A population-level SEIR model for COVID-19 scenarios (updated)

James P. Gleeson[1,3,4,5], Thomas Brendan Murphy[2,3,5], Joseph D. O'Brien[1], David J. P. O'Sullivan[1]

[1] MACSI, Department of Mathematics and Statistics, University of Limerick, Ireland.
[2] School of Mathematics and Statistics, University College Dublin, Ireland.
[3] Insight Centre for Data Analytics, Ireland
[4] Confirm Centre for Smart Manufacturing, Ireland
[5] Irish Epidemiological Modelling Advisory Group (IEMAG).

12 November 2020

Abstract

This technical note summarizes the current version of the population-level susceptible-exposed-infected-removed (SEIR) model that is used by the Irish Epidemiological Modelling Advisory Group (IEMAG) reporting to the National Public Health Emergency Team (NPHET). The model is continuously updated, so this note supersedes previous versions, and will be superseded by future updates.

Change summary: introduced finite difference formulation (Sec. 1.3); updated calibration algorithm using GAMs (Sec. 2.3) and scenario descriptions and initial conditions (Sec. 3); added source code for implementation in R (Sec. 4); updated references.

1 Description of the model

Population-level SEIR models [1–3] assume fully-mixed, homogeneous populations. Despite this simplification, they provide useful information for scenario-based planning, with the potential for further extension of the structure (e.g., to disaggregated age cohorts) in more advanced models.

At each moment of time, each individual in the population is considered to be in one of a discrete number of compartments. The structure of the compartments, and the timescales for individuals to move in and out of compartments, are based on the current understanding of the epidemiology of COVID-19, as evidenced by the extensive literature review and evidence synthesis conducted by [4–6].

1.1 Compartments and parameters

The structure of the model compartments is shown in Figure 1. Informed by the literature review, each parameter is chosen from a uniform distribution with upper and lower limits as listed below. (All timescales are expressed in days. Day 0 of the Irish epidemic is February 28, 2020.)

- L: average latent period; range assumed is 3.9 to 5.9.
- C: average incubation period; range assumed is $\max(L, 4.8)$ to 6.8 (lower limit ensures $C - L$ is nonnegative).

- D: average infectious period; range assumed is $\max(C - L, 5.0)$ to 9.0 (lower limit ensures $D - (C - L)$ is nonnegative).
- h: multiplicative factor for reduction of effective transmission from the Asymptomatic Infected compartment, relative to Symptomatic Infected; range assumed is 0.01 to 0.5.
- i: multiplicative factor for reduction of effective transmission from the Immediate Isolation compartment, relative to Symptomatic Infected; range assumed is 0 to 0.1.
- j: multiplicative factor for reduction of effective transmission from the Post-test isolation compartment, relative to Symptomatic Infected; range assumed is 0 to 0.1.
- f: fraction of infected who are Asymptomatic; range assumed is 0.18 to 0.82.
- τ: fraction of symptomatic cases that are tested; range assumed is 0.5 to 1.0.
- q: fraction of symptomatic cases that self-quarantine from appearance of symptoms until recovery; range assumed is 0 to $1 - \tau$ (upper limit ensures $1 - q - \tau$ is nonnegative).
- T: average wait for test results; range assumed is 1.0 to $\max(5.0, D - C + L)$ (upper limit ensures $D - C + L - T$ is nonnegative).
- N: population, assumed to be 4.9×10^6.

The transmission rate β is assumed to be time-dependent, in order to model the impact of interventions. The relationship between β and the basic reproduction number R_0 for this model is derived in the Appendix.

1.2 Equations

Each compartment in Figure 1 has a corresponding time-dependent variable, which gives the number of individuals that are in that compartment at time t. The dynamics evolve according to the following differential equations:

$$\frac{dS}{dt} = -\beta S \left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)/N \tag{1}$$

$$\frac{dE}{dt} = \beta S \left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)/N - \frac{1}{L}E \tag{2}$$

$$\frac{dI_p}{dt} = \frac{(1-f)}{L}E - \frac{1}{C-L}I_p \tag{3}$$

$$\frac{dI_a}{dt} = \frac{f}{L}E - \frac{1}{D}I_a \tag{4}$$

$$\frac{dI_i}{dt} = \frac{q}{C-L}I_p - \frac{1}{D-C+L}I_i \tag{5}$$

$$\frac{dI_{t1}}{dt} = \frac{\tau}{C-L}I_p - \frac{1}{T}I_{t1} \tag{6}$$

$$\frac{dI_{t2}}{dt} = \frac{1}{T}I_{t1} - \frac{1}{D-C+L-T}I_{t2} \tag{7}$$

$$\frac{dI_n}{dt} = \frac{(1-q-\tau)}{C-L}I_p - \frac{1}{D-C+L}I_n \tag{8}$$

$$\frac{dR}{dt} = \frac{1}{D}I_a + \frac{1}{D-C+L}I_i + \frac{1}{D-C+L-T}I_{t2} + \frac{1}{D-C+L}I_n, \tag{9}$$

