A NEW METHOD FOR CALCULATING THE EFFECTIVE REPRODUCTION NUMBER FOR COVID-19

A TECHNICAL REPORT

Patrick Grice
CloseAssociate (plc)
Wellington, New Zealand
Patrick.Grice@closeassociate.com

Stephen Grice, Ph.D.
CloseAssociate (plc)
Wellington, New Zealand
Stephen.Grice@closeassociate.com

Prof. Richard S. Laugesen
Department of Mathematics
University of Illinois at Urbana-Champaign, USA
Laugesen@illinois.edu

June 1, 2020

ABSTRACT

A new process for estimating $R_{eff}$ for COVID-19 is developed. It combines a deterministic SIR formula for calculating $R_{eff}$ from positive test data and a statistical bootstrapping method for generating confidence intervals.

1 Introduction

SIR modelling is effective at predicting disease spreading through populations [7]. The dominant parameter in all SIR models is the effective reproduction number $R_{eff}$. This effective reproduction number is a function of time $t$, since it changes as the number of Susceptible individuals reduces during an epidemic, and as communities implement measures such as social distancing. Estimating $R_{eff}$ is critical to understanding disease progression in a population, and hence is of intense interest to public health experts and policy officials.

The purpose of this technical report is to present in detail a new “SIR plus Bootstrapping” (SIR+B) process for estimating $R_{eff}(t)$. This new process has been applied to the daily case data for several countries in a separate technical report [2], and results there are compared with statistical methods such as EpiEstim.

The new SIR+B process provides public officials with a rapid and effective tool for knowing whether the reproduction number is increasing or decreasing over time, and whether it exceeds the critical value of 1 for epidemic control.

2 Overview of the method

We proceed backward through the SIR model equations to obtain a formula for the transmission coefficient multiplied by the proportion of Susceptibles in the population. From this one obtains the effective reproduction number $R_{eff}(t)$.

The formula for $R_{eff}(t)$ is then evaluated in terms of the number of Infectious-Tested cases, which we obtain from reported data on the number of confirmed, tested cases per day. Details are presented in Sections [3][4][5]. Sources of error and uncertainty in this SIR stage of the process include a certain amount of educated guesswork in choosing the recovery parameter $\gamma$ and the testing rate $c$, and of course errors due to noisy data.

Model failure is a possibility too. If the SIR model is not applicable, then the method for finding the reproduction number will not give sensible output. This happens when the number of new cases per day is small, which in practice we find to mean less than about 10 per day.
After the first stage of the process has calculated $R_{\text{eff}}$, these values are fed into the second stage, which generates a large number of daily case data sets by the wild bootstrapping method. These data sets yield confidence intervals for $R_{\text{eff}}$.

3 Model equations, and parameter values

Regard the population as having fixed size $N$, and consider the fraction of the population falling into each of the following categories:

- Susceptible
- Infectious-Untested
- Infectious-Tested
- Removed

This SIR system is illustrated in Figure 1.

3.1 Continuous time model

The model equations are most easily stated using continuous time, that is, for a differential equations model, and so we state them that way first. The continuous time model equations are:

$$\frac{dS}{dt} = -\beta S(I_U + I_T) \quad (1)$$
$$\frac{dI_U}{dt} = \beta S(I_U + I_T) - (c + \gamma)I_U \quad (2)$$
$$\frac{dI_T}{dt} = cI_U - \gamma I_T \quad (3)$$
$$\frac{dR}{dt} = \gamma (I_U + I_T) \quad (4)$$

As a consistency check on any simulation, one should be able to compute that

$$S + I_U + I_T + R = 1$$

for all $t$, since the expression on the left has derivative zero.

Notes on the model:
- The model assumes random mixing among the population.
- Infectious individuals might be either symptomatic or asymptomatic.
- The transmission coefficient $\beta = \beta(t)$ is not constant. It is a function of time $t$.

The equations involve several parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmission coefficient</td>
<td>$\beta(t)$</td>
</tr>
<tr>
<td>$I \to R$ transition rate</td>
<td>$\gamma$</td>
</tr>
<tr>
<td>Testing rate for infectious individuals</td>
<td>$c$</td>
</tr>
</tbody>
</table>

The values of $\gamma$ and $c$ are between 0 and 1, and should be chosen based on best available knowledge of disease behavior and testing regimes. In particular, the testing rate $c$ represents the fraction of Infectious-Untested individuals who get tested per day.
The parameters \( c \) and \( \gamma \) appear explicitly in formulas (10) and (12) below for the effective reproduction number. Thus one can readily perform a sensitivity analysis to see what effect different parameter values have on the estimates of the reproduction number. Our numerical investigations show that the output is qualitatively and quantitatively robust across quite a wide range of plausible parameter values; see [2].

