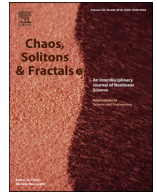




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Frontiers

Insights into the dynamics and control of COVID-19 infection rates

Mark J. Willis^{a,*}, Victor Hugo Grisales Díaz^{b,c}, Oscar Andrés Prado-Rubio^c, Moritz von Stosch^d

^a School of Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

^b Grupo de investigación en Microbiología y Biotecnología MICROBIOTEC, Facultad Ciencias de la Salud, Universidad Libre Seccional Pereira, Belmonte Avenida Las Américas

^c GIANT, Departamento de Ingeniería Química, Universidad Nacional de Colombia - Sede Manizales, Campus La Nubia, Km 9 vía al Aeropuerto La Nubia, Bloque L, Colombia

^d DataHow AG, Zurich, Switzerland

ARTICLE INFO

Article history:

Received 20 May 2020

Accepted 25 May 2020

Available online 28 May 2020

Key words:

COVID-19

Model calibration

Predictive modelling and simulation

Time varying disease transmission rate

ABSTRACT

This work aims to model, simulate and provide insights into the dynamics and control of COVID-19 infection rates. Using an established epidemiological model augmented with a time-varying disease transmission rate allows daily model calibration using COVID-19 case data from countries around the world. This hybrid model provides predictive forecasts of the cumulative number of infected cases. It also reveals the dynamics associated with disease suppression, demonstrating the time to reduce the effective, time-dependent, reproduction number. Model simulations provide insights into the outcomes of disease suppression measures and the predicted duration of the pandemic. Visualisation of reported data provides up-to-date condition monitoring, while daily model calibration allows for a continued and updated forecast of the current state of the pandemic.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Since the emergence of SARS-CoV-2, more commonly known as coronavirus, the related disease (COVID-19) has quickly spread around the world. On the 11th of March 2020, the WHO declared it a pandemic. Numerous features of COVID-19 make control of the pandemic challenging. The virus is new, its dynamics are therefore uncertain, and there are no effective cures or vaccines. Besides, data is incomplete, with the potential of a significant number of unknown (recovered and active) cases and an uncertain incubation period. Current suggestions are that 40 – 50%, with up to 81% of cases only present only mild symptoms, see discussions in [1]. However, a substantial number of known infected individuals eventually develop severe pneumonia, which requires treatment in intensive care units. Without effective measures to control the disease, the number of probable severe cases would disrupt public health systems leaving patients without crucial life support, potentially exacerbating fatality rates [2].

1.1. Non-pharmaceutical interventions

As of 16th May 2020, 4.65 million cases, with 1.7 million recovered patients and 300,000 deaths due COVID-19 have been reported in 213 countries worldwide [3]. All over the world, governments have issued recommendations such as intensified hand hygiene and have taken measures such as closing borders, enforcing lockdowns, etc. These non-pharmaceutical interventions (NPIs) reduce infection rates, keeping the number of severe cases below hospital capacity limit. A strategy popularly referred to as ‘flattening the curve’. Estimates of the effect various types of NPIs have on the reduction of the effective reproduction number are provided in [2] where it is reported that lockdown leads to an average reduction of the transmission of the disease by 50%, school closure, 20% and other measures around 10% (there are relatively large confidence bounds around these figures).

Because of stringent lockdown measures, the disease appears to have been contained in Wuhan, China. Following a lockdown, within a relatively short period (around ten days), the actual number of cases stopped increasing and began to drop exponentially. While China adopted an aggressive strategy of almost total lockdown, the measures adopted in other counties typically have been less “draconian”. However, the containment strategies do appear to be working, with a number of countries, including Germany, Italy

* Corresponding author.

E-mail address: mark.willis@ncl.ac.uk (M.J. Willis).

and South Korea, having passed the peak of cumulative infectious cases.

1.2. Effective management of the reproduction number

In the absence of a proven vaccine, effective management of the long-term return to normality is required. This should consider both the minimisation of death rates and economic impact. Key parameters used by epidemiologists to quantify contagion are the basic and effective reproduction numbers. These dimensionless numbers describe the average number of expected secondary infections generated by each infected person for a completely susceptible population in the absence and presence of controlled interventions, respectively. Epidemiologists track the effective reproduction number through verification of the number of secondary infected arising from confirmed infected cases. This is difficult to estimate accurately; it requires a robust regime of contact tracing and the capacity to perform large numbers of tests to confirm whether individuals are infected.

Current opinion suggests that the COVID-19 has a basic reproduction number¹ $\sim 2 - 3$. NPIs slow the spread of the virus, and as a result, there is a reduction in the effective reproduction number over time. If it is possible to maintain this number below one, then the pandemic will eventually die out. Keeping the effective reproduction number either close to, or below one, will require careful adjustment of NPIs².

