BACHELOR

Growth curve modelling of infectious disease transmission
Using the Verhulst model in times of the COVID-19 pandemic

Heidema, Stan G.A.M.

Award date:
2020

Eindhoven University of Technology
Department of Mathematics and Computer Science

Growth curve modelling of infectious disease transmission

Using the Verhulst model in times of the COVID-19 pandemic

S.G.A.M. Heidema
1252305

Supervisors:
R.A.J. Post
E. R. van den Heuvel

Eindhoven, July 2020

Contents

1 Introduction

In the first half of the year 2020, the COVID-19 virus caused many problems all over the world. Such a pandemic makes people question their safety and health. Countries went on lockdown, supermarkets got hoarded and many hospitals could not handle the stream of patients. Examples of the latter can be found in the countries that are hit hardest at first, China and Italy. Therefore, predicting the number of infections while an epidemic is in its beginning stages is extremely important. Researchers van den Heuvel, Zhan and Regis from TU Eindhoven took on this challenge, and published daily predictions of the COVID-19 pandemic (see `tue.nl`). In their data driven research, the theory of non-linear regression and a logistic growth curve, introduced by Pierre François Verhulst [1], were both fundamental tools. In their technical report, it can be read that van den Heuvel and his colleagues often observed a negative error when applying the statistical Verhulst model to COVID-19 data. This thesis aims at mathematically qualifying the reason of this occurrence. Two of the simplest compartmental models, the susceptible-infected-susceptible (SIS) and the susceptible-infected-recovered (SIR) model as proposed by Kermack and McKendrick in [2], are at the basis of the argument. In particular, it is shown that two important models, first described by P. Verhulst, can be derived from the SIS model. This observation leads to two statistical models, which we refer to as the incidental Verhulst regression model and the cumulative Verhulst regression model. The argument is made that the SIR model is more adequate for modelling COVID-19 infections. We observe that both Verhulst regression models underestimate the final number of infections when applied to epidemics (stochastically) generated by the SIR model. However, there is no way to quantify this bias analytically, since the SIR model has no closed form solution. An alternative statistical model, which has its origin in the SIR model, is proposed to prevent this bias. Finally, the three proposed statistical models are applied to the reported number of COVID-19 infections in both the Netherlands and Sweden. In this section, we also make a prediction for the final number of infections in Sweden.

2 Mathematical models in epidemiology

In this chapter, two simple compartmental models, the susceptible-infected-susceptible (SIS) model and the susceptible-infected-recovered (SIR) model, are introduced. These models are first introduced by Kermack and McKendrick [2] and are of great importance in mathematical epidemiology. The process of analytically solving the SIS model (section 2.1) gives rise to two models, originally discovered by Pierre François Verhulst [1] for modelling population growth, which are defined in section 2.1.1. In addition, the three parameter Verhulst model can also be obtained, which is described in the appendix (section 6.1). These models are suitable models for modelling diseases where people become susceptible after infection. Regarding COVID-19 however, patients produce antibodies, suggesting an at least temporary immunity to the virus [3]. Besides this, there are individuals dying from the virus, which also means that they do not become susceptible after infection. The SIR model [2] is proposed in section 2.2 to capture this problem. In addition, a simulation program is set up in section 2.3 to stochastically model SIR epidemics. Analysis of more complex compartmental models like the susceptible-exposed-infected-recovered (SEIR) ([4], p. 92) model and the susceptible-infected-recovered-susceptible (SIRS) ([4], p. 93) model are outside the scope of this research.

2.1 The SIS Model

One of the simplest disease spreading models is the SIS model, where individuals are either susceptible or infected. In this model, it is assumed infected individuals become susceptible again after recovery. Consider a fixed population of N individuals and let $S(t)$ and $I(t)$ denote the number of susceptible and infected individuals at time t respectively. Since an individual is either susceptible or infected, $N = S(t) + I(t)$ for all t. Further we assume that I_0, the number of infected individuals at $t = 0$, is known and therefore also $S_0 = N - I_0$. The continuous changes in time in these groups are given by:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{bS(t)I(t)}{N} + aI(t) \\ \frac{dI(t)}{dt} &= \frac{bS(t)I(t)}{N} - aI(t) \end{aligned} \tag{1}$$

Here, a and b are non-negative parameters. One can interpret a as the rate at which individuals recover. By setting $a = 0$, we obtain a special case of the SIS model, namely the SI model, where infected individuals stay infected. Further, b can be seen as a contact rate of infected individuals, in which $\frac{S(t)}{N}$ is the probability of meeting a susceptible individual. The dimension of the system can be reduced by one, by using the fact that $S(t) = N - I(t)$ for all t. This yields the following equation, commonly referred to as the logistic equation. ([4], pp. 18-20)

$$\frac{dI(t)}{dt} = rI(t)(1 - \frac{I(t)}{K}) \tag{2}$$

where $r = b - a$ and $K = \frac{Nr}{b}$. The parameter r can be interpreted as the growth rate of the model. Note that r can be either positive or negative, so we consider two cases. A negative growth rate, $r < 0$, causes the disease to disappear after a while. To see this, note that if $r < 0$ then $K < 0$. Therefore, by equation 2:

$$I'(t) \leq rI(t)$$

Solving this differential inequality yields $I(t) = I_0 e^{rt}$, approaching zero. In other words, $I_\infty := \lim_{t\to\infty} I(t) = 0$. Hence, a negative growth rate causes the disease to gradually disappear from the population on its own. When considering a positive growth rate, $r > 0$, there exists a different closed form solution which can be derived using separation of variables.

Theorem 1. Given that $r > 0$, the solution of equation 2 in terms of the initial conditions r, K and I_0 is given by

$$I(t) = \frac{K}{1 + \left(\frac{K - I_0}{I_0}\right) \exp(-rt)} \tag{3}$$

Proof.
Separating the variables in equation 2 yields

$$\frac{1}{I(t)\left(1 - \frac{I(t)}{K}\right)} dt = rdt$$

Now, by applying a partial fraction expansion to the left hand side and integrating both sides, we obtain

$$\int \left(\frac{1}{I(t)} + \frac{1}{K - I(t)}\right) dI(t) = \int rdt$$

Therefore,

$$\log \frac{I(t)}{|K - I(t)|} = rt + C$$

For some integration constant $C \in \mathbb{R}$. It is assumed that the initial condition I_0 is given. Hence,

$$C = \log \frac{I_0}{|K - I_0|}$$

By replacement of C with the above expression and using the properties of the logarithm,

$$\log \frac{I(t)}{|K - I(t)|} - \log \frac{I_0}{|K - I_0|} = \log \left(\frac{I(t)|K - I_0|}{I_0|K - I(t)|}\right) = rt$$

Note that the absolute values can be disregarded because $K - I_0$ and $K - I(t)$ have the same sign. Next, we take the exponent to obtain

$$\frac{I(t)}{K - I(t)} = \frac{I_0}{K - I_0} \exp(rt)$$

Solving for $I(t)$, the desired result is obtained

$$I(t) = \frac{K\frac{I_0}{K-I_0}\exp(rt)}{1+\frac{I_0}{K-I_0}\exp(rt)} = \frac{K}{1+\left(\frac{K-I_0}{I_0}\right)\exp(-rt)}$$

□

2.1.1 Introduction to two Verhulst models

Let $Y(t)$ be the cumulative number of reported infections in a system at time t. Denote by $\Delta Y(t) = Y(t) - Y(t-1)$ the incidental number of reported infections at time t. Based on equation 2, the following model is introduced, referred to as the incidental Verhulst model.

Definition 1. *The incidental Verhulst model*
The expected incidental number of reported infections at time $t \geq 0$ can be assumed to follow the incidental Verhulst model, that is

$$\mathbb{E}\left[\Delta Y(t)|Y(t-1)\right] = rY(t-1)\left(1-\frac{Y(t-1)}{K}\right) \tag{4}$$

where K and r are non-zero parameters

Based on equation 3, we introduce the cumulative Verhulst model, containing two parameters. To avoid ambiguity, we set $L = K$ and $q = r$.

Definition 2. *The cumulative Verhulst model*
The expected cumulative number of reported infections at time $t \geq 0$ can be assumed to follow the cumulative Verhulst model, that is

$$\mathbb{E}\left[Y(t)\right] = \frac{L}{1+\left(\frac{L-Y(0)}{Y(0)}\right)\exp(-qt)} \tag{5}$$

where q and L are both positive parameters.

2.2 The SIR model

This model is an extension to the SIS model, as it introduces a recovered class R. Individuals belonging to the R class are considered to be immune to the virus. Note that this group also contains individuals who passed away after becoming infected. Again, we assume a fixed population of N individuals, and each person belongs to one of the three groups. Hence, $N = S(t) + I(t) + R(t)$ for all t. Another assumption made is that the quantities for each group at the initial point in time, S_0 and I_0 are known. Without loss of generality we set $R_0 = 0$. The differential equations corresponding to the changes in these groups are given by:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\frac{bS(t)I(t)}{N} \\
\frac{dI(t)}{dt} &= \frac{bS(t)I(t)}{N} - aI(t) \\
\frac{dR(t)}{dt} &= aI(t)
\end{aligned} \tag{6}$$

Here, a and b are non-negative. Similarly to the SIS model, we can interpret b as the contact rate of infected individuals, and $\frac{S(t)}{N}$ the probability of meeting a susceptible individual. Again, a corresponds to the rate at which individuals recover. However, recovery has a different definition here, namely individuals becoming immune and part of the R class. The authors of [5] obtained explicit series solutions of the SIR model with the help of the homotopy analysis method. Furthermore, in [6], an exact solution is obtained. However, this is not useful for estimating the original parameters a and b, because the solution is in parametric form.

So, there have been multiple attempts to solve the SIR model, but no closed form solution has been found. Therefore, a similar approach as for the SIS model to obtain explicit models for estimating the number of reported infections is impossible. In section 3.5, an alternative method is proposed based on numerical integration.
The final number of infections, that is $I_\infty + R_\infty = N - S_\infty$ is often of interest when analysing behaviour of epidemics. Therefore, we will look at some properties of these quantities.

2.2.1 Mathematical properties of the SIR model

Note that the number of susceptible individuals is always non-increasing, independently of the initial condition S_0. Since $S(t)$ is monotone and positive we know that

$$\lim_{t\to\infty} S(t) = S_\infty$$

Similarly, the number of recovered individuals also has monotone behaviour, independently of the initial conditions. $R(t)$ is non-decreasing and bounded by the total population size N. Therefore

$$\lim_{t\to\infty} R(t) = R_\infty$$

Such a monotonicity argument cannot be made for the number of infected individuals, which may be increasing at first to some maximum level and then decreasing to zero. Such behaviour happens for example when $\left.\frac{dI(t)}{dt}\right|_{t=0} = (bS_0 - a)I_0 > 0 \iff bS_0/a > 1$. Martcheva also shows that if

$$\lim_{t\to\infty} I(t) = I_\infty$$

then $I_\infty = 0$.
The following theorem will be useful for predicting the final number of susceptible individuals S_∞. Furthermore, R_∞ can be deduced from this equation since $R_\infty = N - S_\infty$. However, theorem 2 is only useful whenever a and b are known. In section 3.5, a method is provided to estimate these parameters from data.

Theorem 2.
When the parameters a, b and the total population size N are given, the final number of susceptible individuals, S_∞, can be expressed in the following way

$$S_\infty = -\frac{aN}{b} W\left(-\frac{bS_0}{aN}\exp\left(-\frac{b}{a}\right)\right) \tag{7}$$

Here, W is the Lambert-W function, defined as the inverse of $f(x) = x\exp x$.