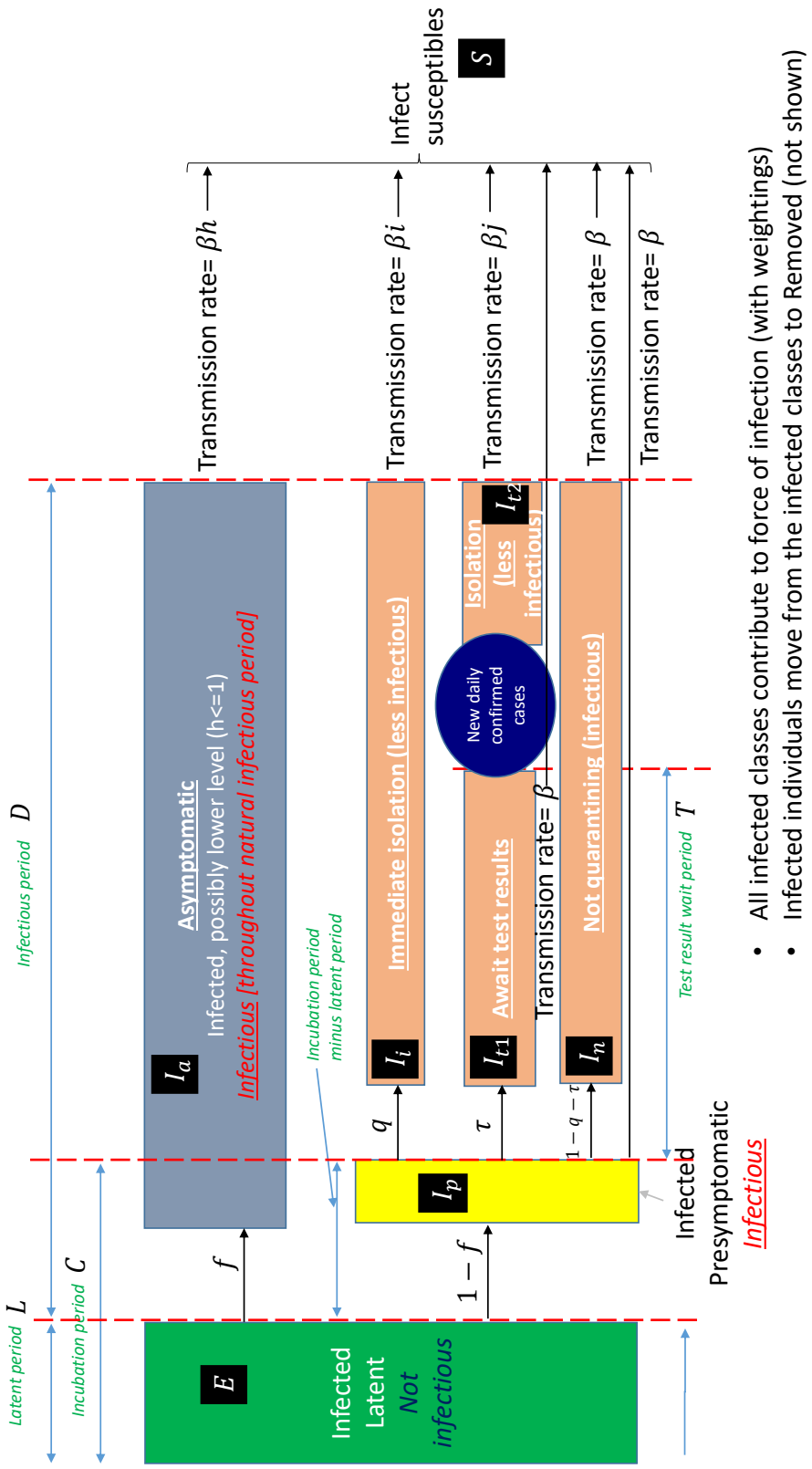

Figure 1: Model compartment structure.

where $S(t)$ is the number of susceptible individuals, $E(t)$ is the number who are exposed, $I_p(t)$ is the number who are pre-symptomatic infected, $I_a(t)$ is the number who are asymptomatic infected, $I_i(t)$ is the number who are symptomatic and self-isolating (without testing), $I_{t1}(t)$ is the number who are symptomatic and waiting for testing, $I_{t2}(t)$ is the number who are in post-test self-isolation, $I_n(t)$ is the number who are symptomatic and not isolating and $R(t)$ is the number who are removed (i.e., recovered from the virus or dead). In addition, we define $C_c(t)$ to be the cumulative number of new cases reported by time t, given by integrating the flux out of the I_{t1} (waiting-for-test) compartment:

$$\frac{dC_c}{dt} = \frac{1}{T} I_{t1} \tag{10}$$

and we also report the number of new daily cases on day n, defined by $c_c(n) = C_c(n) - C_c(n-1)$.