1.3. Hybrid modelling and predictive simulation

In this work, hybrid models of the dynamics of COVID-19 infection rates are developed³. The hybrid model must capture essential system dynamics for calibration, predictive modelling and simulation studies to be meaningful. These are an initial exponential growth in active cases, slowing as the NPIs influence disease transmission. After a peak in the number of cumulative cases, there will typically be a slow decline in active cases, provided the NPIs are not excessively relaxed.

Several efforts around the world are attempting to forecast the behaviour of COVID-19 transmission rates. The methods used include mechanistic transmission models [4], statistical models [5, 6], deterministic epidemiological models [7–9] and a statistical dynamical growth model [10]. In this work, a novel hybrid model is developed that may be used for forecasting and monitoring the effective reproductive number which several countries are using as an indicator to measure the effectiveness of NPIs. As an established backbone of fundamental knowledge, we use a susceptible – infected – removed (SIR) compartmental ordinary differential equations model [11–13]. The equations are augmented with a differential equation that models an exponential decline in the disease transmission rate, and hence the effective reproduction number. Daily monitoring and model calibration (through a customized model parameter estimation) allows classification of the potential

and severity of the outbreak and an assessment of the effectiveness of current NPIs. Model simulation allows predictions to be made of the expected time course of the epidemic, which is conditional upon known current and past reported data.

2. Methods

Up-to-date daily information regarding the number of active, recovered and fatal cases for most countries around the world is available from [3]. Retrieving this information allows daily calibration of the hybrid model, on a country-by-country basis, as well as the monitoring of the status of the pandemic.

2.1. Mathematical modelling

Defining $S(t)$ as the number of susceptible, $I(t)$ the number of infected and $R(t)$ the number of removed (recovered and deceased) individuals, from a constant population N at time t the differential equations of the Susceptible – Infectious – Removed (SIR) model are,

$$\frac{dS(t)}{dt} = -\frac{\beta(t)}{N} \cdot S(t)I(t) \quad (1)$$

$$\frac{dI(t)}{dt} = \frac{\beta(t)}{N} \cdot S(t)I(t) - \gamma \cdot I(t) \quad (2)$$

$$\frac{dR(t)}{dt} = x \cdot \gamma \cdot I(t) \quad (3)$$

In these equations, $\beta(t)$ represents the transmission rate of the virus (day^{-1}) while γ (day^{-1}) is the removal rate of reported infectious individuals, both of which can be time-dependent parameters. The transmission rate varies because of NPIs and the data-driven component of the hybrid model captures the dynamics of disease suppression. Here, a constant removal rate is used. However, it can also change with respect to time: as normally, the chance of recovery in a particular time interval is initially small. That chance of recovery then increases over time [see discussions in, e.g. [14]].

The additional parameter x is included within the equations to account for under-reporting of removed individuals from the infectious class. Emphasis around the world has focused on the reporting of infectious cases (typically hospital admissions) and the subsequent death rates of these patients. A relatively recent report [15] assessed the extent of under-reporting of symptomatic COVID-19 cases globally, and their results indicated significant under-reporting by a large number of countries. Furthermore, this number can be highly variable as case data can be missing, incomplete, incorrect, based on different definitions, or dated (or a combination of all of these)⁴. The reported fraction x and an under-reported fraction $(1 - x)$ maintain the conservation relationship inherent within the SIR model.

Using the model, provided the number of infected cases is relatively small compared to the total population size, so that, $S(t) \approx N$, the effective reproduction number at time t is given by,

$$R_E(t) = \frac{\beta(t)}{\gamma} \quad (4)$$

2.2. Capturing the variation in disease transmission rate

Social distancing and lockdown measures reduce the transmission of the virus and the number of active infected cases. In terms of the SIR model, this translates into a reduction in the value of

⁴ As an extreme example, the UK government currently do not publish cumulative recovered cases.

¹ Although a recent review [22] compared 12 studies published from the 1st of January to the 7th of February 2020 have estimated the basic reproduction number for COVID-19, finding a range of values between 1.5 and 6.68.

² The media have discussed Germany's use of this number to determine its exit strategy frequently. Apparently, Chancellor Angela Merkel uses a daily estimate to track how the gradual easing of social distancing measures and freeing up parts of the German economy.

³ A hybrid semi-parametric (hybrid) model combines fundamental, often referred to as mechanistic, models with data-driven or heuristic-driven models. These models make use of fundamental knowledge and complement the description, where necessary, with nonparametric approaches. The attractive traits of hybrid models are their good extrapolation capabilities, low data requirements and the transparent structure, which helps to develop process understanding [23]. A SIR model provides a fundamental backbone representing overall behaviour of an epidemic, augmented with a data driven component to improve predictive performance.