Proof.
From the system of equations 6, we can obtain

$$\frac{I'}{S'} = \frac{bSI/N - aI}{-bSI/N} = -1 + \frac{aN}{bS}$$

Separating variables yields

$$I' = (-1 + \frac{aN}{bS})S'$$

Integrating leads to

$$I = -S + \frac{aN}{b}\log S + C$$

where C is an arbitrary integration constant. Rewriting, we get

$$I + S - \frac{aN}{b}\log S = C$$

The above equality holds both for (S_0, I_0) and for $(S_\infty, I_\infty) = (S_\infty, 0)$. Recall that $R_0 = 0$, and hence $N = S_0 + I_0$. Therefore, the following relation is obtained, which we will solve for S_∞:

$$I_0 + S_0 - \frac{aN}{b}\log S_0 = S_\infty - \frac{aN}{b}\log S_\infty$$
$$\exp(-\frac{b}{aN}S_\infty)S_\infty = \exp(-\frac{b}{aN}(S_0 + I_0))S_0 = \exp\left(-\frac{b}{a}\right)S_0$$
$$S_\infty = -\frac{aN}{b}W\left(-\frac{bS_0}{aN}\exp\left(-\frac{b}{a}\right)\right)$$

The last equation follows from the fact that for constants $c_1 \neq 0$ and c_2,

$$\exp(c_1 x)x = c_2 \iff x = \frac{W(c_1 c_2)}{c_1}$$

□

2.3 Stochastic SIR model

We have seen a deterministic version of the SIR model in the previous section. Yet, a stochastic SIR model is convenient for a more robust analysis. Since $R(t)$ is determined whenever $S(t)$ and $I(t)$ are, we consider a continuous time Markov chain consisting of two dimensions S and I. Observe that the Markov assumption is satisfied in the system of equations 6, because the future number of susceptible and infected individuals depends only on the number of individuals at the current time. Both $S(t)$ and $I(t)$ take values in the set $\{0, 1, \ldots, N\}$ where $S(t) + I(t) \leq N$ for all t. Hence, the cardinality of the state space equals $\sum_{k=1}^{N+1} k = \frac{1}{2}(N+1)(N+2)$. Using the same notation as in the system of equations 6, we obtain the following transition probabilities for a positive change in time Δt. [7]

$$\mathbb{P}[(S(t+\Delta t), I(t+\Delta t)) = (s+k, i+j) | (S(t), I(t)) = (s, i)] = \begin{cases} bis\Delta t/N + o(\Delta t), & (k,j) = (-1,+1) \\ ai\Delta t + o(\Delta t), & (k,j) = (0,-1) \\ 1 - (bis/N + ai)\Delta t + o(\Delta t), & (k,j) = (0,0) \\ o(\Delta t), & \text{otherwise} \end{cases}$$

Where the little o notation is short hand for functions f satisfying the following property

$$f(\Delta t) = o(\Delta t) \iff \lim_{\Delta t \to \infty} \frac{\Delta t}{f(\Delta t)} = 0$$

In other words, in a period of Δt where $(S(t), I(t)) = (s, i)$, an infection occurs with probability $bis\Delta t/N + o(\Delta t)$ and a recovery with probability $ai\Delta t + o(\Delta t)$. All states where $I(t) = 0$ are absorbing states, because the disease has been terminated. Since this is a birth-death process, the inter-event times can be modeled using the memoryless exponential distribution. The inter-event times of infections, denoted $\{U_j\}$, are independent and exponentially distributed with rate bis/N. Similarly, the inter-event times of recoveries, denoted $\{V_j\}$, are independent and exponentially distributed with rate ai. The following observations will help for setting up the simulation.

$$\min(U, V) \sim \text{Exp}(bis/N + ai) \tag{8}$$
$$\mathbb{P}[U < V] = \frac{bis/N}{bis/N + ai} \tag{9}$$
$$\mathbb{P}[V < U] = \frac{ai}{bis/N + ai} \tag{10}$$

According to (8), inter-event times are exponentially distributed with rate $bis/N + ai$. Further, equations (9) and (10) tell us that such an event is an infection with probability $\dfrac{bis/N}{bis/N + ai}$ and a recovery with probability $\dfrac{ai}{bis/N + ai}$.

3 Statistical methods

To be able to apply the Verhulst regression models to data, statistical frameworks are introduced. In section 3.1, a method for predicting the incidental number of reported infections based on a generalized linear model is proposed. Further, we introduce the method of nonlinear regression in section 3.2, which is appropriate for analysing the cumulative number of reported infections. Both methods are rather basic, but are suitable for our motivation. Other applicable statistical methods include, for example, negative binomial regression and nonlinear regression with heteroscedastic variance.
Sections 3.3 and 3.4 focus on the question: is it possible to apply the two Verhulst regression models to data generated by the SIR model without losing performance? It turns out that problems arise, which are solved by the SIR curve fitting regression model, described in section 3.5.

3.1 Generalized linear models

A generalized linear model is characterized by the underlying distribution and an equation linking the expectation of the variable of interest with a linear combination of the explanatory variables. For more information on generalized linear models, see for instance [8]. Let us consider at a specific example, where the underlying distribution of the data is assumed to be Poisson. This is sensible because the theory will be applied to the incidental number of infections $\Delta Y(t)$, which classifies as counting data. Regression studies the relationship between a variable of interest $\boldsymbol{Y} = [Y_1, \ldots, Y_n]$ with n observations and m explanatory variables $x^{(j)}$. In this model, there are $m+1$ parameters given by $\boldsymbol{\beta} = [\beta_0, \ldots, \beta_m]$. Furthermore, it is assumed that the Y_i's are independent.

Definition 3. *The Poisson regression model*
Given the assumptions above, the general Poisson regression model is given by

$$g\left(\mathbb{E}\left[Y_i\right]\right) = \beta_0 + \beta_1 x_i^{(1)} + \ldots \beta_m x_i^{(m)} \qquad\qquad Y_i \sim \text{Pois}\left(\lambda_i\right)$$

where the one-to-one function $g : \mathbb{R} \to \mathbb{R}$ is called the link function.

3.1.1 Maximum likelihood estimation

Let $\boldsymbol{y} = [Y_1, \ldots Y_n]$ denote a random vector and let the joint probability density function of the Y_i's be

$$\boldsymbol{f}\left(\boldsymbol{y}; \boldsymbol{\beta}\right)$$

The likelihood function $L\left(\boldsymbol{\beta}; \boldsymbol{y}\right)$ is algebraically the same as the density function, but the random vector $\boldsymbol{y}$ is fixed instead of the parameters $\boldsymbol{\beta}$ ([8], pp. 18-20). The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is the value $\hat{\boldsymbol{\beta}}$ that maximizes this likelihood function. That is,

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} L\left(\boldsymbol{\beta}; \boldsymbol{y}\right)$$

Often, since the logarithm is a monotone function, one equivalently maximizes the log-likelihood $\ell(\boldsymbol{\beta}, \boldsymbol{y}) = \log L(\boldsymbol{\beta}, \boldsymbol{y})$. Note that in definition 3, $Y_i \sim \text{Pois}\left(\lambda_i\right)$. In this case, it can be shown that the log-likelihood is globally concave ([9], p. 78). Therefore, the Newton-Raphson method is adequate for numerically determining the maximum.

3.2 Nonlinear regression

Again, consider a variable of interest $\boldsymbol{Y} = [Y_1, \ldots, Y_n]$ with n observations and m explanatory variables $x^{(j)}$.

Definition 4. *The nonlinear regression Model* ([10], p. 1)
The general nonlinear regression model is given by

$$Y_i = f(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)};\ \boldsymbol{\beta}) + \epsilon_i, \qquad\qquad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

Here, f is an appropriate function that depends on the explanatory variables and p parameters $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p]$. The unstructured deviations from the function f are described via the random errors ϵ_i.

Such models differ from multiple linear regression models, as the function f need not be linear in the parameters $\boldsymbol{\beta}$. This makes the class of models more general, but the theory of linear regression does not hold anymore. In particular, linear regression models have an explicit form for the unbiased linear estimator with minimal variance. This is unfortunately not the case for nonlinear regression models. Therefore, the parameters need to be estimated using numerical optimization methods.

3.2.1 Least squares estimation

To get estimates for the parameters $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p]$, one can apply the principle of least squares ([10], pp. 5-6, 10). The sum of squares of residuals SS, defined as

$$SS(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - f(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)};\ \boldsymbol{\beta}))^2 \tag{11}$$

will be minimized. Thus, given the data y_i and $x_i^{(j)}$ $(j = 1, \ldots, m)$, we choose our estimator $\hat{\boldsymbol{\beta}}$ as follows

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} SS(\boldsymbol{\beta})$$

The $n \times p$ matrix of partial derivatives is defined as $A(\boldsymbol{\beta})$ with entries

$$A_i^{(j)}(\boldsymbol{\beta}) = \frac{\partial f(x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(m)};\ \boldsymbol{\beta})}{\partial \beta_j} \qquad \text{and define} \qquad \widehat{A(\boldsymbol{\beta})} = A(\hat{\boldsymbol{\beta}})$$

Further, we estimate the constant variance of the model σ^2, as $\hat{\sigma}^2 = \dfrac{SS(\hat{\boldsymbol{\beta}})}{n-p}$. Now, an estimate of the covariance matrix $V(\boldsymbol{\beta})$ can be derived as $\widehat{V(\boldsymbol{\beta})} = \hat{\sigma}^2 \left(\widehat{A(\boldsymbol{\beta})}^T \widehat{A(\boldsymbol{\beta})} \right)^{-1}$.

3.3 Two regression models

Let $\{Y(t)\}_{t=0}^n$ be a set of n cumulative infections reported at equidistant time epochs. In addition, recall that $\Delta Y(t) = Y(t) - Y(t-1)$ and denote by $\{\Delta Y(t)\}_{t=1}^n$ the set of incidental infections. Using the theory of Poisson regression together with definition 1, the incidental Verhulst regression model is obtained.

Definition 5. *The incidental Verhulst regression model.*
Let K and r be non-zero parameters of the logistic equation. In this model, it is assumed that $\Delta Y(t)|Y(t-1)$ is Poisson distributed with expectation

$$\mathbb{E}\left[\Delta Y(t)|Y(t-1)\right] = rY(t-1)\left(1 - \frac{Y(t-1)}{K}\right)$$

Definition 2 and the theory of nonlinear regression are combined to obtain the cumulative Verhulst regression model.

Definition 6. *The cumulative Verhulst regression model*
Let L and q be the positive parameters of the cumulative Verhulst curve. Further, denote by $\sigma > 0$ the standard deviation of the model.

$$Y(t) = \frac{L}{1 + \left(\dfrac{L - Y(0)}{Y(0)}\right)\exp\left(-qt\right)} + \epsilon_i \qquad\qquad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

Generally, one would expect larger deviations from the expectation as the number of reported infections grow. This phenomenon is somewhat captured in the incidental Verhulst regression model, since the variance of the Poisson distribution equals its mean. However, note that the variance of the cumulative Verhulst regression model is constant. Although this is a problem for the quality of the model, there are some more significant

problems when applying this model to predict infection numbers. We have seen in section 2.1 that these models originate from the SIS model. However, for the analysis of COVID-19, the SIR model is more appropriate since people produce antibodies and, additionally, pass away due to the virus [3]. Therefore, we analyse the performance of these two models, particularly the expectation of the models, when applied to SIR generated data.
To make a consistent comparison, the cumulative number of infections ($Y(t)$) is visualized in the remainder. In the appendix, one can find the $\Delta Y(t)$ variants of these graphs, where the incidental number of infections is shown. Additionally, cumulative analyses are always accompanied in the appendix by figures showing the relative errors of the model fits.

3.3.1 Fitting models to SI generated data

Recall that setting $a = 0$ in the SIR model yields a special case of the SIS model, namely the SI model. In this case, the infected individuals stay in the infected class and do not recover. For completeness, we will show that the model performs well on such a setting. Let us fix, for the remainder of section 3, the parameters S_0, I_0 and b to be: $S_0 = 10000$, $I_0 = 100$, $b = 0.1$. Denote by T the number of analysed data points, where we always consider the **first** T data points. In this setting, this is analogous to T being the number of days after 100 cumulative infections are reported in a country. The number of infections as generated by the SIR model are visualized by crosses, the lines represent the regression models. Further, the dotted lines in the graphs are in correspondence with the chosen values of T.

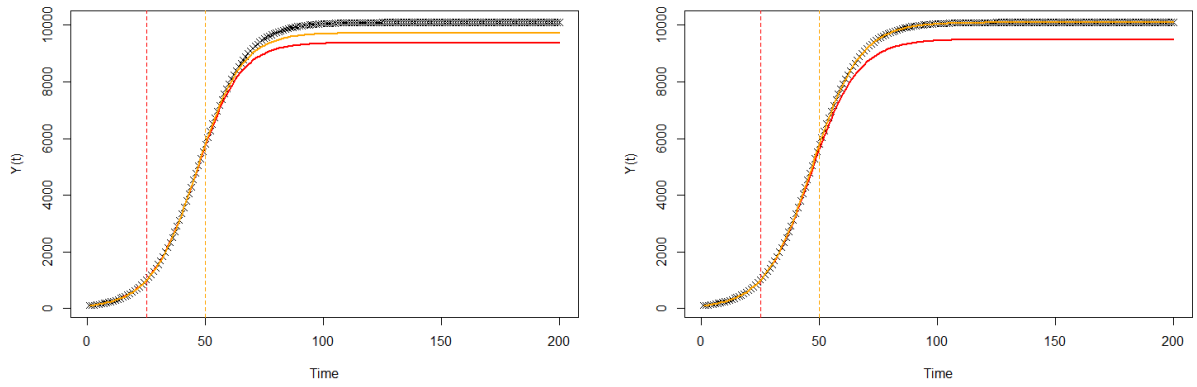

Figure 1: Verhulst regression models fit well on data generated by the SI model
Performances of both incidental (left) and cumulative (right) Verhulst regression models applied to the SI model. Red and yellow lines indicate the performance of the models when fitted until time $T = 25$ and $T = 50$ respectively. Coefficient of determination is high, namely $R^2 > 0.997$ for all scenarios.