3.2 Discrete time model

In the real world, new cases are reported daily, and so it makes sense to use the following discrete time analogue of the model equations, where \( n \) represents the \( n \)-th day and the time step is 1 day:

\[
\begin{align*}
S(n + 1) - S(n) &= -\beta(n)S(n)(I_U(n) + I_T(n)) \\
I_U(n + 1) - I_U(n) &= \beta(n)S(n)(I_U(n) + I_T(n)) - (c + \gamma)I_U(n) \\
I_T(n + 1) - I_T(n) &= cI_U(n) - \gamma I_T(n) \\
R(n + 1) - R(n) &= \gamma(I_U(n) + I_T(n))
\end{align*}
\]

As a consistency check on any simulation, one should be able to compute that

\[ S + I_U + I_T + R = 1 \]

for all \( n \), since the expression on the left is independent of \( n \) (as one seems by summing the left sides of (5)–(8) and observing that the right sides sum to 0).

3.3 Effective reproduction number

The effective reproduction number is defined by

\[
R_{\text{eff}}(t) = \frac{\beta(t)S(t)}{\gamma} \quad \text{or} \quad R_{\text{eff}}(n) = \frac{\beta(n)S(n)}{\gamma},
\]

depending whether the continuous time or discrete time model is used.

4 Estimating \( R_{\text{eff}} \)

Simple SIR model

For illustrative purposes, we begin with a simple SIR differential system

\[
\begin{align*}
S' &= -\beta SI, \\
I' &= \beta SI - \gamma I, \\
R' &= \gamma I,
\end{align*}
\]

which has just a single class of Infectious individuals. The second equation yields

\[
\beta S = \frac{I' + \gamma I}{I},
\]

and so the effective reproduction number for this simple system is

\[
R_{\text{eff}} = \frac{\beta S}{\gamma} = \frac{1}{\gamma} \frac{I' + \gamma I}{I}.
\]

The right side of this formula involves only the proportion of Infectious individuals, which one can hope to observe experimentally.

In practice, not all Infectious individuals are observable, which is why this report follows James et al. [4] in dividing that group into Infectious-Untested and Infectious-Tested categories. We proceed to analyze that model, seeking a formula for \( \beta S \), and hence for the reproduction number.
4.1 Continuous time SIR system with Infectious-Untested and Infectious-Tested

Now we analyze the differential equation model \((1)\)–\((4)\). The model equation \((3)\) can be rearranged to express \(I_U\) in terms of \(I_T\), as

\[
cI_U = I_T' + \gamma I_T. \tag{9}
\]

Substituting this equation into \((2)\) yields

\[
(I_T' + \gamma I_T)' = \beta S(I_T' + (c + \gamma)I_T) - (c + \gamma)(I_T' + \gamma I_T),
\]

which can be rearranged to obtain the effective reproduction number:

\[
R_{\text{eff}}(t) = \frac{\beta(t)S(t)}{\gamma} = \frac{1}{\gamma} \frac{I_T''(t) + (c + 2\gamma)I_T'(t) + \gamma(c + \gamma)I_T(t)}{I_T(t) + (c + \gamma)I_T(t)}. \tag{10}
\]

4.2 Discrete time SIR system with Infectious-Untested and Infectious-Tested

The difference equation model \((5)\)–\((8)\) is analyzed the same way. The model equation \((7)\) can be rearranged to express \(I_U\) in terms of \(I_T\), as

\[
cI_U(n) = I_T(n + 1) - I_T(n) + \gamma I_T(n). \tag{11}
\]

Substituting this equation into the left and right sides of \((6)\) yields

\[
\begin{align*}
(I_T(n + 2) - I_T(n + 1) + \gamma I_T(n + 1)) - (I_T(n + 1) - I_T(n) + \gamma I_T(n)) & = \beta(n)S(n)(I_T(n + 1) - I_T(n) + (c + \gamma)I_T(n)) - (c + \gamma)(I_T(n + 1) - I_T(n) + \gamma I_T(n)).
\end{align*}
\]