$\beta(t)$. Using a first-order linear ordinary differential equation (ODE) to capture the variation in disease transmission rates, models any reduction as an exponential decay,

$$\tau \cdot \frac{d\beta'(t)}{dt} + \beta'(t) = -k\bar{u}(t), \quad \beta'(0) = 0, \quad \bar{u}(0) = 0$$

$$\beta'(t) \geq 0, \quad \forall t$$

$$\beta(t) = \beta_I + \beta'(t) \quad (5)$$

The time-dependent variable $\bar{u}(t)$ models the effect of NPIs. This has a range between zero and one. A value of $\bar{u}(0) = 0$ implies no NPIs and a constant disease transmission rate, i.e. $\beta(t) = \beta_I$. A value of $\bar{u}(t \geq t_c) = 1$ represents the maximum control effort/policies applied in any particular country because of the cumulative effect of NPIs initially introduced to reduce disease transmission. The time t_c is the time when the maximum control measures are introduced. In the differential equation, the time constant τ defines the rate of exponential decrease after the introduction of NPIs, and the gain parameter k is the magnitude of the final reduction in disease transmission rate, $\beta(t)$. As a country relaxes (or strengthens again) the applied NPIs, the value of $\bar{u}(t)$ may be altered to account of any change. Defining a time t_{ri} an estimate can be made of the reduced control effort, $\bar{u}(t \geq t_{ri}) = \hat{u}_{ri}$. This provides a quantitative measure of the relative relaxation of the NPIs in any particular country and updated predictions of the dynamics of the number of cumulative infected individuals.

2.3. Model calibration

To quantify the discrepancy between model response and reported case data, the sum of the squared error between the normalized reported cumulative infected individuals $I_R(t)$ and the model prediction $I(t)$ as well as the reported removed individuals, $R_R(t - t_d)$ and the model prediction $R(t)$ is calculated as,

$$J = \sum_{t=t_1}^{t_2} \left(\frac{I_R(t) - I(t)}{I_{Rmax}} \right)^2 + \sum_{t=t_1}^{t_2} \left(\frac{R_R(t - t_d) - R(t)}{R_{Rmax}} \right)^2 \quad (6)$$

The sum is over the time-period t_1 (the initial outbreak reaching exponential growth) and t_2 (current time), I_{Rmax} is the maximum of the reported infected individuals and R_{Rmax} the maximum reported removed individuals. The parameter t_d represents a time delay associated with the reporting of recovered / deceased individuals. This delay accounts for the time it takes for confirmation of deaths, recoveries, or the validation of data from tests.

To calibrate the hybrid model, a gradient-based optimisation algorithm, *lsqnonlin*⁵, minimises (6) by adjusting unknown model parameters. This involves repeated numerical solution of the ODEs, coded in Simulink to accommodate the time delay term and integrated using MATLAB's built-in routine *ode15s* for efficient integration of stiff ODEs. The use of *lsqnonlin* allows the additional MATLAB function *nlparci*⁶ to calculate the 95% confidence intervals on the model parameters.

Using a constant value for the removal rate⁷, $\gamma = \frac{1}{8} \text{day}^{-1}$ [16] successful minimisation of (6) using reported case data, pro-

vides an estimate of the numerical values of the hybrid model parameters, β_I , τ , k , x , the time delay, t_d and their associated confidence bounds. For countries where NPI relaxations are being introduced, given the timing of these relaxations, t_{ri} , an estimate can also be made of the reduced control effort, $\bar{u}(t \geq t_{ri}) = \hat{u}_{ri}$ during model calibration.

2.4. Data screening for model calibration

As discussed in [17] initially, prior to an outbreak becoming fully established, daily new case numbers are relatively low and the cumulative number of infectious cases typically grows at a sub-exponential rate [18]. However, an established outbreak driven mainly by community transmission is more likely after around one hundred reported cumulative cases [19]. Following [17], it is assumed that the initial time of the outbreak as the day when there are (as close as possible) one hundred reported cumulative cases in a particular country. This provides the start-time for data visualisation as well as the initial conditions for the ODEs.

For each country, the timing of the introduction of the most severe NPI's defines the control time t_c and hence, $\bar{u}(t \geq t_c) = 1$. While most countries introduced NPIs in a staged manner, e.g. suggesting improved hygiene regimes, closing entertainment venues, closing schools, to full lockdown measures, we assume that the speed of introduction of these measures was generally rapid such that the subtleties of any fast ramp in interventions would not be discernible from the inaccuracies in the data. However, we note that recent work [20] quantify the effect of staged interventions in Germany, where large events were cancelled, followed by the closing of schools and many shops prior to a final intervention on March 23rd 2020.

3. Results

To demonstrate and discuss our predictive modelling, we use reported case data from Spain, Germany, Sweden and South Korea. Predictions for a selection of other countries are in the supplementary material. These selected countries have a reasonable COVID-19 testing capacity (test coverage per thousand residents is over 15), which would imply reasonably robust data. As with any data-driven approach, it will only be possible to have representative and reliable results if the data is also reliable.

Fig. 1 presents the actual cumulative numbers of active cases as well as the hybrid model prediction for Spain using data up to 16th May 2020. Around three weeks ago, Spain 'flattened the curve', and since then the number of cumulative cases has been decreasing.