3.3.2 Goodness-of-fit of models on SIR generated data

Let us assess the goodness-of-fit of models 5 and 6 on data generated by the SIR model using a general approach. To increase the robustness of the analysis, we use the simulation model that is described in section 2.3, performing 1000 runs. The following two parameter sets are chosen: $\{a = 1/50, T = 200\}$ and $\{a = 1/15, T = 300\}$. Hence, in contrast with the previous section, more complete curves are analysed. Figure 2 shows the variation in the simulation runs and the incidental and cumulative model fits to two extreme epidemics for both parameter sets. Additionally, in table 1, the coefficients of determination of both models applied to the simulations are summarized. We observe high mean coefficients of determination with small standard errors and good fits to extreme SIR curves. So, based on this analysis, there is no reason to reject the goodness-of-fit of the Verhulst regression models when applied to data generated by the SIR model. Generally however, predicting the number of infections is done in an earlier stage of the epidemic. In sections 3.3.3 and 3.4 we will observe that not having access to the full curve leads to problems.

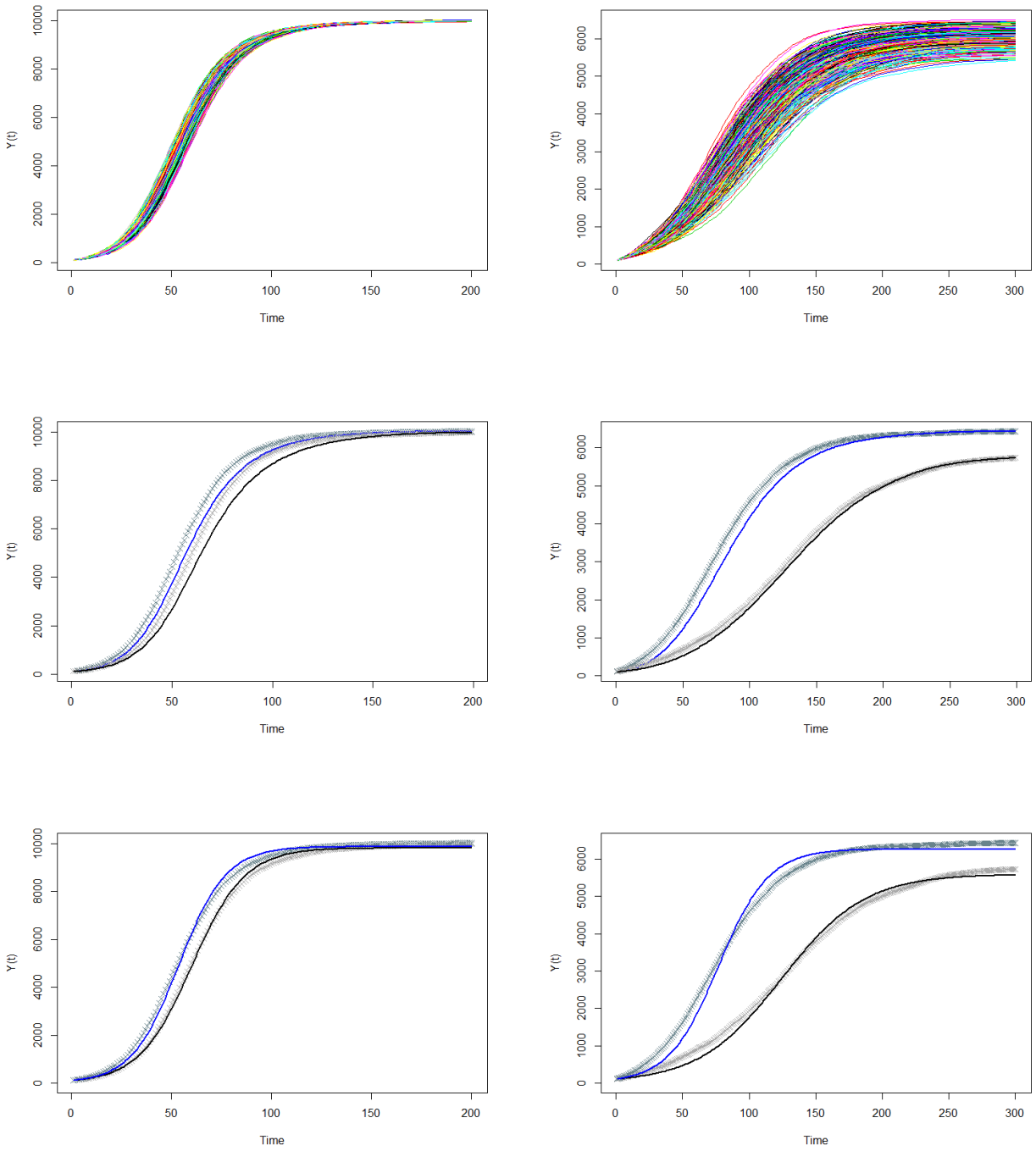

Figure 2: Application of Verhulst regression models to SIR simulations
1000 SIR simulation runs (top row) with $a = \frac{1}{50}$ (left column) and $a = \frac{1}{15}$ (right column).
The middle row and the bottom row show the fit of the incidental and the cumulative Verhulst regression model respectively on two extreme epidemics, a heavy epidemic (light blue crosses) and a mild epidemic (grey crosses). The corresponding fits to these data points are given by blue (heavy epidemic) and black (mild epidemic) lines.

Regression model	a	Mean R^2	SE R^2
Incidental Verhulst	1/50	0.998	$3.85 \cdot 10^{-4}$
Incidental Verhulst	1/15	0.998	$9.26 \cdot 10^{-4}$
Cumulative Verhulst	1/50	0.999	$2.03 \cdot 10^{-4}$
Cumulative Verhulst	1/15	0.997	$1.29 \cdot 10^{-3}$

Table 1: Summary of the coefficients of determination R^2 of model fits to 1000 SIR simulations

3.3.3 Intuitive symmetry argument

As in definition 6, let $r > 0$, $L > 0$, $Y(0) \in \mathbb{N}$ and consider $f(t) = \dfrac{L}{1 + \left(\dfrac{L - Y(0)}{Y(0)}\right) \exp(-qt)}$. Then

$$f'(t) = \frac{LqY(0)\,(L - Y(0)) \exp(qt)}{(L + Y(0)\,(\exp(qt) - 1))^2}$$

Under the mild assumption $Y(0) < L$, which holds for our parameter choices, there is a point of symmetry t_{sym} in this derivative, such that $f'(t - t_{\text{sym}}) = f'(t + t_{\text{sym}})$ for all $t \in \mathbb{R}$. Namely,

$$t_{\text{sym}} = \frac{1}{q} \log\left(\frac{L - Y(0)}{Y(0)}\right)$$

Furthermore, notice that $f(t_{\text{sym}}) = \frac{L}{2}$.
Of course, the number of infections in a SIS setting follows this pattern, e.g. figure 1. However, when we take into account that people become immune to the virus, as in the SIR model, this property is not necessarily preserved for the total number of reported infections $(I + R)$. Generally, flow into the recovery class in a SIR model causes the total number of reported infections to grow faster after the maximal point of the derivative. Therefore, if the full SIR curve is not available, a negative bias of the Verhulst regression models is unavoidable. This observation is visualized by plotting the derivative of the number of reported infections in the SIR model, that is $I' + R' = bSI$, and comparing it with the symmetric continuation of the derivative, i.e., mirroring the data at its maximal value. This (non-)symmetry property shown in figure 3 is a fundamental difference between the SIS and the SIR model. In the next section we will see what this implies for fitting the Verhulst regression model on data generated by the SIR model.

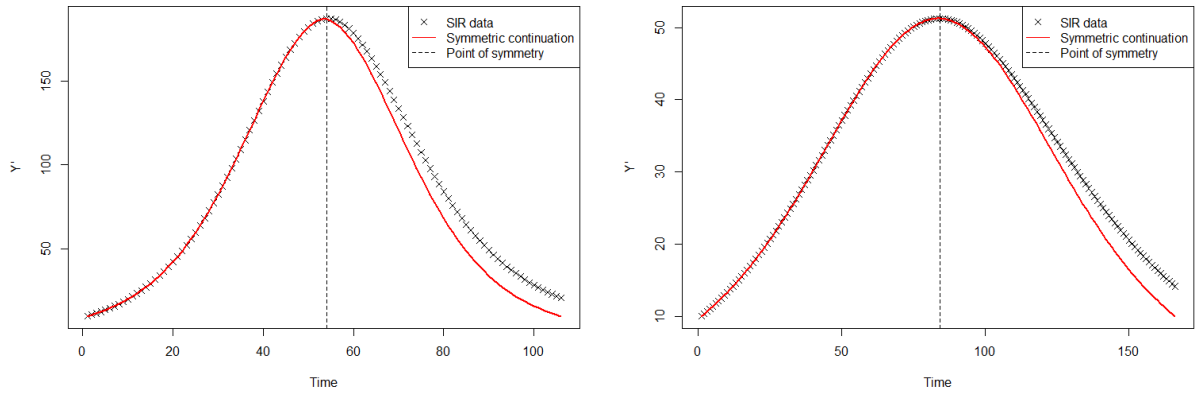

Figure 3: Non-symmetry of the incidental number of infections in the SIR model
Incidental number of infections in a SIR model with $a = \frac{1}{50}$ (left) and $a = \frac{1}{15}$ (right). The red line represents the symmetric continuation of the data, i.e. the red line mirrors the data at its maximal value.

3.4 Time varying performance of models on SIR generated data

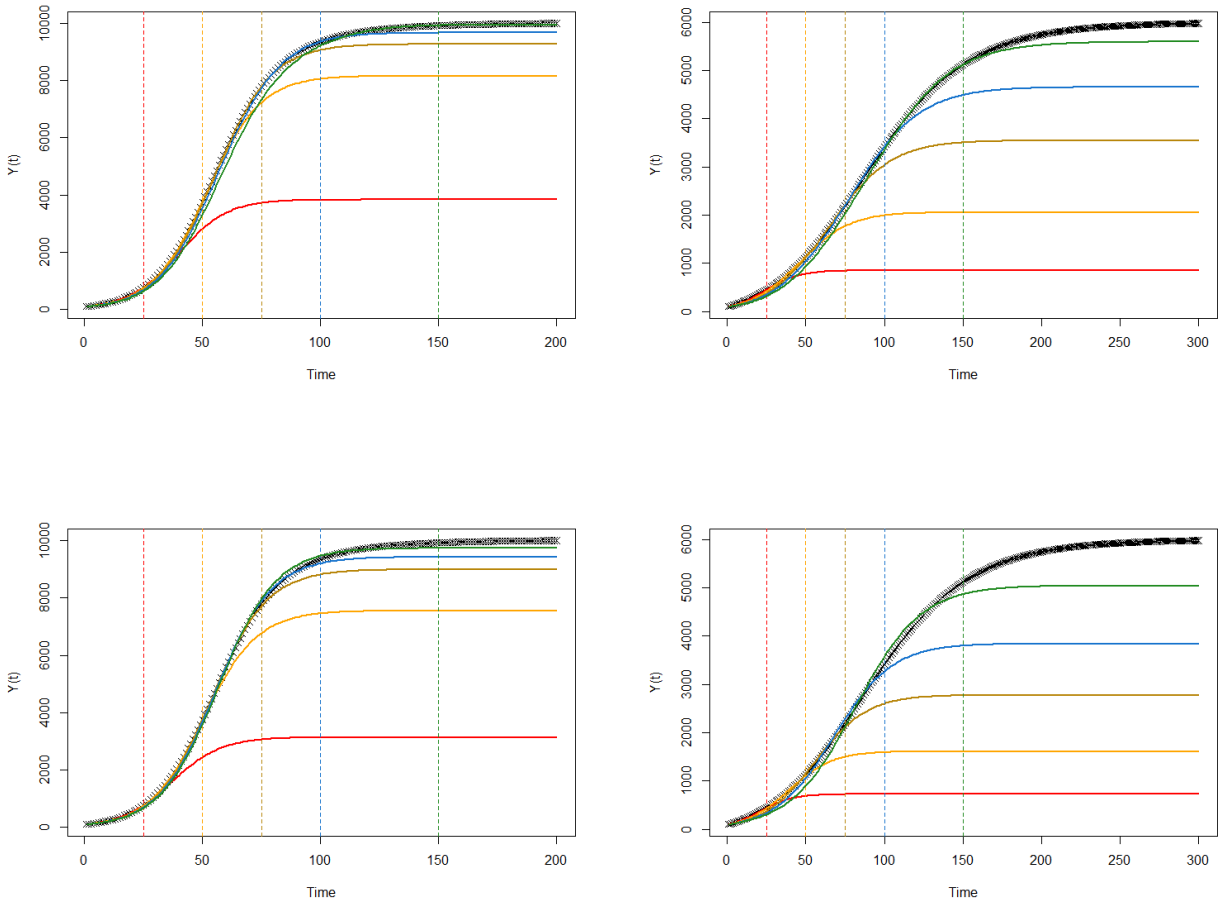

Figure 4: Verhulst regression models are negatively biased when applied to early SIR epidemics Time-varying performance of both incidental (top row) and cumulative (bottom row) Verhulst regression models applied to the SIR model with $a = \frac{1}{50}$ (left column) and $a = \frac{1}{15}$ (right column) when fitted until time T. Values for T are indicated by dotted lines and equal 25 (red), 50 (yellow), 75 (brown), 100 (blue) and 150 (green).