Rearranging the last equation enables us to evaluate the effective reproduction number \(R_{\text{eff}}(n) = \beta(n)S(n)/\gamma\), finding

\[
R_{\text{eff}}(n) = \frac{1}{\gamma} \frac{I_T(n + 2) - 2I_T(n + 1) + I_T(n)}{I_T(n + 1) - I_T(n) + (c + \gamma)I_T(n)}. \tag{12}
\]

This formula depends explicitly on the testing rate \(c\) and “recovery or death” parameter \(\gamma\), as well as on the Infectious-Tested data \(I_T(n)\).

To make use of formula \((12)\) in practice, one needs values for \(I_T(n)\) on some sequence of days. The initial day is not necessarily when the disease began spreading, and need not be when the first case was diagnosed. Section 5 explains how we estimate values of \(I_T\) from the testing data.

Note. The numerator and denominator of \((12)\) should be positive. If they turn out to be negative, then either the real-world data we are relying on is too noisy, or else the real-world epidemic is not following an SIR model. For example, the denominator can be rewritten as \(I_T(n + 1) - (1 - c - \gamma)I_T(n)\). Thus if the number of real-world Infectious-Tested cases decreases in a single day by more than fraction \(c + \gamma\) of its value, one knows the denominator is negative and the reported data for the epidemic is not following the SIR model.

5 Estimating Infectious-Tested data \(I_T\) from daily case numbers

Fully real-world data is not available for the number of Infectious-Tested individuals, since those people are not being tested every day to determine exactly when they recover. Estimated values for \(I_T\) are obtained instead by smoothing out the daily new-case counts over nearby days, and then taking a rolling sum to obtain the cumulative total of Infectious-Tested cases. To estimate the recoveries and deaths, a fraction \(\gamma\) of the cases is subtracted daily, in keeping with the model equation \((7)\).

In more detail, write \(i(n)\) for the number of new Infectious-Tested cases reported on day \(n\), and smooth out fluctuations in the data by averaging over two or three days on either side, as follows. To average over two days on either side, use the 5-point Newton–Cotes rule

\[
j(n) = \frac{1}{90} \left[ 7i(n - 2) + 32i(n - 1) + 12i(n) + 32i(n + 1) + 7i(n + 2) \right].
\]

To average over three days on either side, use the 7-point Newton–Cotes rule

\[
j(n) = \frac{1}{840} \left[ 41i(n - 3) + 216i(n - 2) + 27i(n - 1) + 272i(n) + 27i(n + 1) + 216i(n + 2) + 41i(n + 3) \right].
\]
Comment. In most situations we expect the 7-day formula to do a better job of smoothing out the data. But if, for example, the daily case numbers suddenly drop precipitously, then the shorter 5-day formula will do a better job of preserving that behavior. Note that the averaging process uses some information from the near future and past, and so on the first few and last few days of the data series, a shorter averaging rule such as Simpson’s rule should be employed.

The estimated $I_T$ value on day $n+1$ is now computed from its value on day $n$ by the update formula

$$I_T(n+1) = I_T(n) + \frac{1}{N} j(n) - \gamma I_T(n), \quad n = 1, 2, 3, \ldots,$$

with the initial value $I_T(1) = i(1)/N$ coming simply from the number of cases on that first day. The update formula is based on model equation (7), with the newly Infectious-Tested term $cI_T(n)$ in the model equation being replaced in the update formula by the smoothed new-case fraction $j(n)/N$, which we obtained from real-world data.

The resulting function $NI_T(n)$ gives the estimated number of Infectious-Tested individuals on day $n$.

6 Confidence intervals — quantifying the stability of the deterministic method

Uncertainty in the daily case data

Random errors are introduced when the SIR+B method is applied to noisy real world data.

The data we have used comes from three sources: the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [9], the European Centre for Disease Prevention and Control (ECDC; an agency of the European Union) [11], and USAFacts, a “not-for-profit, nonpartisan civic initiative providing the most comprehensive and understandable government data” [12].

There are differences between these data sets which we cannot explain.