At the peak, Spain had around 100,000 cumulative infected cases, and the hybrid model predicts it will take a significant time to reduce the number of active cases. The red line shows the model prediction of cumulative active cases ($R_{\text{squared}} = 99.2\%$). Currently, this indicates that this will not reach a lower threshold of 100 cases (our chosen threshold) until between 10th November 2020 and the 19th December 2020. Note that this assumes that there is no relaxation in current NPIs⁸. At this stage, and as may be expected, the prediction of the time to reach the chosen threshold of cases is not particularly accurate with a wide range in the final forecast because of the model confidence bounds. However, through continuous monitoring of case data and continual model re-calibration, we will obtain improved predictions.

model parameters altered the value of β_I whereas the other model parameters remaining relatively constant. Moreover, this random variation in γ had no detrimental effect on the minimum value of the cost function (6).

⁸ Spain started tentatively relaxing some NPIs around the 4th May 2020 although we have not accounted for this in our model. We initially ran a model allowing the adjustment of the level of, u , from the 4th May 2020 and the confidence bounds indicated that this term was not significantly different from zero.

⁵ This is a nonlinear least squares algorithm based on trust-region-reflective method (default) that handles bound constrained optimization, available within MATLAB's optimisation toolbox

⁶ Function available in MATLAB's statistics toolbox. Calculation of the parameter confidence intervals is based on the covariance matrix and t-student test for normal distribution.

⁷ The simultaneous estimation of unique numerical values for the model parameters β_I and γ is not possible. Consider equation (2), with $\beta(t) = \beta_I$ and $S \approx N$. This implies, $\frac{dI}{dt} \approx (\beta_I - \gamma)I$. Therefore, it is only possible to identify the relative difference between the two parameters. Monte Carlo simulations verified this. Randomly adjusting γ , around the mean value of $\gamma = \frac{1}{8} \text{day}^{-1}$ and identifying the remaining

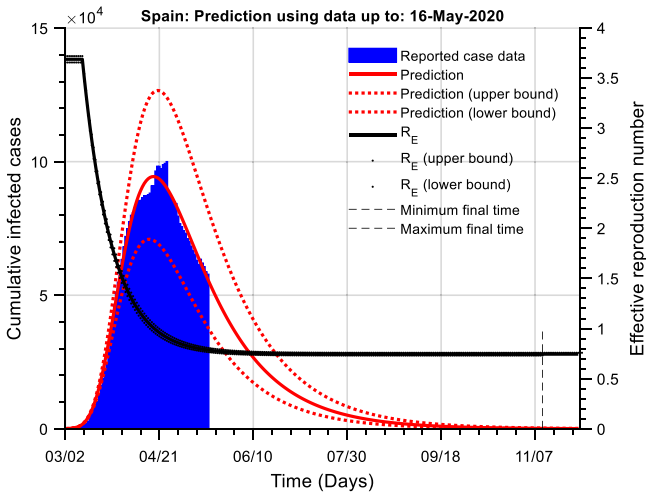


Fig. 1. The cumulative number of active infected cases as a function of time (Spain). The red line is the prediction of the actual number of cumulative active cases (blue) using our dynamic model. The black line is an estimate of the effective reproduction number. The dotted lines show the predictor confidence intervals. Simulation of our dynamic model allows for a prediction of the time when the cumulative infected cases is lower than a threshold; this threshold was set at 100 cases. The current estimate is that the number of cases will be lower than this threshold around 10th November 2020 – 19th December 2020.

The current calibrated model constants obtained using the Spanish data are,

β_I	k	τ	$x(\%)$	t_d
0.460 ± 0.004	0.368 ± 0.002	15.74 ± 1.12	41.62 ± 1.12	1.215 ± 0.822

These provide an estimate of the basic reproduction number, $R_0 = 3.68 \pm 0.03$ and a current effective reproduction number⁹ of, $R_E = 0.74 \pm 0.01$. The model calibrations would also indicate a significant under-reporting of removed individuals. While this number is not required to predict the number of active cumulative cases, it is necessary for model calibration to the reported number of recovered patients. Note also that the estimated time constant of the disease suppression dynamics is relatively large at around 2 weeks.

Fig. 2 shows the cumulative active cases and a predictive simulation for Germany. As may be observed, the curve of cumulative active cases has passed their initial peak and has been declining for a significant number of days. Indeed, Germany first began to relax NPIs around April 20th 2020. Therefore, to calibrate the hybrid model, as well as estimating the unknown model parameters an estimate is also made of the relative reduction in NPIs using, $\bar{u}(t \geq t_{ri}) = \hat{u}_{ri}$. Assuming t_{ri} is April 20th 2020 gives the following calibrated model constants,

β_I	k	τ	$x(\%)$	t_d	$\Delta \hat{u}_{ri}$
0.395 ± 0.002	0.358 ± 0.006	17.62 ± 0.90	53.14 ± 0.69	5.05 ± 0.837	-0.11 ± 0.03

These provide an estimate of the basic reproduction number, $R_0 = 3.16 \pm 0.1$ and a current effective reproduction number¹⁰ of, $R_E = 0.61 \pm 0.03$. Note, again that the model calibrations would also indicate a significant under-reporting of removed individuals.