Figure 4 shows the bias that both models have when partial SIR data is available. As expected by the symmetry argument of the previous section, the biases are less significant when more data is available. Further, higher recovery rates seem to produce more significant biases, which is expected since in such cases, the SIR model deviates more from the SIS model. In the technical report of van den Heuvel, Regis and Zhan (see `tue.nl`), a similar result is observed when applying a three parameter Verhulst regression model to the reported number of COVID-19 infections in the Chinese provinces. The definition and the results of this model can be read in the appendix (section 6.1).

3.5 The SIR regression model

Since there is no closed form solution of the SIR model in the literature, we turn to methods of numerical integration. In particular, the Runge-Kutta 4-th order method is applied to the system of equations 6. Raissi et al. [11] suggest an efficient estimation procedure, employing an unconstrained non-linear optimization

algorithm such a the Nelder-Mead algorithm which searches for a local minimum using a regression approach. Let us have a sequence of n reported cumulative infections $Y_1, Y_2, \ldots, Y_n$ where the times between reporting are equal. To estimate these quantities using the SIR model, we set the number of reported cumulative infections to the sum of infections and recoveries in the SIR model. In other words, $Y_k = I_k + R_k$ for $k = 1, \ldots, n$. Furthermore, let $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ be the initial estimates, by numerically integrating the SIR model for some starting values a_{start} and b_{start}. Then, following the method of least squares, the Nelder-Mead algorithm is applied to minimize the sum of squares of residuals, over the parameters a and b.

$$(\hat{a}, \hat{b}) = \underset{(a,b)\geq 0}{\arg\min} \sum_{k=1}^{n} (Y_k - \hat{Y}_k(a, b))^2 \tag{12}$$

This procedure yields estimates $\hat{a}$ and $\hat{b}$, from which we can derive $\hat{S}_\infty$ and $\hat{R}_\infty$ with the help of equation 7. The multivariate delta method can be applied to get confidence on these quantities, see section 6.2.
An alternative model where the sum of squares of the incidental number of infections is minimized can be constructed in a similar manner. However, the differences in results that such a model has with respect to this proposed model are marginal. In the two Verhulst regression models, the total population size of the system, N, is captured in the free parameters. In this model however, the total population size is fixed as $N = S_0 + I_0$. It can be considered as a benefit, since it makes for a richer model. However, the total population size is not always easy to derive, since the assumption that everyone in a country is either susceptible or infected at $t = 0$ can be questioned.
To test the performance of this model, it is applied to the 1000 simulation runs described in section 3.3.2. Crude starting values $a_{\text{start}} = b_{\text{start}} = 1$ are used. Since the error made by the SIR regression model can be either positive or negative, analysing the squared relative error is appropriate. The same procedure is also conducted by analysing the relative error, which can be found in the appendix (figure 17).
Again, higher recovery rates cause worse model fits. Nonetheless, when comparing these results to the results of section 3.4 (also see figure 14), a solid improvement can be observed.

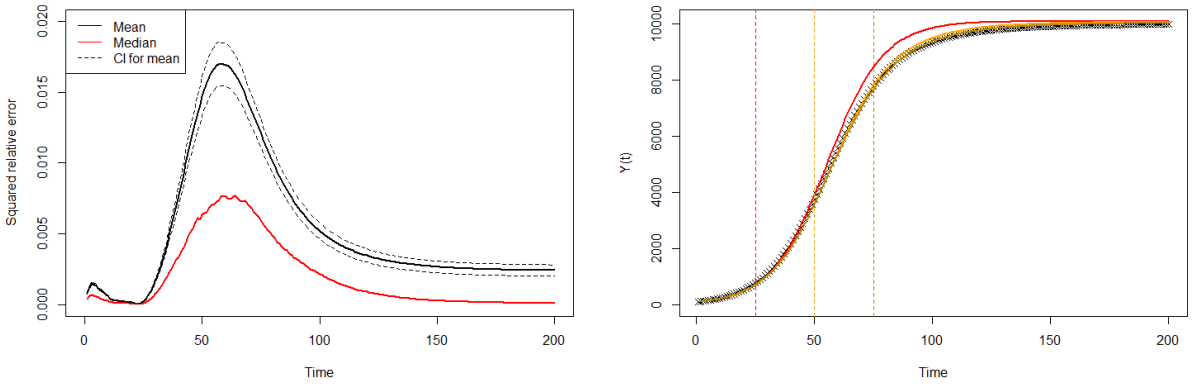

Figure 5: Analysis of SIR regression model applied to SIR simulations with low recovery rate (left): Median and (95% confidence interval of the) mean of the squared relative errors of the SIR regression model applied to the 1000 SIR simulation runs, with $a = \frac{1}{50}$. Model is fitted until time $T = 25$.
(right): Model fits for an average SIR curve with $a = \frac{1}{50}$ when fitted until time T. Values for T are indicated by dotted lines and equal 25 (red), 50 (yellow) and 75 (brown).

Algorithm 1 Estimating SIR parameters

Initialize S_0, I_0 from data
Set a_{start} and b_{start}
function O(a,b) ▷ O: Objective Function
    Integrate(SIR) ▷ Numeric integration using RK4
    return $SS(a, b)$ = Sum of squares of residuals ▷ See equation 11
procedure NELDER-MEAD(O, a_{start}, b_{start})
    Minimize O(a, b)
    return Estimates for a, b and $SS(a, b)$

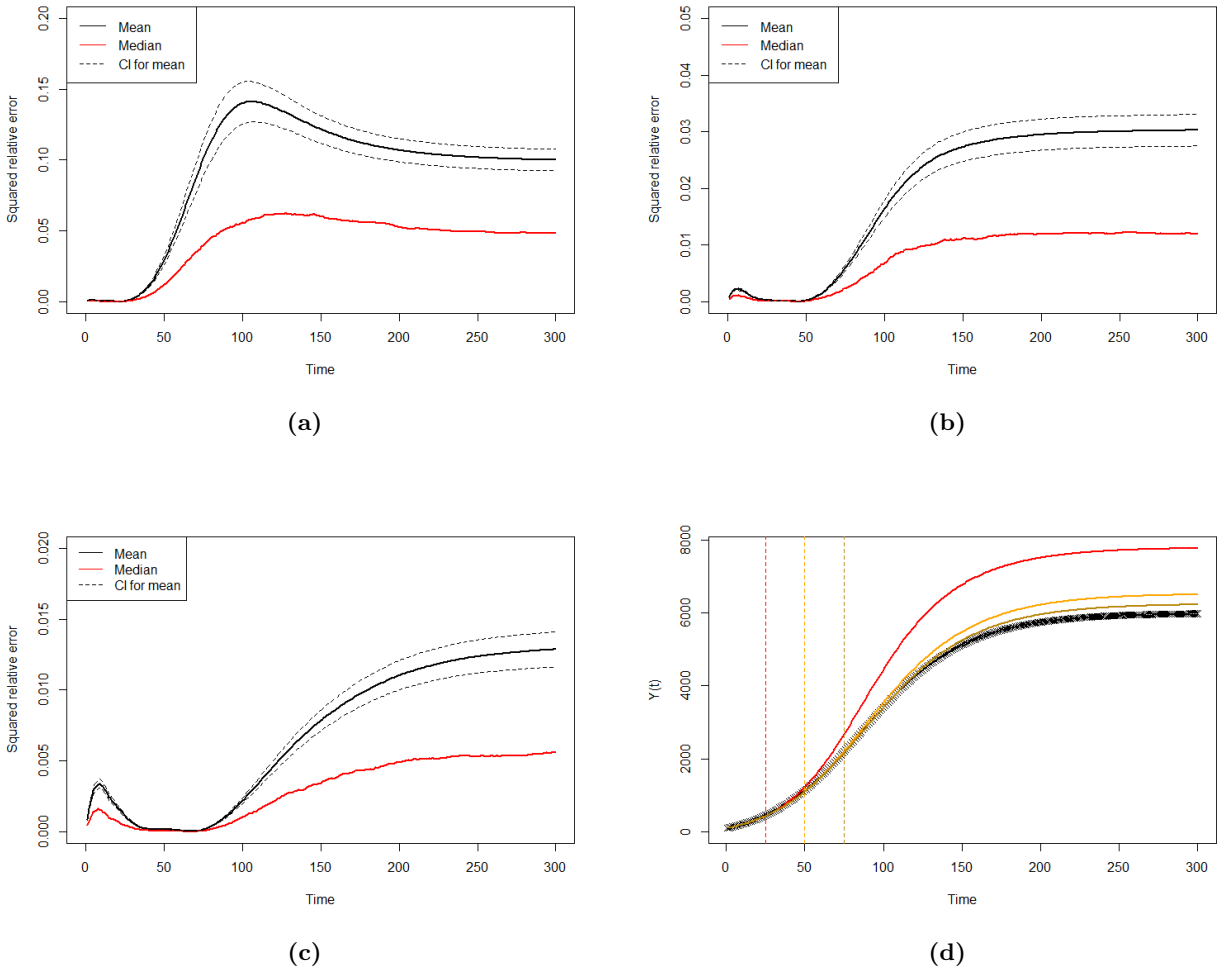

Figure 6: Analysis of SIR regression model applied to SIR simulations with high recovery rate (a, b, c): Median and (95% confidence interval of the) mean of the squared relative errors of the improved model applied to the 1000 SIR simulation runs, with $a = \frac{1}{15}$. Figures (a), (b) and (c) correspond with $T = 25, 50, 75$ respectively.
(d): Model fits for an average SIR curve with $a = \frac{1}{15}$ when fitted until time T. Values for T are indicated by dotted lines and equal 25 (red), 50 (yellow) and 75 (brown).

4 Case study: COVID-19 pandemic

The ultimate goal of this research is to improve predictions regarding the number of reported COVID-19 infections per country. Therefore, we will compare the Verhulst regression models (definitions 5 and 6) to the SIR regression model (section 3.5) when applied to empirical data. The data is obtained by Johns Hopkins University, who have been tracking the reported number of infections of all countries since 22/01/2020 at `https://coronavirus.jhu.edu/`. The data set, containing the reported infections until 14/07/2020, is retrieved on 15/07/2020.
An important remark is that the reported numbers are only fractions of the total infections, which disagrees with our model assumptions. Furthermore, governmental interventions affect the spread of COVID-19, resulting in time-varying parameters. In the models proposed in this paper, all parameters are assumed to be independent of time. Post et al. [12] argue that, among the considered European countries, Sweden is the country where the interventions had the least effect. Therefore, besides the Netherlands, we analyse the data of Sweden. The analysis is started whenever there are at least 100 reported infections, hence $t_0 = \arg\min_t\{Y(t) \geq 100\}$, and set $Y(0) = I_0 = I_{t_0}$ (recall $R_0 = 0$). Further, it is assumed that N equals the total population size of the country and $S_0 = N - I_0$.
Note that each of these assumptions can be questioned. However, this section is still valuable for an empirical comparison of the three regression models. Based on the retrieved data set, the Netherlands and Sweden both reported more than 100 cumulative infections for the first time on 06/03/2020.

4.1 The Netherlands

According to CBS (`cbs.nl`), the Netherlands had a population size of 17.422.992 inhabitants at the start of March 2020. Figure 7 shows the (time-varying) performances of the three regression models introduced in this thesis when applied to the reported number of COVID-19 infections in the Netherlands. When considering partial data, both Verhulst regression models underestimate the COVID-19, which is in line with section 3.4. However, the data is also poorly modeled by the SIR regression model.