Compare, for example, the French daily new cases as reported by the CSSE and the ECDC (Figure 2). An extreme outlier appears in the CSSE data around 12 April, with over 25,000 new cases reported in just one day. The CSSE justifies its methodologies on github [[https://github.com/CSSEGISandData/COVID-19/issues/2459](https://github.com/CSSEGISandData/COVID-19/issues/2459)] for example the decision to include probable cases as well as confirmed cases — “France has made the explicit decision to not test these cases and they will not be reflected in the official case count. We believe this masks the true extent of the infectious spread in the country and should be captured by those interested in understanding the case burden across time and space.” It is unknown whether the methods of the CSSE correctly adjust for this so called masking of probable cases.

The CSSE French data set [9] has other problems too. On 19 April, the cumulative number of reported cases actually drops, which should of course be impossible. Other drops occur on 22 April, 29 April and 14 May. Further, the number of daily new cases seems unusually volatile in certain periods, such as late April.

Most countries and regions suffer from such problems with the data, although rarely to such an extent as with France.

These issues with the data suggest that a certain amount of smoothing of the data is desirable, thus justifying the use of Newton–Cotes in the previous section. Even with smoothing, a certain variability will remain in the data along with the underlying disease signal, and for that reason we apply bootstrapping.

Bootstrapping method

The wild bootstrap method is used to estimate 50% and 95% confidence intervals for $R_{eff}$. The term bootstrapping is used for any technique involving random re-sampling with replacement. Davidson et al. [11] state, “The wild bootstrap is based on an idea suggested by Wu (1986), and has been explored in detail by Hardle (1989, 1990) and Mammen (1992). The effectiveness of the wild bootstrap, particularly for studentized coefficients, was demonstrated by Mammen (1993).” Bootstrapping methods have been used in epidemiological models that account for the delay between case onset date and notification date [10].

The advantage of bootstrapping methods is that they can be used when the observed process and the nature of the measurement errors is not fully understood. We chose the wild bootstrap method because the number of daily new cases can be unusually volatile in certain periods (for example in the CSSE France data set), and hence the assumption that the measurement errors are normally distributed is not appropriate.

Let us describe wild bootstrapping in general, and then explain how it is used in this paper. To generate bootstrapped observations, first consider a linear model $Y = X\beta + \epsilon$ and fit it by the method of least squares as $y = X\beta + u$, where $X$ is the design matrix, $\beta$ is the least squares predictor, $Y$ is the vector of true values, $y$ is the vector of observations, $\epsilon$ is the (unobservable) error vector and $u$ is the vector of residuals. Then one randomly re-samples the errors, as follows.
The bootstrapping is "wild" because the error of the $i$-th observation is modelled with an independent random variable $F_i$. These variables are chosen [5] such that the expectations satisfy $E(F_i) = 0$, $E(F_i^2) = u_i^2$ and $E(F_i^3) = u_i^3$ where $u_i$ is the $i$-th residual.

Mammen provides three different choices for the random variable $F_i$. The one we choose to work with is $F_i = u_i U_i$ where

$$U_i = (\delta_1 + V_{i,1}/\sqrt{2})(\delta_2 + V_{i,2}/\sqrt{2}) - \delta_1 \delta_2,$$

with $\delta_1 = (3/4 + \sqrt{17/12})^{1/2}$ and $\delta_2 = (3/4 - \sqrt{17/12})^{1/2}$ and where $V_{i,1}$ and $V_{i,2}$ are independent normal random variables with mean 0 and variance 1.

We want to apply this wild bootstrapping method to the log of the daily case data. Unfortunately, some of the residuals are not well defined, because on some days no new cases were reported (meaning we would need to take the log of zero), or the cumulative number of reported cases goes down rather than up, which is impossible. Mammen’s $F_i$ is not well defined for these exceptional observations in the data, and so we need to create somehow residual values for those days. Another issue is that having a small residual $u_i$ on a certain day does not mean we have more certainty in the underlying data for that day (since the data is somewhat noisy). Thus it would seem unjustified to always force a small variance by using the value $u_i$ in the bootstrapped $F_i$.