1.3 Finite-difference formulation

The system of differential equations (1)–(10) may be solved numerically using, for example, a simple forward-Euler scheme. This scheme approximates the derivative dx/dt by $(x_{n+1} - x_n)/\Delta$, where Δ is the finite timestep and x_n is the discrete-time approximation to $x(n\,\Delta)$, the value of $x(t)$ at time $t = n\,\Delta$. Expressing the dynamical system (1)–(10) in the general form

$$\frac{dx}{dt} = F\left(x(t)\right), \tag{11}$$

where $x(t)$ represents the vector of all unknowns and using the finite-difference approximation enables the solution x_n to be determined from the initial condition x_0 by iteration:

$$x_{n+1} = x_n + \Delta\, F\left(x_n\right) \qquad \text{for } n = 0, 1, 2, \ldots. \tag{12}$$

To ensure accuracy of this finite-difference formulation, it is important that the timestep Δ chosen be sufficiently small: the convergence of the finite-difference solution to the differential equation solution occurs in the limit $\Delta \to 0$. Testing of the finite-difference results for the system (1)–(10) shows that a value of Δ equal to 0.1 days gives sufficient accuracy.

2 Calibration of the model

2.1 Data

The confirmed positive case data was extracted from the Computerised Infectious Disease Reporting (CIDR) database hosted by the Heath Protection Surveillance Centre (HPSC). For each observation, the event date and lab specimen collected date were stored. The event date of the cases was used to calibrate the model, as outlined in Section 2.3.

2.2 Negative Binomial Generalized Additive Model

The daily case counts $c_c(t)$ at day t, as described in Section 2.1, are modelled by a negative binomial random variable

$$c_c(t) \sim \mathsf{NegB}(g(t), \theta), \tag{13}$$

where $g(t)$ is the expected number of cases on day t and θ is the over dispersion parameter.

To model the mean parameter of the negative binomial distribution, we use a thin-plate regression spline [7],

$$\log g(t) = \beta_0 + \sum_{k=1}^{K} \beta_k B_k(t), \tag{14}$$

where $(\beta_0, \beta_1, \ldots, \beta_K)$ are unknown parameters and $\{B_k(t) : k = 1, 2, \ldots, K\}$ are thin-plate spline basis functions; the value of K is chosen to be large enough to achieve a satisfactory goodness of fit. The resulting model is a negative binomial generalized additive model (GAM) [8, 9].

To account for parameter uncertainty, the model was fitted in a Bayesian framework to the case data using the `brms` R package [10–12]. The `brms` R package interfaces with `Stan` [13] to generate samples from the posterior distribution for the model parameters.

The samples of the model parameters from the posterior distribution can be used to generate posterior realizations of mean curve, $g(t)$, and the cumulative cases curve, $G(t) = \int_0^t g(s)ds$, using (14).

2.3 Calibration algorithm

We calibrate the SEIR model with data from February 29th 2020 to November 11th 2020, using the GAM fit described in Section 2.2. The process we describe here is similar in principle to the method described in [14] (and references therein) for the SIR model, but complicated by the additional compartments of this SEIR model.

For each choice of model parameter values and for each realization $G(t)$ of the data-fitting function, the SEIR equations can be inverted to determine the time-varying transmission rate $\beta(t)$ that would yield precisely $C_c(t) = \mu(t)$. The steps in this inversion algorithm are as follows:

1. Assuming $C_c(t) = G(t)$, solve the algebraic equation (10) for $I_{t1}(t)$.
2. Using $I_{t1}(t)$ from step 1, solve the differential equation (7) for $I_{t2}(t)$.
3. Using $I_{t1}(t)$ from step 1, solve the algebraic equation (6) for $I_p(t)$.
4. Using $I_p(t)$ from step 3, solve the differential equation (5) for $I_i(t)$.
5. Using $I_p(t)$ from step 3, solve the differential equation (8) for $I_n(t)$.
6. Using $I_p(t)$ from step 3, solve the algebraic equation (3) for $E(t)$.
7. Using $E(t)$ from step 6, solve the differential equation (4) for $I_a(t)$.
8. Combine equations (1) and (2) to give

$$\frac{dS}{dt} = -\frac{dE}{dt} - \frac{1}{L}E \tag{15}$$

   and, using $E(t)$ from step 6, solve this differential equation for $S(t)$.