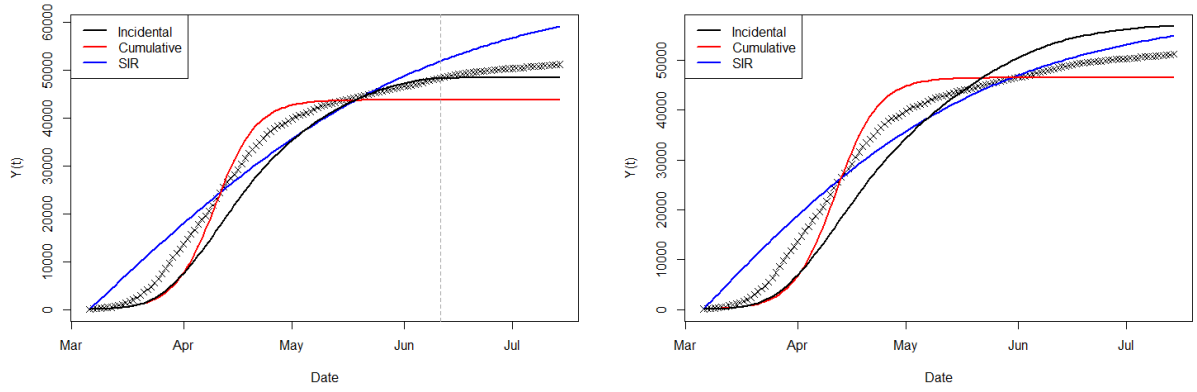

Figure 7: Comparison of three models applied to reported Dutch COVID-19 infections
Model fits of the incidental Verhulst regression model (black), cumulative Verhulst regression model (red) and the SIR curve fitting model (blue) to the cumulative number of reported COVID-19 infections in the Netherlands. Models are fitted until 11/06/2020 (left, 11/06/2020 indicated by dotted grey line) and until the last date in the data set, 14/07/2020 (right).

4.2 Sweden

Statistics Sweden (`scb.se`) states that Sweden houses 10.341.503 inhabitants at the start of March 2020. Similarly to the previous section, figure 8 shows the (time-varying) performances of the three regression models when applied to the reported number of COVID-19 infections in Sweden. Again, we observe an

underestimation of both Verhulst regression models when fitted to partial data. Since the model fit of the SIR regression model is good (also see table 2 and figure 21), theorem 2 will be applied. This yields estimates $\widehat{S_\infty} = 10147446$ and $\widehat{R_\infty} = 194057$. Approximate 95% confidence intervals are obtained using the multivariate delta method, described in detail in section 6.2 and are given by:

$$S_\infty \in [10140373, 10154518] \qquad \text{and} \qquad R_\infty \in [186985, 201130]$$

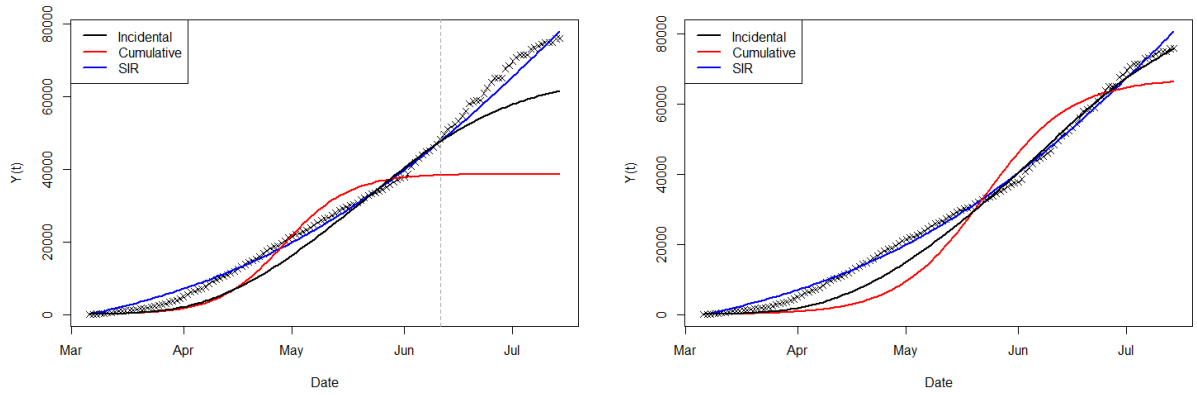

Figure 8: Comparison of three models applied to reported Swedish COVID-19 infections
Model fits of the incidental Verhulst regression model (black), cumulative Verhulst regression model (red) and the SIR curve fitting model (blue) to the cumulative number of reported COVID-19 infections in Sweden. Models are fitted until 11/06/2020 (left, 11/06/2020 indicated by dotted grey line) and until the last date in the data set, 14/07/2020. (right)

5 Discussion

This thesis provides an introduction to two simple compartmental models for modelling infectious disease, the SIS model and the SIR model. Particularly, it is shown that Verhulst models, originally used for modelling population growth, can be obtained from the SIS model. Statistical frameworks, such as the theory of generalized linear models and of nonlinear regression, are introduced to properly define these Verhulst models. When applying such models on simulations of the SIR model, good model fits are observed. However, these positive results do not generalize when applying the Verhulst models to SIR epidemics which are not in their final stages. The most significant problem is that these Verhulst models are negatively biased towards the end of the SIR epidemic. An important remark is that there is no way to obtain an exact closed form expression of this bias, since the SIR model can not be solved exactly using analytical methods. Intuitively, a fundamental reason for the occurrence of this bias is the symmetry property of the total number of infections, which is preserved in the SIS model but not in the SIR model. To tackle this problem, the SIR regression model is proposed, which has an improved performance when applied to the SIR simulations that are in their early stages. The goal of this research is to improve predictions regarding the cumulative number of reported COVID-19 infections per country. Therefore, the three regression models are applied to empirical data, the number of reported COVID-19 infections of the Netherlands and Sweden. Again, the negative bias of the Verhulst models when applied to partial data is present. However, especially the results of the Netherlands suggest that the assumptions of the SIR model are too strong to realistically model the COVID-19 epidemics per country. An important limitation of this research when considering the COVID-19 pandemic is that in all three models, the parameters are assumed to be constant over time. However, partly due to the profound governmental measures [12], this assumption is very questionable. Whether the time-varying nature of the parameters is a primary cause of the problems observed in the case study could be further investigated, for example by implementing time-varying parameters to the simulation program and conducting a similar analysis of the regression models.

Many choices such as the type of compartmental models and of the statistical frameworks have been made. It is important to note that although these choices are not arbitrary, many different options exist, some of which are proposed throughout the thesis, where similar analysis methods are also adequate.
This thesis focused mainly on the mathematical modeling of infectious disease, without taking too many specifics of COVID-19 into account. This makes it such that the models are generalisable to other infectious diseases. Another advantage is that whenever there is little knowledge about a certain disease, as was the case for COVID-19 in the early months of 2020, the models can still be applied. However, realistically implementing disease specific information, such as a value for the recovery rate of patients a, can improve the model.

References

[1] Verhulst PF. Notice sur la loi que la population poursuit dans son accroissement. Correspondance Mathématique et Physique. 1838;10:113–121.

[2] Kermack W, McKendrick A. Contributions to the mathematical theory of epidemics-I. 1991;53(1):700–721.

[3] Long QX, Liu BZ, Deng HJ, Wu GC, Deng K, Chen YK, et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. Nature Medicine. 2020;.

[4] Martcheva M. An Introduction to Mathematical Epidemiology; 2013.

[5] Khan H, Mohapatra RN, Vajravelu K, Liao SJ. The explicit series solution of SIR and SIS epidemic models. Applied Mathematics and Computation. 2009;215(2):653–669. Available from: http://dx.doi.org/10.1016/j.amc.2009.05.051.

[6] Harko T, Lobo FSN, Mak MK. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. Applied Mathematics and Computation. 2014;236(0):184–194.

[7] Allen LJS. A primer on stochastic epidemic models : Formulation , numerical simulation , and analysis. Infectious Disease Modelling. 2017;2(2):128–142. Available from: http://dx.doi.org/10.1016/j.idm.2017.03.001.

[8] Dobson AJ, Barnett AG. An introduction to generalized linear models, third edition; 2008.

[9] Winkelmann R. Econometric analysis of count data. Springer Berlin Heidelberg; 2008.

[10] Ruckstuhl A. Introduction to Nonlinear Regression. ZHAW Zürcher Hochschule für Angewandte Wissenschaften. 2010;(October). Available from: https://www.ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/AdvancedStudiesinAppliedStatistics/course-material/robust-nonlinear/nlreg16E.pdf.

[11] Raissi M, Ramezani N, Seshaiyer P. On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. Letters in Biomathematics. 2019;0(0):1–26. Available from: https://doi.org/23737867.2019.1676172.

[12] Post RAJ, Regis M, Zhan Z, van den Heuvel ER. How did governmental interventions affect the spread of COVID-19 in European countries? medRxiv. 2020;p. 2020.05.27.20114272. Available from: http://medrxiv.org/content/early/2020/05/29/2020.05.27.20114272.abstract.

6 Appendix

6.1 Three parameter Verhulst model

Making sensible assumptions on $Y(0)$ when predicting the number of reported COVID-19 infections is important when applying this model. To avoid this problem of the original Verhulst model (definition 2, an alternative model is also used in practice.
Since generally $0 < I_0 \ll N$, the assumption $I_0 < K = N\left(1 - \frac{a}{b}\right)$ often holds. Notice that in that case, we can reparametrize equation 3 with $\beta = r$, $M = K$, and $\alpha = \frac{1}{r} \log \frac{M - I_0}{I_0}$. This leads to the following expression, which is more flexible to data, referred to as the three parameter Verhulst model.

Definition 7. *The three parameter Verhulst model*
The expected cumulative number of reported infections at time $t \geq 0$ can be assumed to follow the three parameter Verhulst model, that is

$$\mathbb{E}\left[Y(t)\right] = \frac{M}{1 + \exp\left(-\beta\left(t - \alpha\right)\right)} \tag{13}$$

where $\alpha \in \mathbb{R}$ and β and M are positive.

Using the theory of nonlinear regression, we introduce the three parameter cumulative Verhulst regression model. This model was the fundamental model that van den Heuvel, Zhan and Regis used in their research for predicting the number of COVID-19 infections (see `tue.nl`). The time-varying performance on data generated by the SIR model (figure 9) shows similar results as we have seen in section 3.4.

Definition 8. *The three parameter cumulative Verhulst regression model*
Let $M > 0$, $\beta > 0$ and $\alpha \in \mathbb{R}$ be the parameters of the three parameter Verhulst curve. Further, denote by $\sigma > 0$ the standard deviation of the model.

$$Y(t) = \frac{M}{1 + \exp\left(-\beta\left(t - \alpha\right)\right)} + \epsilon_i \qquad\qquad \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