Fig. 2 also shows the model predictions ($R_{\text{squared}} = 99.4\%$), with associated confidence bounds; indicating that the cumulative number of active cases would reach a lower threshold of 100 around the 22nd – 30th August 2020. The estimate of the relative NPI relaxation since April 20th 2020 is around 11 %. To obtain this figure,

⁹ Currently, [4] report this value as 0.88 (0.75 – 1.06)

¹⁰ Current media reports place it closer to one. Currently [4] report this value as 0.76 (0.53 – 1.07).

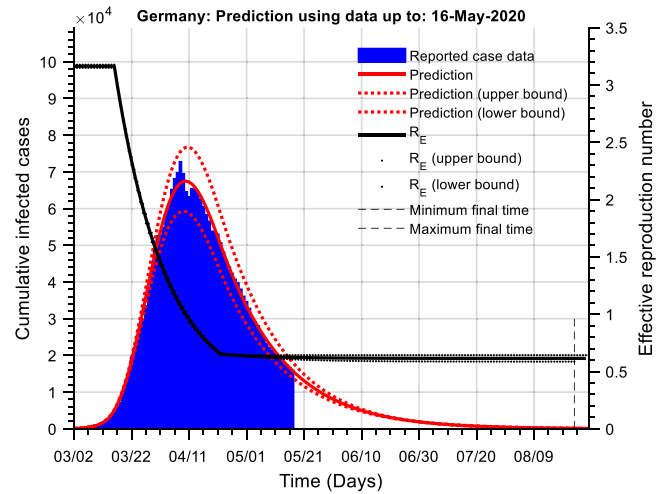


Fig. 2. The cumulative number of active infected cases as a function of time (Germany). The red line is the prediction of the actual number of cumulative active cases (blue) using our dynamic model. The black line is an estimate of the effective reproduction number. Simulation of our dynamic model allows for a prediction of the time when the cumulative infected cases is lower than a threshold; this threshold was set at 100 cases. The current estimate is that the number of cases will be lower than this threshold around the 22nd – 30th August 2020.

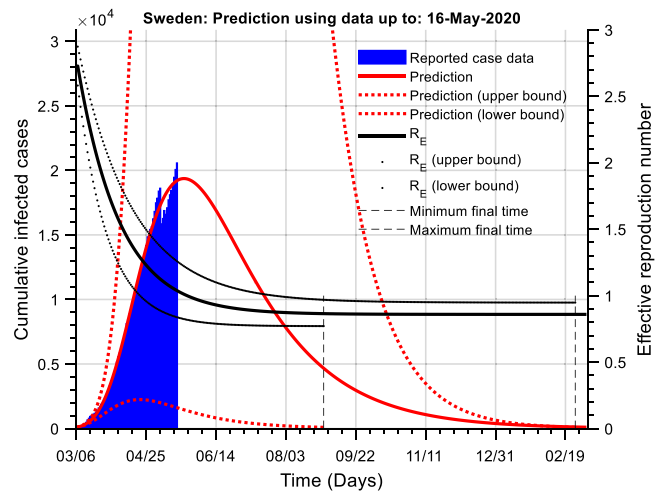


Fig. 3. The cumulative number of active infected cases as a function of time (Sweden). The red line is the prediction of the actual number of cumulative active cases (blue) using our dynamic model. The black line is an estimate of the effective reproduction number. The confidence bounds are large, as is the estimation of the effective reproduction number and the time to achieve a lower threshold of 100 cases. The current estimate is that the number of cases will be lower than this threshold around December 2020 – mid February 2021.

we calibrate the model initially assuming that $\bar{u}(t \geq t_c) = 1$ and estimate the reduction from this number as, $\Delta \hat{u}_{ri}$ giving, $\bar{u}(t \geq t_{ri}) = 1 + \Delta \hat{u}_{ri}$. Model calibration is relatively insensitive to a change in the assumed day that the NPIs were relaxed. Model calibration was still possible when t_{ri} was varied by five days on either side of April 20th 2020. However, this does affect the predicted time to get to the lower threshold of 100 active cumulative cases.

Fig. 3 presents the cumulative active cases as well as their prediction for Sweden ($R_{\text{squared}} = 99.4\%$). Sweden is the only European country not to implement full lockdown measures. They did implement NPIs, where the measures in Sweden include closing of universities and high schools and asking older and at-risk residents to avoid social contact while keeping restaurants and primary schools open. As may be observed from the figure, the curve of cumulative active cases may peak in the coming days.