9. Using the results of steps 2, 3, 4, 5, 7 and 8, solve equation (1) for the inferred time-dependent transmission rate $\beta(t)$:

$$\beta(t) = -\frac{N}{S}\frac{dS}{dt}\left(I_p + hI_a + iI_i + I_{t1} + jI_{t2} + I_n\right)^{-1}. \tag{16}$$

These steps are implemented using the finite-difference approximation of Sec. 1.3 for each derivative term. Note that each of the equations (3)–(8) can be expressed in the general form

$$\frac{dC}{dt} = w_{pc} r_p P - r_c C, \tag{17}$$

where $P(t)$ is the number of individuals in the "parent" compartment and $C(t)$ is the number in the "child" compartment. The rates r_p and r_c are associated with the parent and child compartments, respectively, while w_{pc} is a weight that is specific to the parent-child pair. Applying the forward-Euler scheme described in Sec. 1.3 to equation (17) yields the relationship

$$\frac{C_{n+1} - C_n}{\Delta} = w_{pc} r_p P_n - r_c C_n, \tag{18}$$

where C_n and P_n are the finite-difference approximations to $C(n\Delta)$ and $P(n\Delta)$, respectively. If the values of the child variable C_n are known over a time range, we can solve for the parent variable by rearranging equation (18) to obtain

$$P_n = \frac{1}{w_{pc} r_p} \left(\frac{C_{n+1} - C_n}{\Delta} + r_c C_n \right) \text{ for } n = 1, 2, \ldots. \tag{19}$$

Conversely, if the parent variable P_n is known over a time range, along with an initial value $C_0 = C(0)$ of the child variable, then the child variable can be found for all times by iteration:

$$C_{n+1} = C_n + \Delta \left(w_{pc} r_p P_n - r_c C_n \right) \text{ for } n = 0, 1, \ldots. \tag{20}$$

2.4 Model-inferred reproduction number

Using the linear relationship between transmission rate and reproduction number for this model, given by equation (22) in the Appendix, each realization of the inferred $\beta(t)$ from equation (16) can be converted to an equivalent time-varying model-inferred reproduction number R_t. Figure 2 shows the results of this calculation, using 1000 realizations of the GAM fits and associated model parameters, as described in Sec. 3.4 below. This method of estimating a reproduction number is specific to the model, and the estimates should be understood as applying within the context of the SEIR model only. Similar to more standard measures for reproduction number, when the model-inferred R_t exceeds one the number of infections will rise exponentially, while the number of infections decays when R_t is less than one. Note the very high values of this model-inferred R_t in the early stage of the epidemic: this is because the SEIR model is not accurate in the early stages of an epidemic, and the model-inferred R_t values should not be taken as indicative of the true value of the reproduction number in the early stages. However, as the figures in Section 3 demonstrate, the SEIR approach yields good fits to the data and so it can be used for comparison of possible future scenarios.

3 Scenario analysis

For each set of model parameters, the calibration algorithm detailed in Section 2.3 yields a time-varying transmission rate $\beta(t)$ that is consistent with the GAM-fitted data up to the current date. Forward predictions of the model may then be examined under assumptions about how the transmission rate (or, equivalently, the reproduction number) may behave in the future.

To numerically solve the differential equations we require initial conditions that are consistent with the data. The inversion algorithm described in Sec. 2.3 generates the initial values for variables

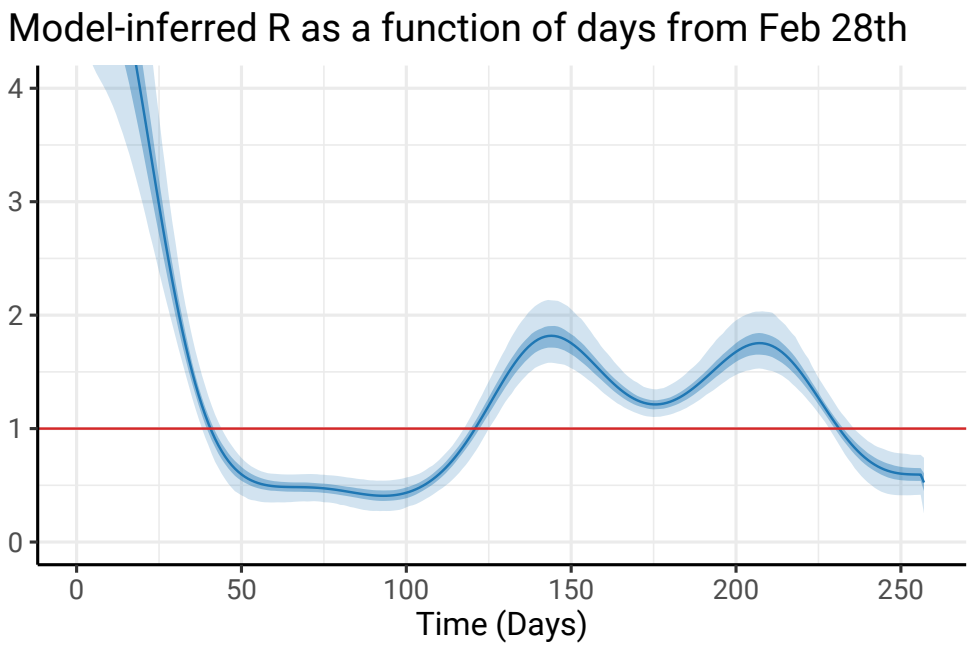

Figure 2: Model-inferred R_t up to November 11th, as described in Section 2.3. The curve indicates the mean (over 1000 realizations as described in Sec. 3.4); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) credible intervals. In this and subsequent figures time is measured in days from February 28th, 2020.