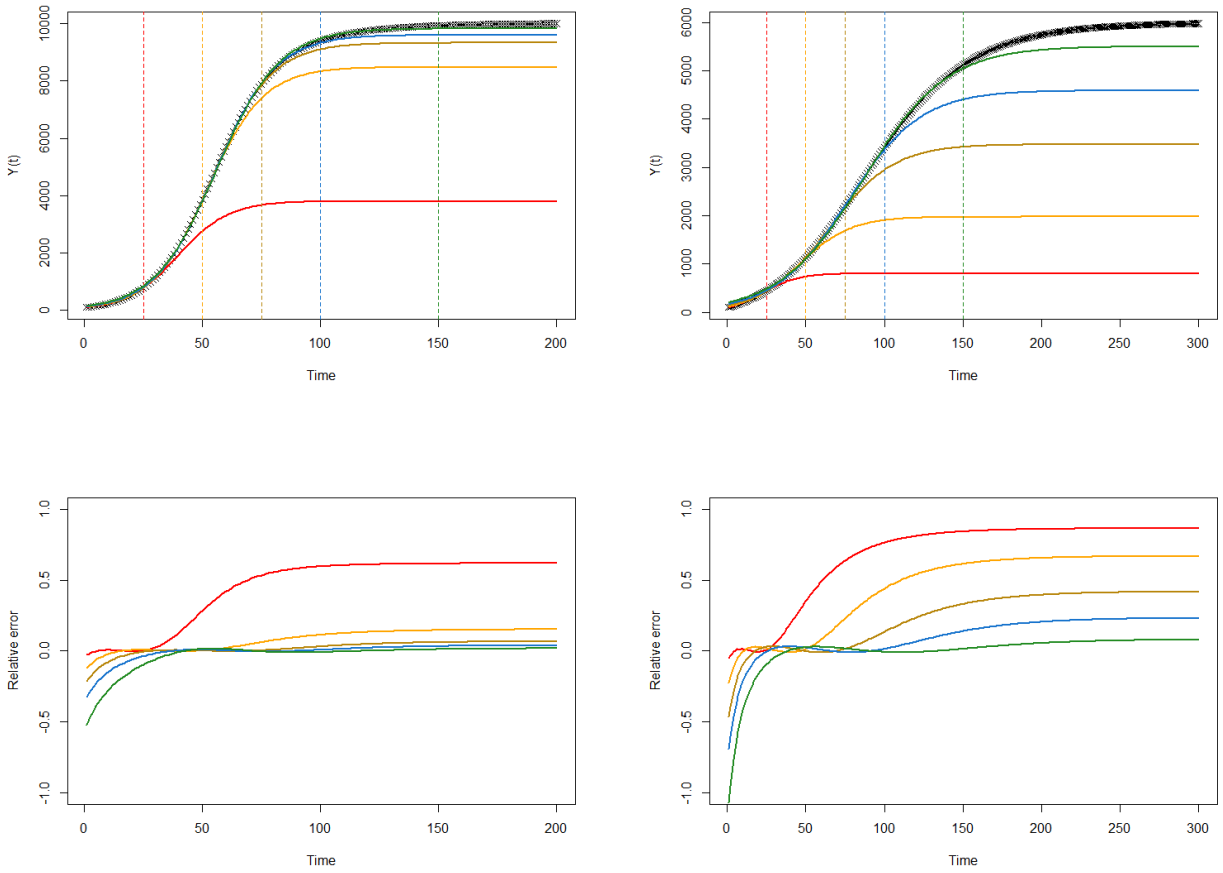

Figure 9: The three parameter Verhulst regression model is negatively biased when applied to SIR epidemics in early stages
Time varying performance of the three parameter cumulative Verhulst regression model (top row) for $a = \frac{1}{50}$ (left column) and $a = \frac{1}{15}$ (right column) when fitted until time T. Values for T are indicated by dotted lines and equal 25 (red), 50 (yellow), 75 (brown), 100 (blue) and 150 (green). The bottom row shows the relative error of the model.

6.2 Multivariate delta method for the SIR regression model

Using section 3.2.1, we estimate the $n \times 2$ matrix of partial derivatives A by

$$\widehat{A(a,b)} = \lim_{\delta \to 0} \begin{bmatrix} \dfrac{\hat{Y}_1(\hat{a}+\delta,\hat{b}) - \hat{Y}_1(\hat{a}-\delta,\hat{b})}{2\delta} & \dfrac{\hat{Y}_1(\hat{a},\hat{b}+\delta) - \hat{Y}_1(\hat{a},\hat{b}-\delta)}{2\delta} \\ \vdots & \vdots \\ \dfrac{\hat{Y}_n(\hat{a}+\delta,\hat{b}) - \hat{Y}_n(\hat{a}-\delta,\hat{b})}{2\delta} & \dfrac{\hat{Y}_n(\hat{a},\hat{b}+\delta) - \hat{Y}_n(\hat{a},\hat{b}-\delta)}{2\delta} \end{bmatrix}$$

This yields an estimate of the covariance matrix $V(a,b)$.

$$\widehat{V(a,b)} = \hat{\sigma}^2 \left(\widehat{A(a,b)}^T \widehat{A(a,b)} \right)^{-1} \qquad \text{with} \qquad \hat{\sigma}^2 = \frac{SS(\hat{a},\hat{b})}{n-2}$$

Assuming that the joint distribution of $\hat{a}$ and $\hat{b}$ is well approximated by a normal distribution, i.e.

$$\begin{bmatrix} \hat{a} - a \\ \hat{b} - b \end{bmatrix} \overset{d}{\approx} \mathcal{N}(0, \widehat{V(a,b)})$$

Recall from theorem 2 that

$$S_\infty = -\frac{aN}{b} W\left(-\frac{bS_0}{aN} \exp\left(-\frac{b}{a}\right)\right)$$

By taking partial derivatives with respect to a and b (recall that S_0 and I_0 are considered fixed), one obtains the gradient of S_∞; ∇S_∞.

$$\nabla S_\infty = \begin{bmatrix} \frac{\partial S_\infty}{\partial a} \\ \frac{\partial S_\infty}{\partial b} \end{bmatrix} = \begin{bmatrix} -\frac{W\left(\phi\left(a,b\right)\right)\left(aW\left(\phi\left(a,b\right)\right)+b\right)N}{ab\left(W\left(\phi\left(a,b\right)\right)+1\right)} \\ \frac{W\left(\phi\left(a,b\right)\right)\left(aW\left(\phi\left(a,b\right)\right)+b\right)N}{b^2\left(W\left(\phi\left(a,b\right)\right)+1\right)} \end{bmatrix} \quad \text{where} \quad \phi\left(a,b\right) = -\frac{bS_0}{aN}\exp\left(\frac{-b}{a}\right)$$

Further, define the estimate of ∇S_∞, $\widehat{\nabla S_\infty}$ as

$$\widehat{\nabla S_\infty} = \begin{bmatrix} -\frac{W\left(\phi\left(\hat{a},\hat{b}\right)\right)\left(\hat{a}W\left(\phi\left(\hat{a},\hat{b}\right)\right)+\hat{b}\right)N}{\hat{a}\hat{b}\left(W\left(\phi\left(\hat{a},\hat{b}\right)\right)+1\right)} \\ \frac{W\left(\phi\left(\hat{a},\hat{b}\right)\right)\left(\hat{a}W\left(\phi\left(\hat{a},\hat{b}\right)\right)+\hat{b}\right)N}{\hat{b}^2\left(W\left(\phi\left(\hat{a},\hat{b}\right)\right)+1\right)} \end{bmatrix}$$

According to the multivariate delta method,

$$(\widehat{S_\infty} - S_\infty) \overset{d}{\approx} \mathcal{N}\left(0, (\widehat{\nabla S_\infty})^T \widehat{V(a,b)} \widehat{\nabla S_\infty}\right)$$

From which we can derive approximate confidence intervals for the quantities S_∞ and R_∞.