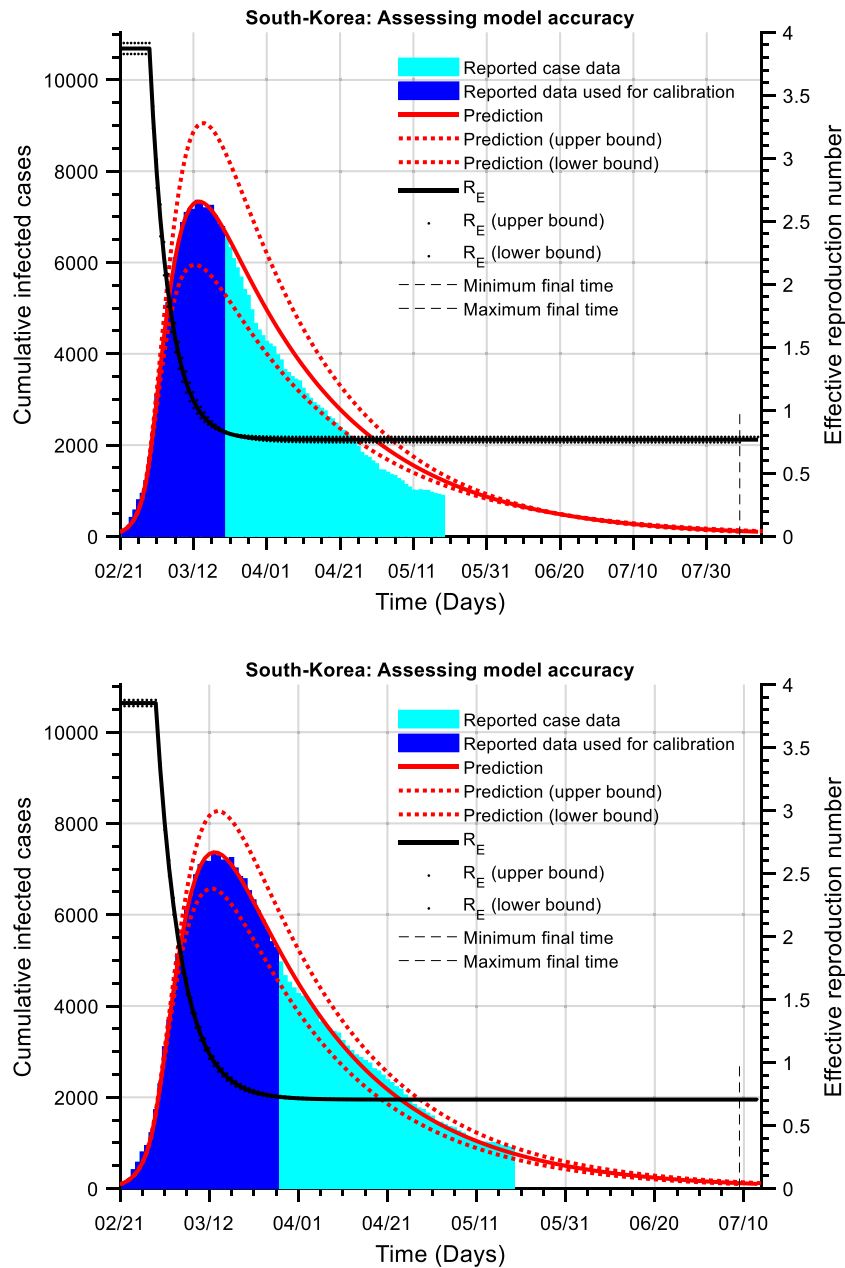


Fig. 4. The cumulative number of active infected cases as a function of time (South Korea). The red line is the prediction of the actual number of cumulative active cases (blue and cyan) using our dynamic model. Model calibration using the blue data (data up to the 19th April 2020). Comparison of model simulations to the assumed ‘unknown’ reported case data (cyan). The black line is an estimate of the effective reproduction number.

In terms of model calibration, there are relatively large confidence bounds associated with the parameter estimates,

β_1	k	τ	$x(\%)$	t_d
0.341 ± 0.018	0.234 ± 0.007	30.28 ± 8.70	17.20 ± 2.05	9.92 ± 2.43

In turn, this leads to large bounds in the estimate of the cumulative number of effective cases, the effective reproduction number and the time to reach a lower threshold of 100 cumulative active cases. In the fullness of time, these issues will be resolved. Once the reported number of cases have passed the peak and they continue to decline, it will be possible to obtain a significantly more robust estimation of the model parameters and improved model predictions. Currently, the hybrid model parameters indicate the basic reproduction number, $R_0 = 2.73 \pm 0.15$ and a current effective reproduction number close to one.

3.1. Long-range forecasts are always uncertain

If there were a first law of forecasting, it would be that long-range forecasts are always uncertain. In the present context, this may arise because of future changes in NPIs, which, when known to occur have to be accommodated in future predictions. To demonstrate the forecasting accuracy of the hybrid model, we use reported case data from South Korea and Germany. South Korea have successfully ‘flattened the curve’; they acted rapidly and decisively to suppress the disease through contact tracing using mobile phones and a rigorous testing regime, and have far fewer cases and deaths than the majority of countries around the world [21].