I_{t_1}, I_p, E and S as part of the solution to the algebraic equations in steps 1, 3, 6 and 8, respectively. The remaining variables (I_{t_2}, I_i, I_n and I_a) are each assumed to have an initial value of zero.

Using these initial conditions, the finite-difference approximation of the differential equation system is solved under several assumed scenarios by controlling the time-varying transmission rate $\beta(t)$, as described below.

3.1 Scenario 1: $R = 0.5$ from November 11th

Figure 3 shows results for a scenario where the effective reproduction number R_t is given by the inversion of the model (as in Fig. 2) up to November 11th, with R_t then assumed to be 0.5 thereafter.

3.2 Scenario 2: $R = 0.9$ from November 11th

Scenario 2 assumes that the value of R_t from November 11th is 0.9, giving a slower decrease in cases (see Fig. 4) than Scenario 1.

3.3 Scenario 3: $R = 0.5$ from November 11th to December 2nd, with $R = 1.4$ thereafter

In this scenario, the value of R_t is assumed to be 0.5 from November 11th until December 2nd, with an increase of R_t to 1.4 thereafter, leading to an increase in the number of cases, see Fig. 5.

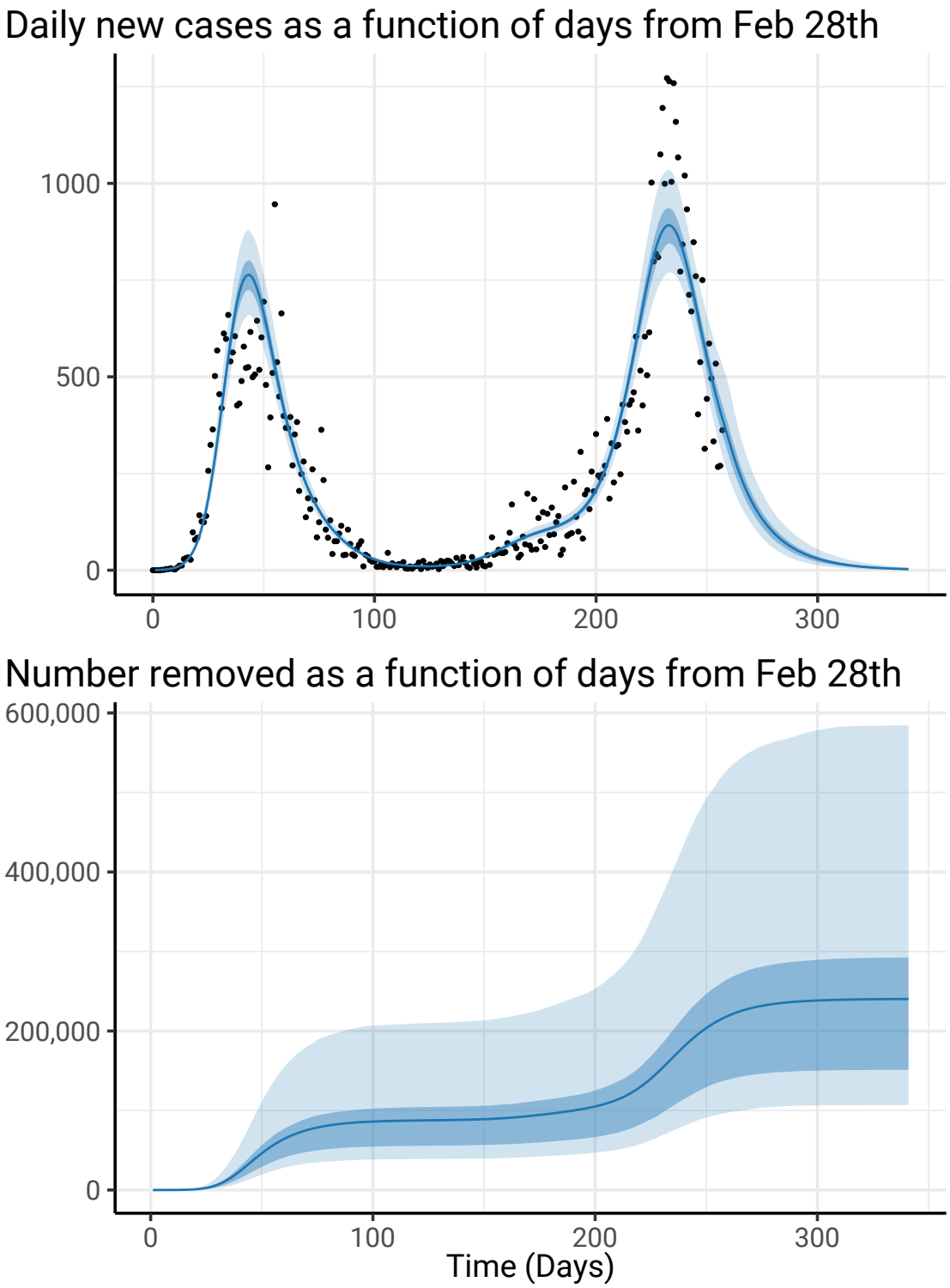

Figure 3: Scenario 1 assumes $R = 0.5$ from November 11th. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 1000 realizations as described in Sec. 3.4); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) credible intervals.