To address these issues and handle all data points in a consistent fashion, a modified residual $u_i^*$ is defined below. First randomly choose a residual $u_{j(i)}$ from the 5 closest observations for which the residual is well defined (inclusive of the $i$-th observation), with each residual having equal probability of being chosen. If the $i$-th residual is not defined, choose from the nearest two residuals on the left and two on the the right that are well defined. Several exceptional situations must be handled, for example if the residual of a neighboring data point is also not defined. Such points can be dealt with by hand as necessary. Second, normalise the chosen residuals by letting

$$u_i^* = \sigma(u_{j(i)} - \bar{u})$$

where $\bar{u}$ is the mean of the values $u_{j(1)}, \ldots, u_{j(n)}$ and $\sigma$ is the standard deviation of the well defined residuals $u_i$ divided by the standard deviation of the $u_{j(i)}$. This normalization ensures that

$$\frac{1}{n} \sum_{i=1}^{n} u_i^* = 0, \quad \frac{1}{n} \sum_{i=1}^{n} (u_i^*)^2 = \frac{1}{m} \sum_{i} (u_i - \bar{u})^2,$$
where \( \pi \) is the mean of the residual values \( u_i \) that are well defined, with \( m \) being the number of such residuals. We then define \( F^*_i = u_i^* U_i \) and generate a bootstrapped observation at the \( i \)-th data point by randomly re-sampling \( \epsilon^*_i \) from \( F^*_i \) and computing \( y^*_i = X^* \beta + \epsilon^*_i \).

When the model is fitted with the least squares, the residual vector is automatically orthogonal to the columns of the design matrix:

\[
X^T u = X^T (y - X \beta) = X^T y - X^T X (X^T X)^{-1} X^T y = 0,
\]

where we used the standard formula (13) for \( \beta \). Note the vector of modified residuals \( u^* \) need not be orthogonal to the column space.

The Bootstrap method is summarised as

1. Construct the vector of observations and the design matrix \( X \) as follows. The rows of \( X \) correspond to the observations and the columns correspond to the linear polynomial coefficients.
   - Take the first and last day where at least 10 new cases were reported and exclude the data on either side. Label the first day as day 0, the next day as day 1, and so on, up to the last day, which we call day \( \ell \).
   - Exclude the days where less than 1 case was reported, leaving a list of days \( t = (t_1, \ldots, t_m) \) at which we have valid data. (So, \( t_1 = 0 \) and \( t_m = \ell \).) Take the log of the case data on these days, and collect those numbers in a vector \( y = (y_1, \ldots, y_m) \).
   - Select the knot points by choosing the 5th element of \( t \) and subsequently every 7th element. Let the number of knots be \( p \).
   - Initialise \( X \) to have \( m \) rows and \( p + 2 \) columns. Set the entries of the first column to be equal to 1. Set the second column to be equal to \( t \). Each remaining column corresponds to a knot, as follows. To fill in the \( i \)-th column for \( i \geq 3 \), set it to be equal to \( t \), and then subtract the \( (i - 2) \)-nd knot from each entry, and then zero out any negative entries.
   - Solve the least squares problems by setting
     \[
     \beta = (X^T X)^{-1} X^T y. \tag{13}
     \]
     In practice, \( \beta \) is computed by solving the system \( (X^T X) \beta = X^T y \).

2. Compute the residuals \( u = y - X \beta \). At the \( i \)-th data point (inclusive of the days where less than 1 case was reported), randomly choose \( u_{j(i)} \) from the nearest neighbours, compute \( u^*_i \) and define \( F^*_i = u^*_i U_i \) for \( i = 1, \ldots, n \) (see above).

3. Generate bootstrapped observations by randomly re-sampling the error \( \epsilon_i \) from \( F^*_i \) to arrive at \( y^*_i = X_{ij} \beta_j + \epsilon_i \).

4. Compute \( R_{\text{eff}} \) for the bootstrapped data set.

5. Repeat steps 2 - 4 a large number of times (e.g. 1000 times). Smooth the raw and bootstrapped data on the linear scale 3 times with the 3-point Newton-Cotes rule, the 7-point Newton-Cotes rule and the 3-point Newton Cotes rule respectively.

6. Compute \( R_{\text{eff}} \) for the raw data and bootstrapped data. Construct a 50% and 95% confidence interval by taking the 5th, 25th, 75th and 95th percentiles of the bootstrapped \( R_{\text{eff}} \)’s for each day.

The wild bootstrap produces artificial daily case data which compares well to the raw data. For example, Figure 3 shows the daily case data for France (data source ECDC) together with daily case data generated by bootstrapping.