Fig. 4 shows a prediction of the number of cumulative active cases (red curve) and the effective reproduction number (black curve) of the disease.

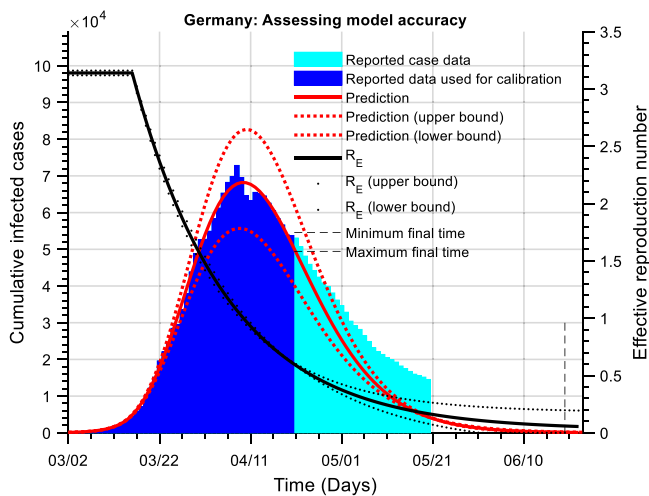


Fig. 5. The cumulative number of active infected cases as a function of time (Germany). The red line is the prediction of the actual number of cumulative active cases (blue and cyan) using our dynamic model. Model calibration using the blue data (data up to the 19th April 2020). Comparison of model simulations to the assumed 'unknown' reported case data (cyan). The black line is an estimate of the effective reproduction number. Observe that inaccurate model predictions as the forecast does not account for the relaxation of NPIs, which have a clear and observable effect on the dynamics of the number of active cases.

The hybrid model was calibrated using reported case data up to one week (top figure) then two weeks (bottom figure) after the flattening of the peak. Observe the accuracy of the model predictions. Model calibration using data up to one week after the turning point produces an inaccurate forecast of the number of active cases although the dynamics show a similar trend. However, with the additional data, the calibrated model provides a much-improved forecast of the actual number of infected cases.

The hybrid model provides predictions of the disease suppression dynamics and time to reach a low threshold of cases within the bounds of ~30 and ~50 days, using data for model calibration up to one and two weeks after the turning point in the cumulative number of active cases. The estimated times being 7th – 27th August 2020 (using data up to one week after the peak) and with the use of additional data (two weeks after the peak), this estimate changes to 28th June – 7th July 2020. After, recalibration of the hybrid model using data up to the 19th May 2020 the current estimate is the 13 – 19th July 2020.

Fig. 5 shows the results of a model calibration for Germany using reported case data up-to the 19th April 2020 (before relaxation of NPIs). While the future forecast captures the qualitative behaviour of the response, the predicted number of cases are inaccurate (compare the prediction to the subsequently reported actual data). The model predicts that the number of cases will reduce at a faster rate, but because of NPI relaxation, this rate has actually slowed. However, as demonstrated in Fig. 2 daily re-calibration of the hybrid model can account for any observable changes, caused, for example, through relaxation of current NPIs. This allows an up to date forecast to be made of the future decline in cumulative active cases.

4. Discussion and conclusions

We describe a hybrid model of the dynamics of cumulative active cases because of COVID-19 and a methodology for the model calibration. Model simulation allows predictions to be made of the expected time course of the epidemic, which is conditional upon available current and past reported data. We have highlighted the trend in the cumulative number of active cases in Spain, Ger-

many, Sweden and South Korea. In the supplementary material, we provide additional figures demonstrating the status of the pandemic in a selection of other countries (Italy, France, Switzerland, Turkey, USA and Colombia). Our results would indicate that the data-driven component of our hybrid model accurately captures an exponential decline in disease transmission.

We elected to keep our hybrid model comparatively simple using the established *SIR* model as a fundamental backbone to a data driven component that modelled an exponential decline in disease transmission rate. We find that this exponential decline adequately represented by a first order linear ordinary differential equation. Model calibration using reported case data requires the determination of four model parameters (β_i , k , τ , x) as well as a time delay associated with the reporting of removed cases, t_d . As NPIs are relaxed we can introduce additional model terms that may be used to quantify the relative relaxation of NPIs. Defining a time t_{ri} that an NPI relaxation occurs then an estimate can be made of the reduced control effort, $\hat{u}(t \geq t_{ri}) = \hat{u}_{ri}$. We demonstrated how this affects future predictions of the number of active cases using reported case data from Germany. As more case data becomes available, we will rigorously test the viability of including phased NPI changes as part of the hybrid model calibration.