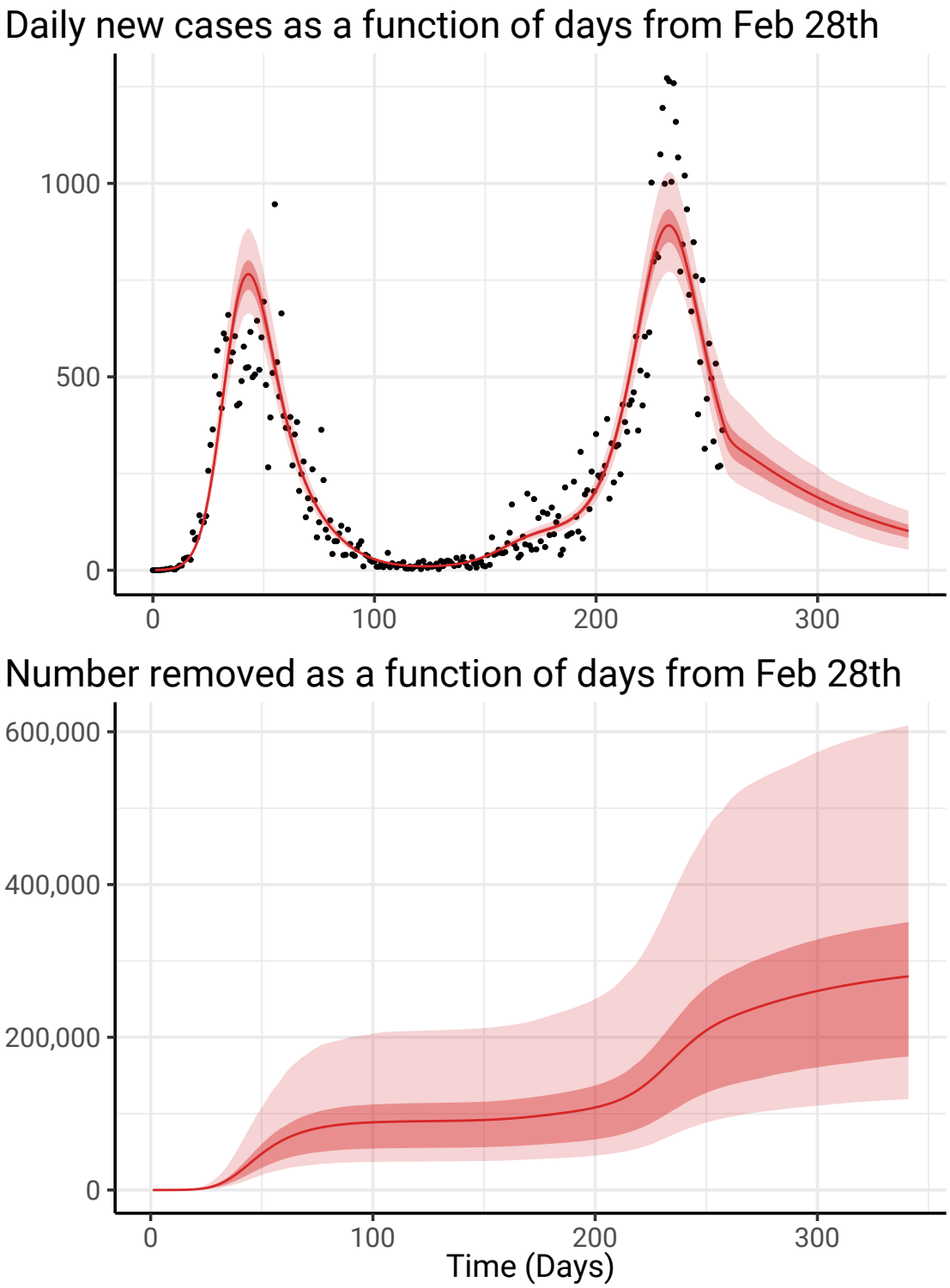

Figure 4: Scenario 2 assumes $R = 0.9$ from November 11th. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 1000 realizations as described in Sec. 3.4); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) credible intervals.