7 Results

In a separate technical report Grice et al. [2] have used the SIR+B process to calculate \( R_{\text{eff}} \) for several countries. In that technical report the authors compared the SIR+B results with the results of other statistical approaches. We refer the reader to the technical report to assess the value of the SIR+B process when applied to real world data.

8 Conclusion

A SIR+B process for estimating the effective reproduction number from daily case data has been developed. The process has the following desirable attributes:

- it is fast, taking about 0.1 seconds per country analysed, when coded in Python on a standard desktop machine,
Figure 3: Top: ECDC daily new cases (France) versus synthetic daily case data summarised with box plots for each day. The whiskers extend to the top and bottom 5% of the data. The hollow squares show the daily case data smoothed with a 7 point Newton-Cotes smoothing scheme (the smoothing scheme was applied on the linear scale). Bottom: The top plot on the log scale.

- it is based directly on the epidemiological SIR model, giving an analytic formula for the effective reproduction number,
- it involves many fewer assumptions than other methods such as EpiEstim,
- the calculation of $R_{\text{eff}}$ depends on only two parameters, $\gamma$ and $c$,
- it is relatively insensitive to those parameters (see [2]).

Thus the paper demonstrates the potential of the SIR+B process to inform public health policy in the immediate future.

A Using Synthetic data to reconstruct deterministic $R_{\text{eff}}$ values

In this appendix, we test the deterministic part of the SIR+B process on a numerically exact solution of the SIR system that features rapid changes in $R_{\text{eff}}(t)$. Consider the SIR system (5)–(8) over a full year ($1 \leq n \leq 365$), with parameter values $\gamma = 1/15$, $c = 1/15$, and initial conditions

$$S(1) = 1 - 10/(5 \times 10^6), \quad I_U(1) = 10/(5 \times 10^6), \quad I_T(1) = 0, \quad R(1) = 0, \quad D(1) = 0.$$  

That is, the epidemic begins with 10 infected people (untested) in a population of 5 million.

The transmission coefficient $\beta(n)$ that is applied to the system is chosen to exhibit rapid changes between two levels, as shown in Figure 4 corresponding to normal societal conditions ($\beta/\gamma = 2.5$) and severe lockdown conditions ($\beta/\gamma = 0.5$). These values correspond to the “$R_0$” values that a fully susceptible population ($S = 1$) would experience, under these conditions.
Figure 4: Left: plot of the applied reproduction number $\beta(n)/\gamma$ in the test of the method, in Section A. Right: The Infectious-Untested $I_U$ (red, higher peak) and Infectious-Tested $I_T$ (black, lower peak) curves computed for that discrete time SIR system.

Figure 5: Left: plot of the exact $R_{\text{eff}}(n)$, which equals the applied reproduction number $\beta(n)/\gamma$ multiplied by the computed Susceptible fraction $S(n)$. Right: the effective reproduction number $R_{\text{eff}}(n)$ computed numerically from (12), by using the rounded, exact new-case value $i(n+1) = \text{Round}(NcI_U(n))$ for day $n+1$ in order to get approximate values for $I_T$ as in Section 5. The computed values on the right correctly pick out the oscillations in $R_{\text{eff}}$, showing that the method has adequate resolution even in the face of repeated changes in the applied reproduction number (i.e., changes in lockdown conditions).

Solving the discrete time SIR system numerically yields the plots of $I_U(n)$ and $I_T(n)$ shown on the right of Figure 4. The effective reproduction number $R_{\text{eff}} = \beta S/\gamma$ is plotted on the left of Figure 5. It drops off somewhat, because $S(n)$ decreases with time.

This oscillating $R_{\text{eff}}$ is detected by the method, on the right of Figure 5. Notice the method is not reliable until about day 30. Before that time, the number of daily new tested cases is very small (less than 10), and our rounding of the underlying solution to give integer-valued case numbers is enough to throw off the method. After day 30 the number of cases gets larger, and the method starts giving sensible estimates for the reproduction number.

This synthetic result gives reason to believe that the method can reproduce $R_{\text{eff}}$ for real world data also. It is important to note, though, that the daily case numbers in this example are derived from an exact computed solution to an SIR system, so that even after rounding, the approximate $I_T$ values being fed into formula (12) are high on signal and low on noise. Real data for real populations will not be as favorable, as is seen when using the method to estimate $R_{\text{eff}}(t)$ using real-world data [2].
References