6.3 Section 3.3.1 supplementary figures

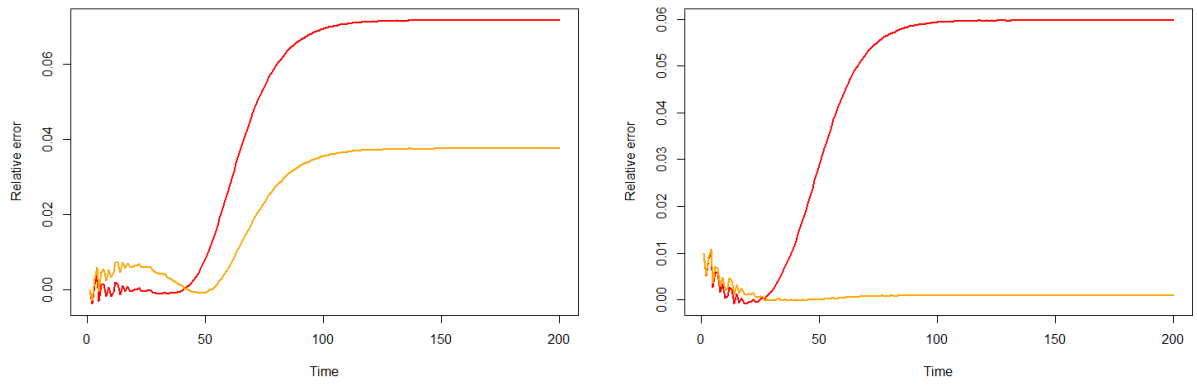

Figure 10: Relative errors of the model fits in figure 1

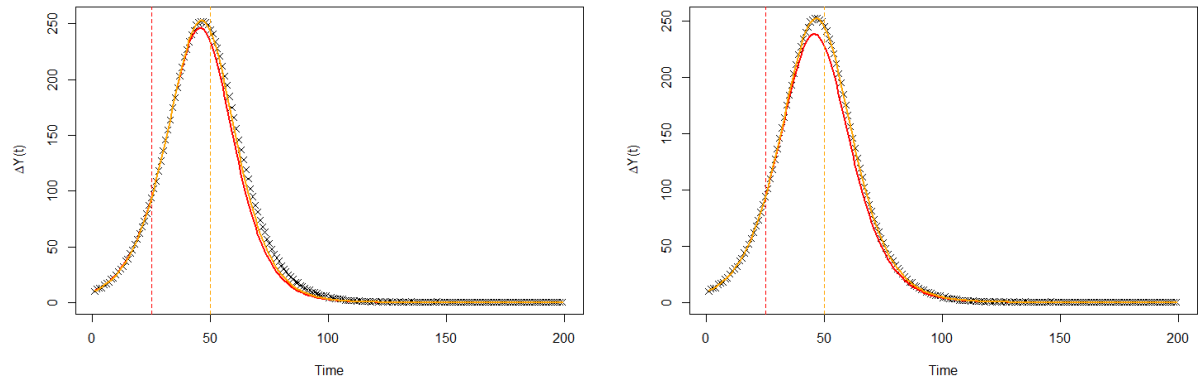

Figure 11: $\Delta Y(t)$ variant of the model fits in figure 1

6.4 Section 3.3.2 supplementary figures

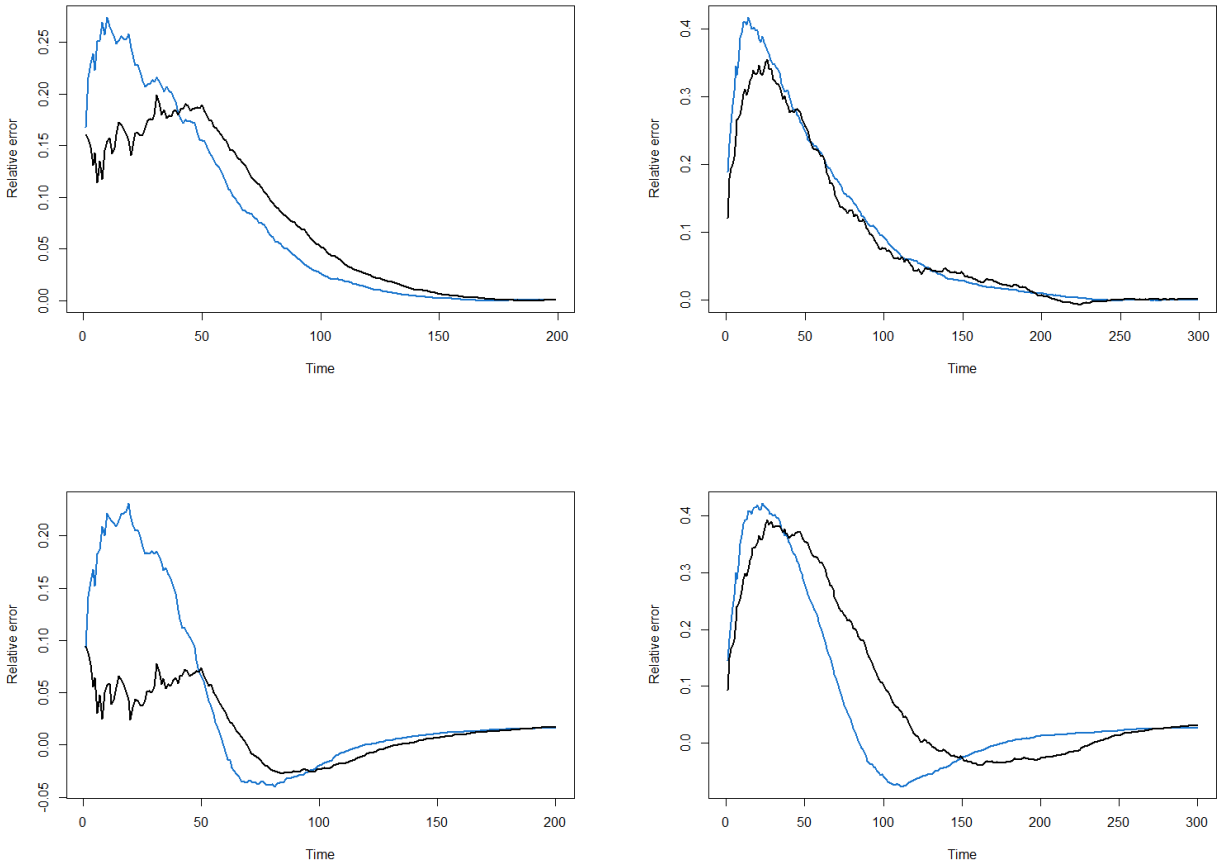

Figure 12: Relative errors of model fits in figure 2

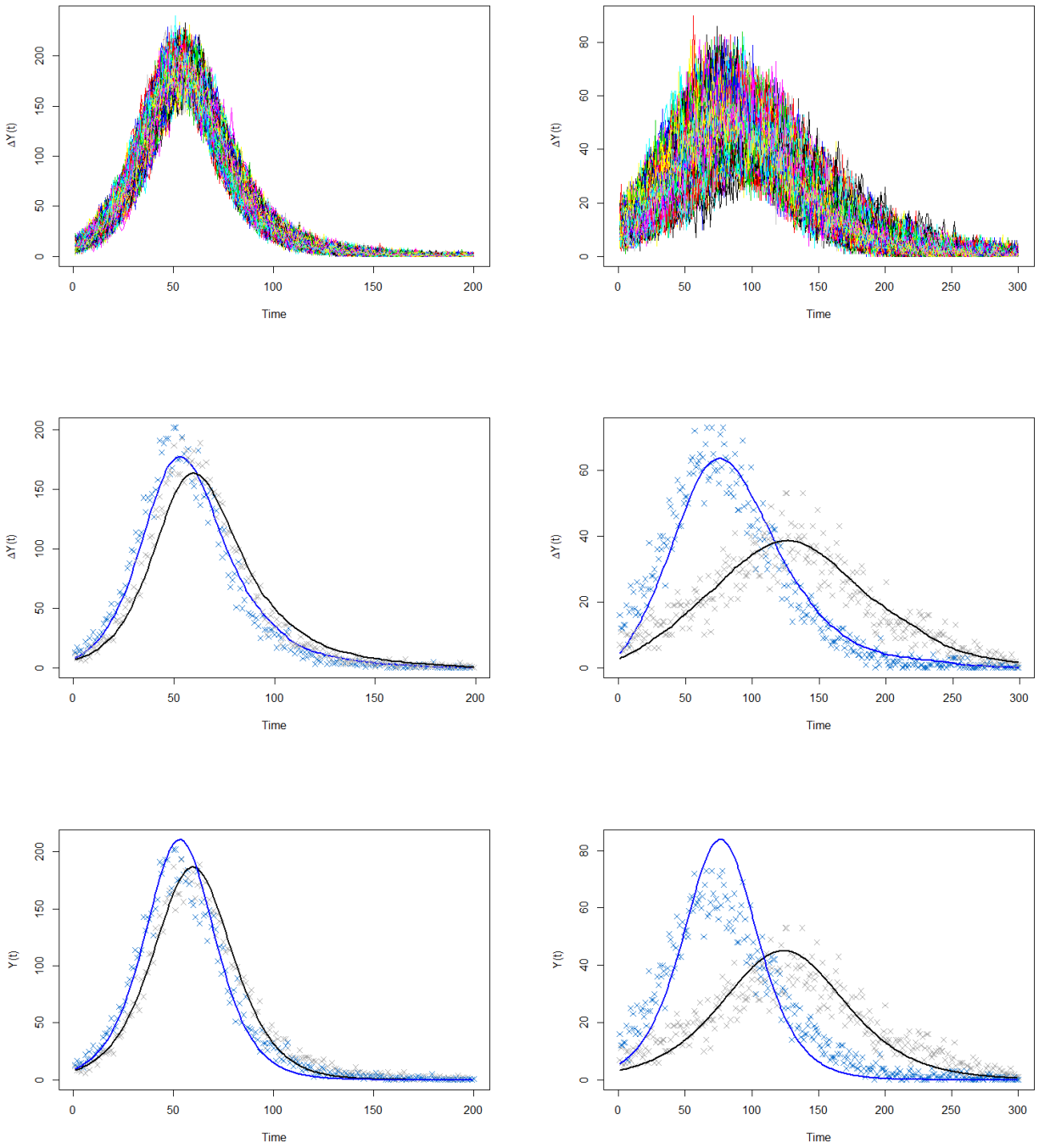

Figure 13: $\Delta Y(t)$ variant of model fits in figure 2

6.5 Section 3.4 supplementary figures

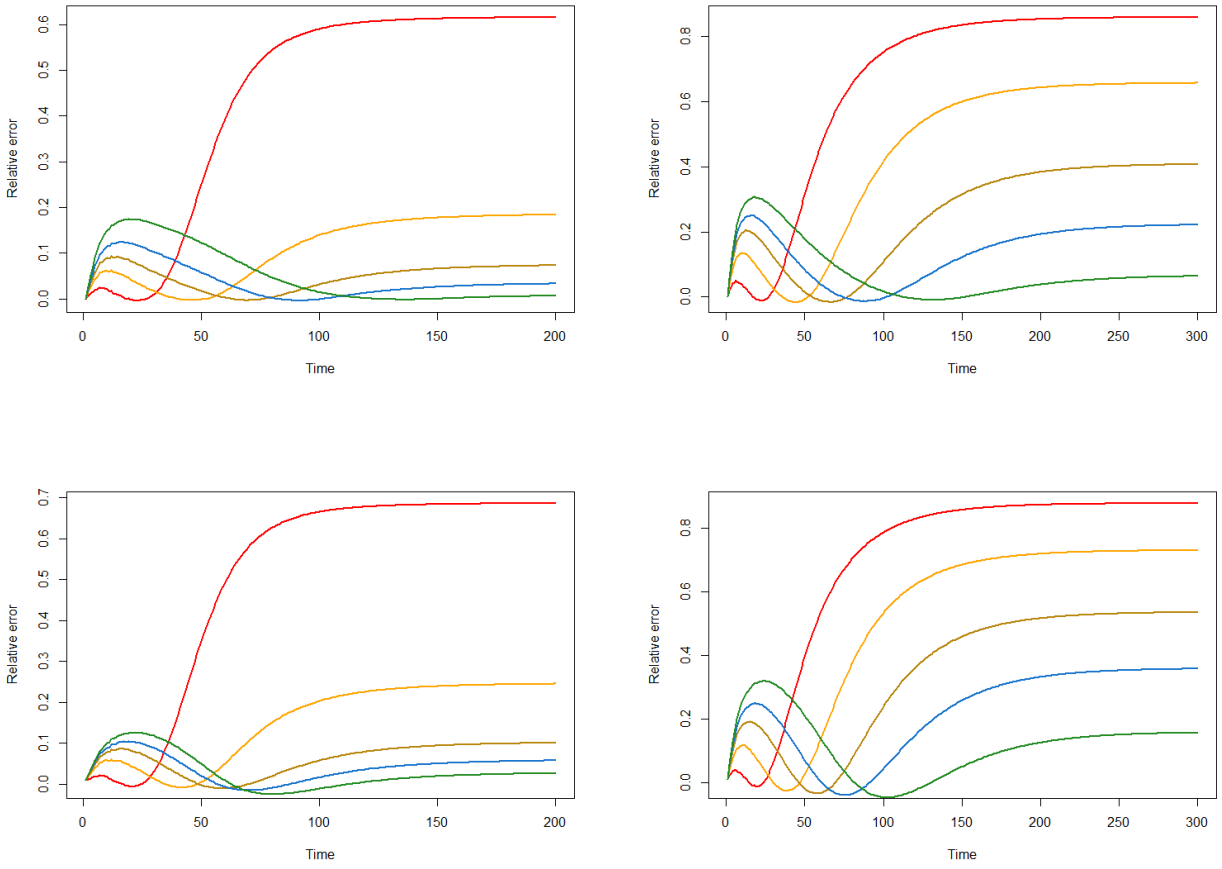

Figure 14: Relative errors of model fits in figure 4

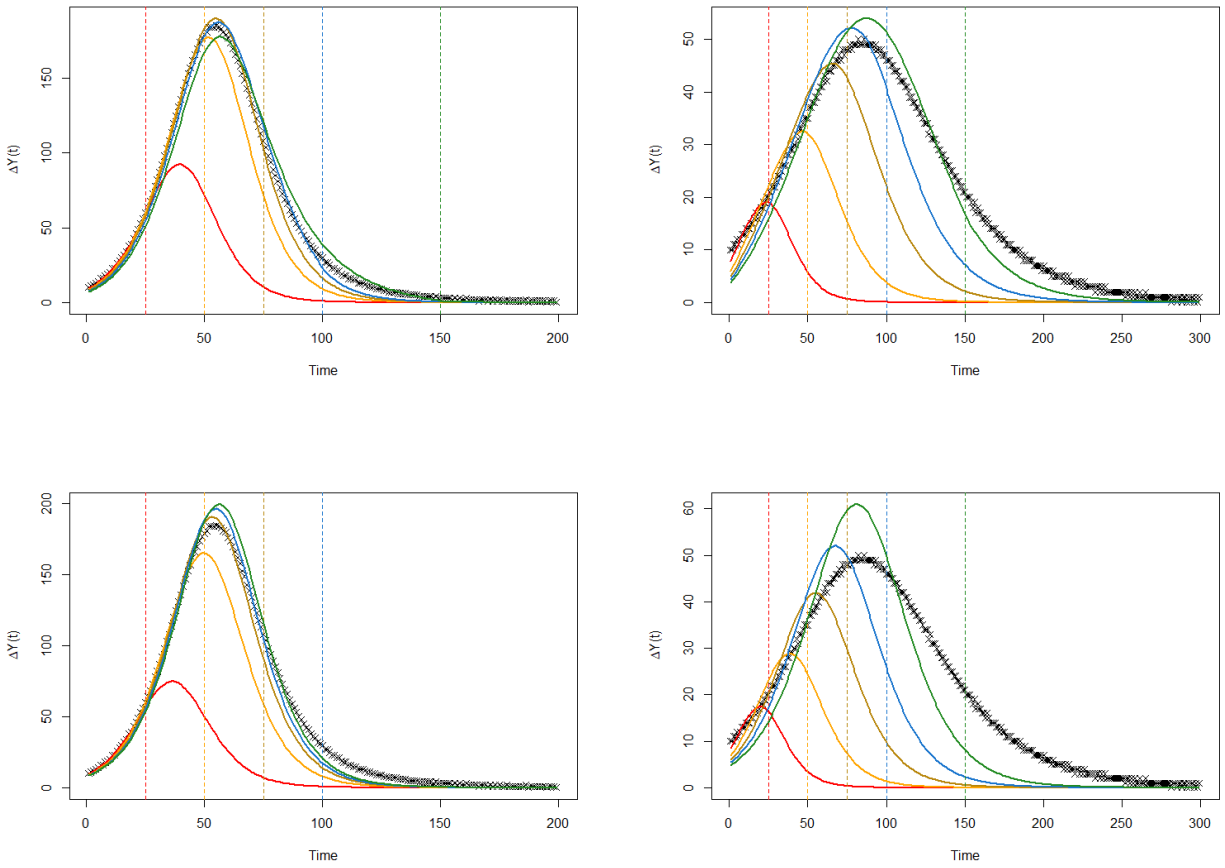

Figure 15: $\Delta Y(t)$ variant of model fits in figure 4

6.6 Section 3.5 supplementary figures

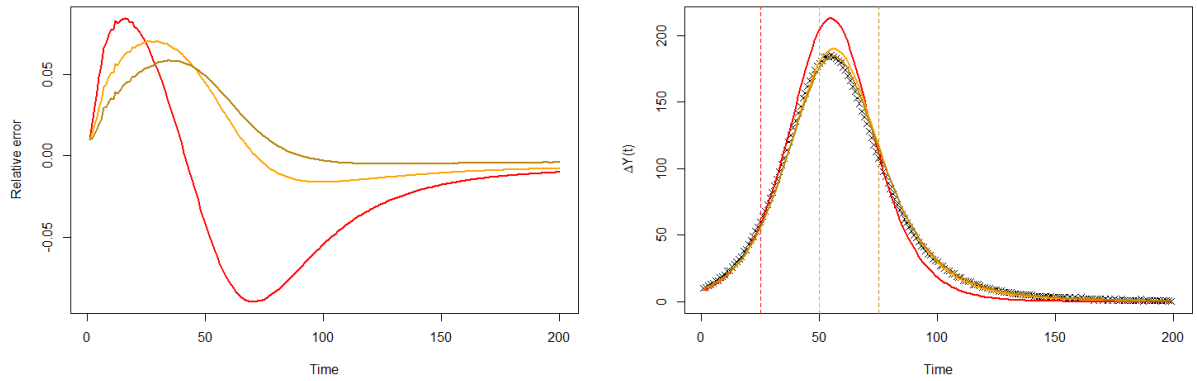

Figure 16: Relative error (left) and $\Delta Y(t)$ variant (right) of model fits in figure 5b