Regarding alternative work that considers forecasting the dynamics of the reduction in the number of active cases of COVID-19, we note that [9] has adopted a similar approach to the estimation of the time it would take to achieve a lower threshold of active cases. However, calibration of their *SIR* model only considers daily cases with respect to time, recovered data is not considered and the method does not explicitly account for a varying transmission rate in the model equations or the tracking of, $R_E(t)$.

In conclusion, our hybrid modelling approach allows the prediction of future infected case number scenarios as well as providing a measure of the effect of the relaxation of government interventions. Using a linear ODE, we capture the dynamics associated with disease suppression. For the countries we consider, the estimated time constant, τ , associated with the disease suppression dynamics is large (over 2 weeks for both Spain and Germany). In making a change to a dynamic system, the time constant provides an indication of the speed of the response and it takes approximately 3 – 4 time constants to move from the original to the final state. This would suggest that the easing of lockdown measures would be a difficult task requiring patience and careful planning.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.icarus.2017.10.034](https://doi.org/10.1016/j.icarus.2017.10.034).

References

- [1] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, "How will country-based mitigation measures influence the course of the COVID-19 epidemic?" *The Lancet*, vol. 395, no. 10228. Lancet Publishing Group, pp. 931–934, 21-Mar-2020.
- [2] N. M. Ferguson et al., "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," 2020.
- [3] *Worldometer Coronavirus Cases*. *Worldometer 2020:1–22*.
- [4] I. C. The MRC Centre for Global Infectious Disease Analysis, "Short-term forecasts of COVID-19 deaths in multiple countries," 2020. [Online]. Available: <https://mrc-ide.github.io/covid19-short-term-forecasts/index.html>. [Accessed: 19-May-2020].
- [5] U. of G. Institute of Global Health, "COVID-19 Forecasting," 2020. [Online]. Available: <https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/>. [Accessed: 19-May-2020].
- [6] U. of T. COVID-19 modeling consortium, "COVID-19 Mortality Projections for US States and Metropolitan Areas," 2020. [Online]. Available: <https://covid-19.tacc.utexas.edu/projections/>. [Accessed: 19-May-2020].

- [7] C. Analytics, "DELPHI Epidemiological Case Predictions," 2020. [Online]. Available: <https://www.covidanalytics.io/projections>. [Accessed: 19-May-2020].
- [8] U. Statistical Machine Learning Lab, "UCLAML Combating COVID-19," 2020. [Online]. Available: <https://covid19.uclaml.org/>. [Accessed: 19-May-2020].
- [9] J. Luo, "When will COVID-19 end - Data-Driven Estimation of end dates," 2020. [Online]. Available: <https://ddi.sutd.edu.sg/when-will-covid-19-end>.
- [10] N. L. Los alamos, "COVID-19 Confirmed and Forecasted Case Data." [Online]. Available: <https://covid-19.bsvgateway.org/>. [Accessed: 19-May-2020].
- [11] Kermack W, McKendrick A. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character Aug. 1927;115(772):700–21.
- [12] Kermack W, McKendrick A. Contributions to the mathematical theory of epidemics. II.—The problem of endemicity. Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character Oct. 1932;138(834):55–83.
- [13] Kermack W, McKendrick A. Contributions to the mathematical theory of epidemics. III.—Further studies of the problem of endemicity. Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character Jul. 1933;141(843):94–122.
- [14] Lloyd AL. Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics. Theor. Popul. Biol. Aug. 2001;60(1):59–71.
- [15] Russell TW. Using a delay-adjusted case fatality ratio to estimate under-reporting | CMMID Repository. et al 2020. [Online]. Available https://cmmid.github.io/topics/covid19/global_cfr_estimates.html Accessed: 19-May-2020.
- [16] Sreevalsan-Nair J, Rani R, Pritesh V, Ghogale R. Analysis of clinical recovery-period and recovery rate estimation of the first 1000 COVID-19 patients in Singapore. medRxiv 2020.
- [17] Binny RN, Hendy SC, James A, Lustig A, Plank MJ, Steyn N. Effect of Alert Level 4 on effective reproduction number: review of international COVID-19 cases. medRxiv 2020;04:30.20086934 May 2020.
- [18] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, "Mathematical models to characterize early epidemic growth: A review," *Physics of Life Reviews*, vol. 18. Elsevier B.V., pp. 66–97, 01-Sep-2016.
- [19] Mishra PK, Mishra S. A deductive approach to modeling the spread of COVID-19. medRxiv 2020;03:26.20044651 Mar. 2020.
- [20] Dehning J, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science (80-). May 2020:eabb9789.
- [21] Campbell C. How South Korea Is Beating Coronavirus Without a Lockdown. Time 2020.
- [22] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med. 2020:1–4 2020.
- [23] Azevedo CR, et al. Hybrid Semiparametric Modeling: A Modular Process Systems Engineering Approach for the Integration of Available Knowledge Sources. In: Kenett RS, Swarz RS, Zonnenshain A, editors. *Systems Engineering in the Fourth Industrial Revolution*. Wiley; 2019. p. 345–73.