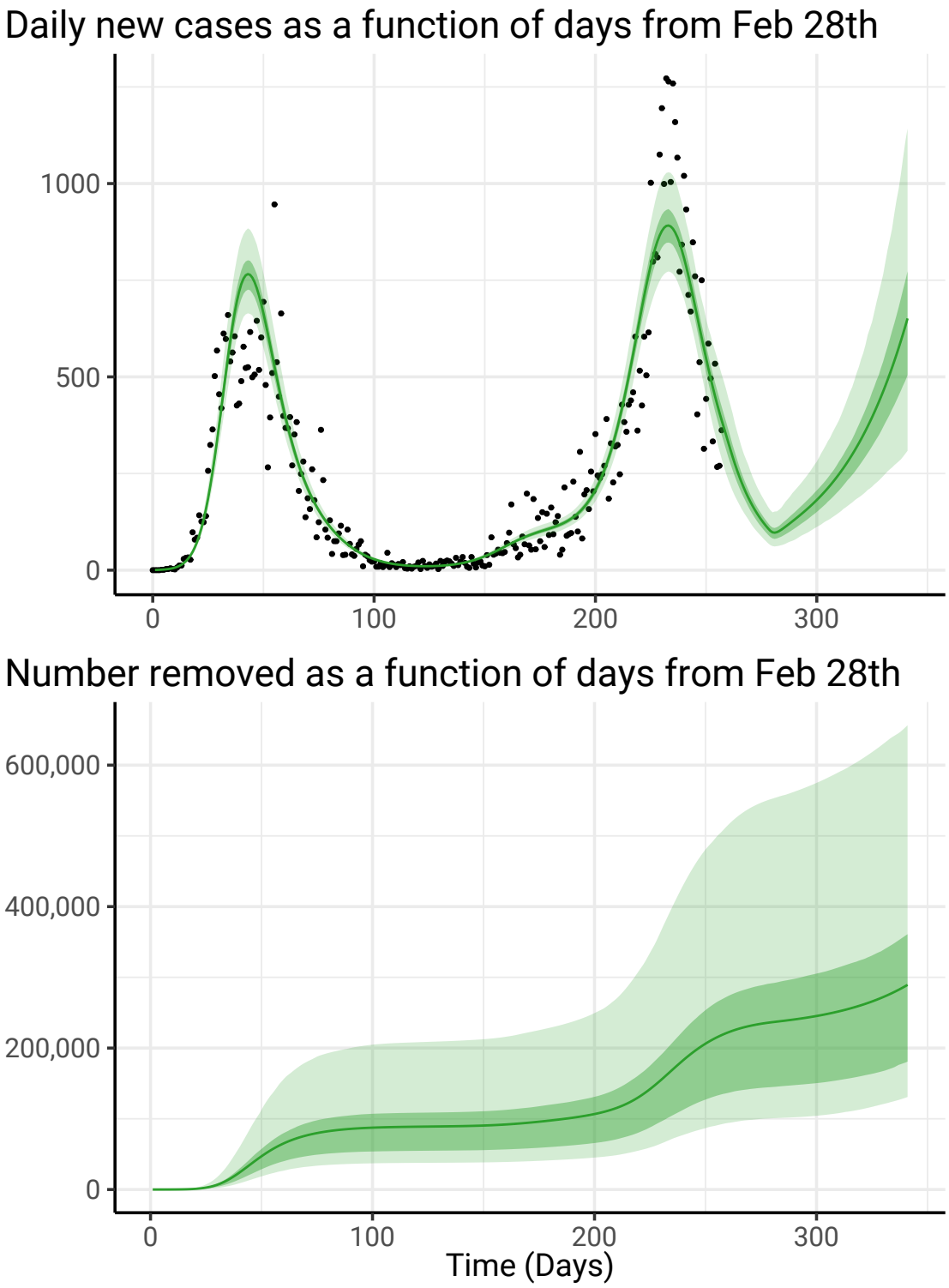

Figure 5: Scenario 3 assumes $R = 0.5$ from November 11th until December 2nd, with $R = 1.4$ thereafter. Top panel shows confirmed new cases per day; bottom panel is the number of removed individuals (this includes both deaths and those who recovered from the virus). The curve indicates the mean (over 1000 realizations as described in Sec. 3.4); shaded regions show the 50% (quantiles 0.25 to 0.75) and 95% (quantiles 0.025 to 0.975) credible intervals.

3.4 Uncertainty quantification

The uncertainty shown with the quantiles in Figures 3 through 5 is due to the ranges of parameter values as specified in Section 1.1 and to the uncertainty in the GAM fits. A sample of 1000 posterior realizations of the curves are generated from the GAM fitting procedure (Sec. 2.2), and for each realization, draws from each distribution of the set of model parameters (Sec. 1.1) are used to generate scenario outputs. However, it is important to note that in each case it is assumed that the reproduction number is exactly as specified in the scenario description. Comparing the results for the three scenarios, it is clear that the effect of uncertainty in future R values dominates the uncertainty due to the range of possible values for the other parameters of the model.

4 Source code

R code that implements the calibration algorithm described in Sec. 2.3 and runs an example scenario is available for download from `https://github.com/obrienjoey/ireland_covid_modelling`. The differential equations are approximated by finite-difference equations using the forward Euler method described in Sec. 1.3, with timestep of 0.1 day; this timestep was found to be sufficiently small to give high accuracy.

