by the other modelling strategies in the Belgian consortium. Nonetheless, in August evidence presented itself that the model was rather negative due to the hospitalisation probabilities used. The probabilities used for $p_h$ were still the ones regulating the March wave. However the increase of the summer period was due to the younger age groups significantly reducing the hospitalisation probability. Since the model does not take future behavioural changes into account, it cannot compensate at the time of the forecast for future knot points. As a result, the model is primarily equipped for short time projections. During the March wave the long term projection was fairly accurate since the Italian proxy gave an idea of the Lockdown effect which was by the proxy taken into account.

### *5.4 September period*

The current month shows hospitalisations which are daily rising. The doubling period on September 17th is estimated to 15.2503 days with a $R_e = 1.1994$. This is a worrying situation as the linear drift term is expect to continue its increasing slope. As a result, one expects on September 27th to observe a doubling time of 8.9104 days leading to a $R_e = 1.3650$. This implies that the doubling period becomes 41.57% shorter over 10 days. This is consistent with the exponential growth pattern whose rate changes exponentially as well.

The forecast estimates for October a hospitalisation load 1126 occupied beds with a 95% uncertainty interval of [924, 1342]. The peak is estimated by early November with an estimate of 6777 occupied beds within a large prediction uncertainty of [4215,9573]. Besides the large prediction uncertainty, efforts must be made to counter the increase observed. Such efforts have been shown to be effective in decreasing the summer growth.

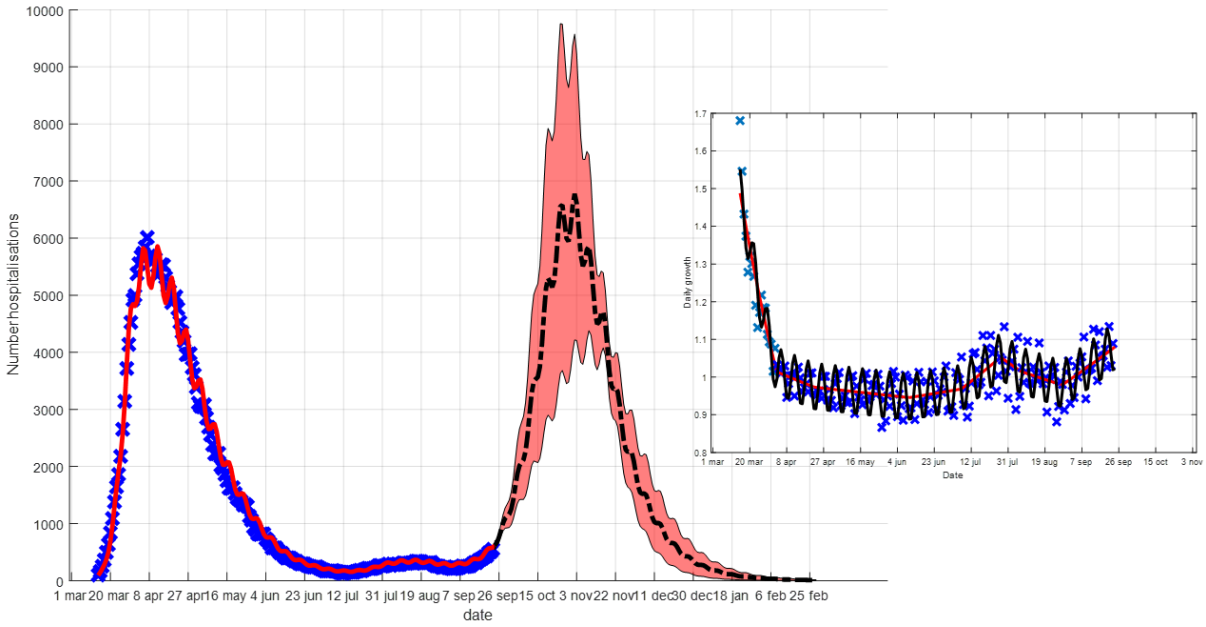


*Figure 8: Forecast determined by the model based on the Hospitalisation growth pattern observed since August 28th*

## 6 Conclusions

In this technical report a SIR-based time series approach is described which is computationally light and allows black-box modelling of an infection wave. It is data-driven and does not require a long time series to obtain insight in the potential solution path of the infection wave. Moreover it is capable of adapting to incoming changes of the data-properties by the introduction of knot points.

Forecast models for an infection wave are used to assess the impact of current observations in order to intervene and return to a user-defined path of preference since every disease spread is driven by human behaviour. Thus, models can advice human behaviour to maximally suppress diseases from spreading.

## References

[1] Daley and Gani. 1999. “Epidemic Modelling: An Introduction”, *Cambridge University Press* , UK
[2] Murray 2002. “Mathematical Biology: An Introduction”, *Springer* , Germany
[3] N. Franco, Covid-19 Belgium: Extended SEIR-QD model with nursery homesand long-term scenarios-based forecasts from school opening, technical note, University of Namur, 2020.
[4] T.W. Alleman, J. Vergeynst, E. Torfs, D.I. Gonzalez, I. Nopens and J.M. Baetens,A deter-ministic, age-stratified, extended SEIRD model for assessing the effect of non-pharmaceuticalinterventions on SARS-CoV-2 spread in Belgium, medRxiv, doi:10.1101/2020.07.17.20156034.
[5] S. Abrams, J. Wambua, E. Santermans, et al.,Modeling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories, medRxiv, doi:10.1101/2020.06.29.20142851.
[6] P. Coletti, P. Libin, O. Petrof, S. Abrams et al.,COVID-19 report on a meta-population model for Belgium: a first status report, https://www.uhasselt.be/Images/DSI/report_meta.pdf.
[7] Akaike, H. (1974), "A new look at the statistical model identification", IEEE Transactions on Automatic Control, 19 (6): 716–723.

[8] B. Porat, Digital Processing of Random Signals – Theory and Methods, Dover Publications, NY USA, 1994

[9] Giannakis, G.B. “Cyclostationary Signal Analysis” Digital Signal Processing Handbook Ed. Vijay K. Madisetti and Douglas B. Williams Boca Raton: CRC Press LLC, 1999

[10] 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.

[11] K Barbe, J Schoukens, R Pintelon, IEEE transactions on signal processing 59 (10), 4635-4647, 2011.

[12] M. G. Roberts & J. A. P. Heesterbeek, Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection, Journal of Mathematical biology, 55, 2007