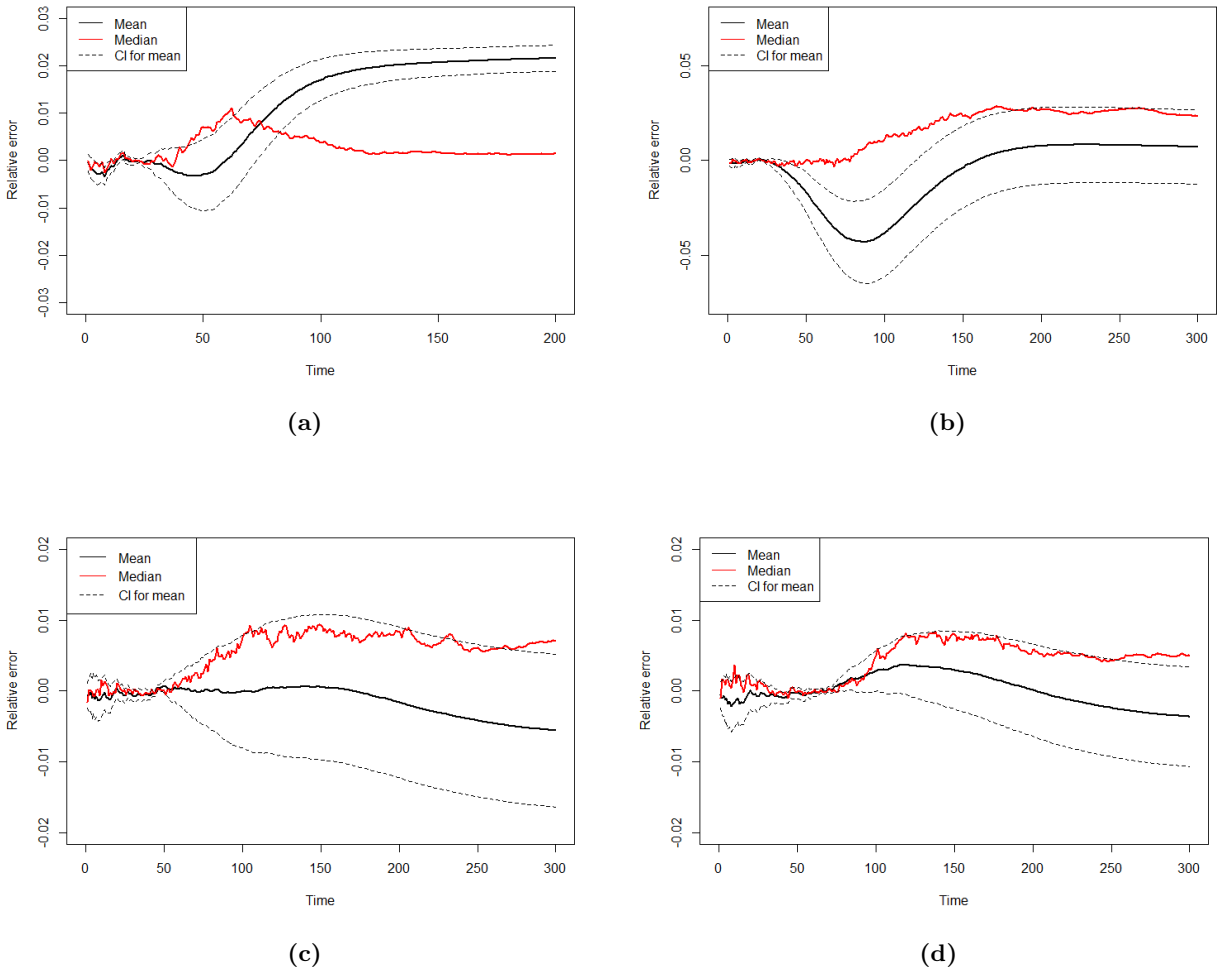

Figure 17: Alternative analysis of SIR regression model applied to SIR simulations
Median and (95% confidence interval of the) mean of the relative errors of the SIR regression model applied to the 1000 SIR simulation runs. Corresponding parameters are given by

(a): $a = \frac{1}{50}, T = 25$ **(b):** $a = \frac{1}{15}, T = 25$ **(c):** $a = \frac{1}{15}, T = 50$ **(d):** $a = \frac{1}{15}, T = 75$

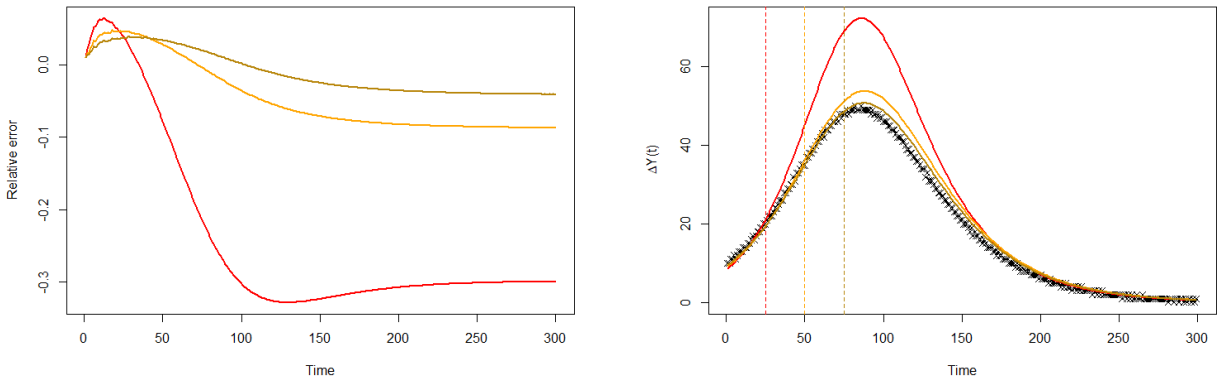

Figure 18: Relative errors (left) and $\Delta Y(t)$ variant (right) of model fits in figure 6d

6.7 Section 4 supplementary figures and table

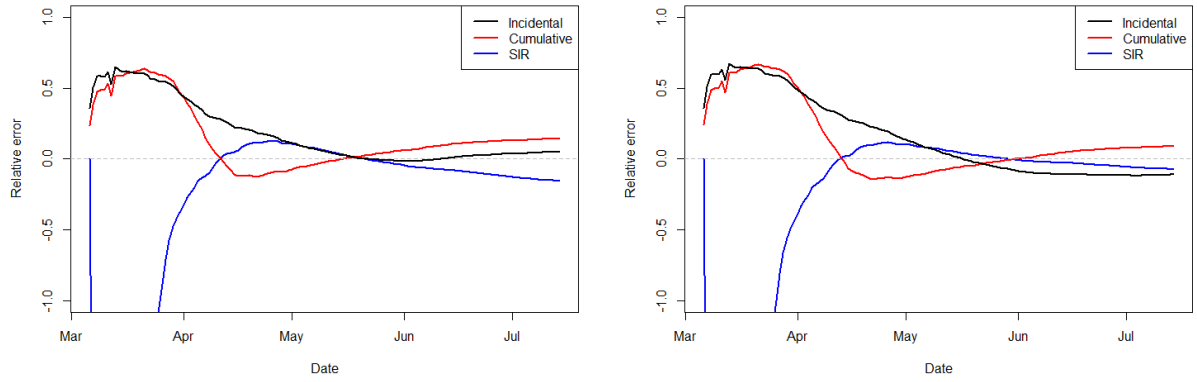

Figure 19: Relative errors of model fits in figure 7

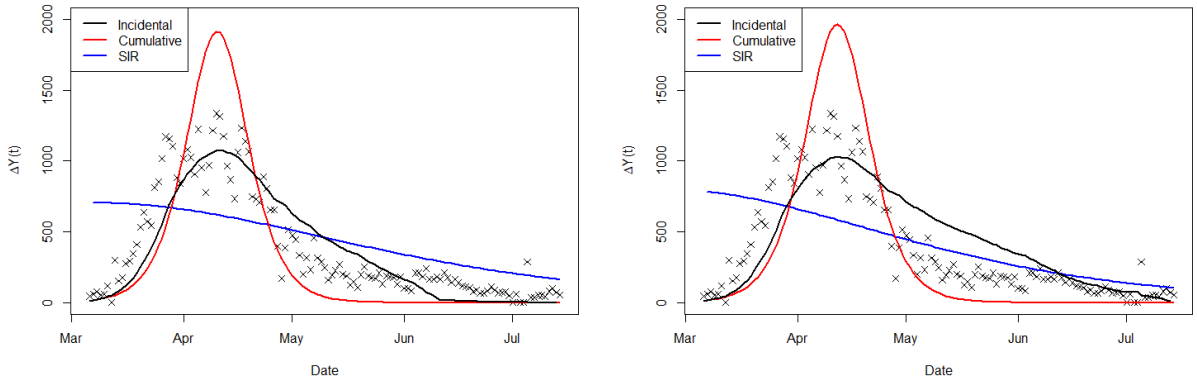

Figure 20: $\Delta Y(t)$ variant of model fits in figure 7

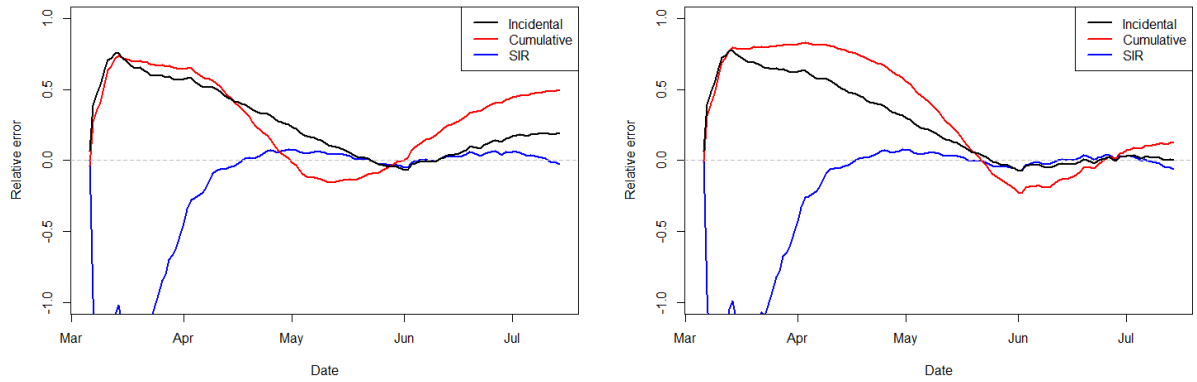

Figure 21: Relative errors of model fits in figure 8

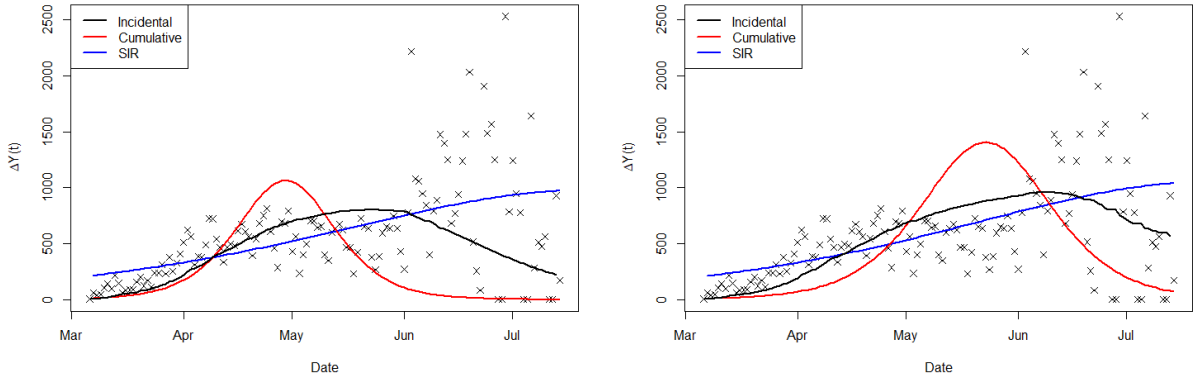

Figure 22: $\Delta Y(t)$ variant of model fits in figure 8

Regression model	Country	Fitted until	R^2
Incidental Verhulst	Netherlands	11/06/2020	0.992
Cumulative Verhulst	Netherlands	11/06/2020	0.980
SIR curve fitting	Netherlands	11/06/2020	0.976
Incidental Verhulst	Netherlands	14/07/2020	0.979
Cumulative Verhulst	Netherlands	14/07/2020	0.983
SIR curve fitting	Netherlands	14/07/2020	0.985
Incidental Verhulst	Sweden	11/06/2020	0.988
Cumulative Verhulst	Sweden	11/06/2020	0.895
SIR curve fitting	Sweden	11/06/2020	0.998
Incidental Verhulst	Sweden	14/07/2020	0.995
Cumulative Verhulst	Sweden	14/07/2020	0.972
SIR curve fitting	Sweden	14/07/2020	0.998

Table 2: Coefficients of determination (R^2) of the model fits displayed in figures 7 and 8