Acknowledgements

The model was constructed with the advice of members of the Irish Epidemiological Modelling Advisory Group. This work includes input from researchers funded by Science Foundation Ireland (grant numbers SFI/16/IA/4470, SFI/18/CRT/6049, SFI/12/RC/2289_P2, SFI/16/RC/3918, SFI/16/RC/3835, SFI/12/RC/2275_P2, SFI/15/IA/3090).

Appendix

To calculate the basic reproduction number for the model we follow the approach explained in Section 2.2 of [15]. The value of R_0 is given by the spectral radius of the next generation matrix, which can be written as FV^{-1}. Defining the vector of relevant dynamical variables as $\mathbf{x} = \{E, I_p, I_a, I_i, I_{t1}, I_{t2}, I_n\}$, the matrix F is zero except for its first row, which is $(0, \beta, h\beta, i\beta, \beta, j\beta, \beta)$. The corresponding matrix V is

$$\begin{pmatrix} \frac{1}{L} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{(1-f)}{L} & \frac{1}{C-L} & 0 & 0 & 0 & 0 & 0 \\ -\frac{f}{L} & 0 & \frac{1}{D} & 0 & 0 & 0 & 0 \\ 0 & -\frac{q}{C-L} & 0 & \frac{1}{D-C+L} & 0 & 0 & 0 \\ 0 & -\frac{\tau}{C-L} & 0 & 0 & \frac{1}{T} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{T} & \frac{1}{D-C+L-T} & 0 \\ 0 & -\frac{(1-q-\tau)}{C-L} & 0 & 0 & 0 & 0 & \frac{1}{D-C+L} \end{pmatrix}. \tag{21}$$

Calculating the eigenvalues of FV^{-1} then yields the relationship

$$\frac{R_0}{\beta} = (f-1)((i-1)q(C-L) + (j-1)\tau(C-L+T)) + \\ + D(f(h-iq-j\tau+q+\tau-1) + (i-1)q + (j-1)\tau + 1). \quad (22)$$

It can prove useful to also consider this formula is terms of contributions from each of the infected compartments. With a contribution to R_0 defined as the fraction of infectives moving through the compartment multiplied by the average time spend in the compartment multiplied by the transmission rate for that compartment, the respective contribution from each compartment is as follows:

- Contribution from I_a is $fDh\beta$.
- Contribution from I_p is $(1-f)(C-L)\beta$.
- Contribution from I_i is $(1-f)q(D-C+L)i\beta$.
- Contribution from I_{t1} is $(1-f)\tau T\beta$.
- Contribution from I_{t2} is $(1-f)\tau(D-C+L-T)j\beta$.
- Contribution from I_n is $(1-f)(1-q-\tau)(D-C+L)\beta$.

Summing these contributions gives equation (22).

References

[1] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[2] Emilia Vynnycky and Richard White. *An introduction to infectious disease modelling*. Oxford University Press, 2010.

[3] Mason A Porter and James P Gleeson. Dynamical systems on networks. *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, 4, 2016.

[4] John Griffin, Miriam Casey, Áine Collins, Kevin Hunt, David McEvoy, Andrew Byrne, Conor McAloon, Ann Barber, Elizabeth Ann Lane, and Simon More. Rapid review of available evidence on the serial interval and generation time of COVID-19. *BMJ Open*, 10(11):e040263, 2020.

[5] Conor McAloon, Áine Collins, Kevin Hunt, Ann Barber, Andrew W Byrne, Francis Butler, Miriam Casey, John Griffin, Elizabeth Lane, David McEvoy, Patrick Wall, Martin Green, Luke O'Grady, and Simon J More. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10(8):e039652, 2020.

[6] Andrew William Byrne, David McEvoy, Aine B Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth A Lane, Conor McAloon, Kirsty O'Brien, Patrick Wall, Kieran A Walsh, and Simon J More. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *BMJ Open*, 10(8):e039856, 2020.

[7] Simon N. Wood. Thin-plate regression splines. *Journal of the Royal Statistical Society Series B*, 65(1):95–114, 2003.

[8] Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B*, 73(1):3–36, 2011.

[9] Simon N. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2nd edition, 2017.

[10] Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1):1–28, 2017.

[11] Paul-Christian Bürkner. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411, 2018.

[12] Paul-Christian Bürkner, Jonah Gabry, and Sebastian Weber. brms: Bayesian regression models using 'Stan', 2020. R package: Version 2.14.0.

[13] Stan Development Team. Stan modeling language users guide and reference manual, 2020. Version 2.21.2. https://mc-stan.org.

[14] Anna Mummert. Studying the recovery procedure for the time-dependent transmission rate(s) in epidemic models. *Journal of Mathematical Biology*, 67(3):483–507, 2013.

[15] Jane M Heffernan, Robert J Smith, and Lindi M Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.