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.
Uncertainty quantification in Covid-19 spread: Lockdown effects

Ana Carpio a, b, Emile Pierret b

a Departamento de Matemática Aplicada, Universidad Complutense, 28040 Madrid, Spain
b CMLA, ENS Paris-Saclay, 91190 Gif-sur-Yvette, France

ARTICLE INFO

Keywords:
SEIJR models
Covid-19
Numerical simulation
Bayesian inference
Uncertainty quantification

ABSTRACT

We develop a Bayesian inference framework to quantify uncertainties in epidemiological models. We use SEIJR and SJIR models involving populations of susceptible, exposed, infective, diagnosed, dead and recovered individuals to infer from Covid-19 data rate constants, as well as their variations in response to lockdown measures. To account for confinement, we distinguish two susceptible populations at different risk: confined and unconfined. We show that transmission and recovery rates within them vary in response to facts, and that the diagnose rate is quite low, which leads to large amounts of undiagnosed infective individuals. A key unknown to predict the evolution of the epidemic is the fraction of the population affected by the virus, including asymptomatic subjects. Our study tracks its time evolution with quantified uncertainty from available official data, limited, however, by the data quality. We exemplify the technique with data from Spain, country in which late drastic lockdowns were enforced for months during the first wave of the current pandemic. In late actions and in the absence of other measures, spread is delayed but not stopped unless a large enough fraction of the population is confined until the asymptomatic population is depleted. To some extent, confinement can be replaced by strong distancing through masks in adequate circumstances.

Introduction

Since the outbreak of the current Covid-19 pandemic [1,2], Health Services worldwide report daily data about the status of the epidemic, which serve as a guide for the design of non-pharmaceutical interventions [3,4]. An increasing number of mathematical studies assess the efficacy of different policies [4–9]. Moreover, mathematical models and data analysis are employed to estimate relevant epidemiological parameters [4,10–14] and to try to forecast the evolution [15–21]. While some of this research is based on direct data analysis [4,13], machine learning techniques [15,21] or empirical laws for different populations [9], the use of balance equations to predict population dynamics is a common approach.

After the pioneering work of Kermack and McKendrick [22], SIR type models have become a standard tool in epidemiological studies [23]. The specific structure of the selected models depends on the available information and on assumptions about the epidemic spread [24]. Basic SIR models involve populations of susceptible S, infected I, and recovered R individuals, expecting immunity of the latter [8,14,17,25]. SEIR variants distinguish also the individuals exposed to the virus E, which may become infective [11,19]. Immunity of the recovered is suppressed in SEIRS systems [18,20]. During the 2002–04 SARS (Severe Acute Respiratory Syndrome) outbreak, these models were adapted to describe the SARS epidemic in different countries by singling out the diagnosed infective D [10,26], becoming SELIR or SJIR models. Diagnosed individuals are isolated. The virus SARS-CoV-2 responsible for the illness Covid-19 belongs to the same family as the virus SARS-CoV, responsible for SARS. The epidemics triggered by them share some features, such as the role of asymptomatic individuals in superspread events, see [12] for a quantification of the fraction of asymptomatic population during Covid-19 spread following this approach. Here, we will study the effect of confinement measures on Covid-19 spread by distinguishing two susceptible SEIJR populations: confined and non confined.

To have a predictive value, we must fit the model parameters to available data. This can be done applying optimization or adjoint-based data assimilation techniques to reduce the difference between recorded data and model predictions for selected parameters [10,14], for instance. However, data for epidemiological studies are subject to many sources of noise and uncertainty. In the case of the current Covid-19 pandemic, different countries, and regions within them, define the diagnosed, recovered and dead individuals they count in their official reports in different ways. The number of dead individuals may refer only to patients who die in hospitals or include also deaths at homes and care homes. Furthermore, the death of covid patients with previous health issues may be officially attributed to other causes. On the other
hand, the number of diagnosed individuals may refer only to cases confirmed by a PCR (Polymerase chain reaction) test or include also positive antibody tests, or probable cases with compatible symptoms and clinical history. Moreover, the results of tests may arrive with a variable delay, which results in fluctuations and exclusions. Undated cases may not be counted at all. Tests repeated for the same individuals may be counted as different. Additionally, the number of tests performed varies largely over the weeks due to supply shortages and changes in local testing policies, and the accuracy of the tests employed may fluctuate, yielding false negatives or positives.

Uncertainty in the data propagates to any predictions based on them. Instead of fixing specific guesses for the model coefficients, it is convenient to explore approaches that quantify uncertainty [7–9,12,27]. Unlike most work which does not distinguish undocumented and documented infected individuals, here we follow the SEIJR approach and compare data to model predictions of diagnosed infected \( J \) may fluctuate, yielding false negatives or positives. Changes in local testing policies, and the accuracy of the tests employed cases may not be counted at all. Tests repeated for the same individual may fluctuate, yielding false negatives or positives.

The uncertainty in the data propagates to any predictions based on them. Instead of fixing specific guesses for the model coefficients, it is convenient to explore approaches that quantify uncertainty [7–9,12,27]. Unlike most work which does not distinguish undocumented and documented infected individuals, here we follow the SEIJR approach and compare data to model predictions of diagnosed infected \( J \) models for SARS and Covid-19 type epidemics

SEIJR models involving populations of susceptible (S), exposed (E), infective (I), diagnosed (J), and recovered (R) individuals were proposed in [26] to study the spread of the 2002–04 SARS outbreak. Here, we will adapt them to describe contention measures for Covid-19. Considering two populations \( S_1 \) and \( S_2 \) of different susceptibility, the model takes the form:

\[
\begin{align*}
\frac{dS_1}{dt} &= -\beta S_1(t) \left( I(t) + q E(t) + \epsilon J(t) \right), \\
\frac{dS_2}{dt} &= -\beta p S_2(t) \left( I(t) + q E(t) + \epsilon J(t) \right), \\
\frac{dE}{dt} &= \beta (S_1(t) + p S_2(t)) \left( I(t) + q E(t) + \epsilon J(t) \right) \frac{N}{N} - k E(t), \\
\frac{dI}{dt} &= k E(t) - (\alpha + \gamma_1 + \delta) I(t), \\
\frac{dJ}{dt} &= \alpha I(t) - (\gamma_2 + \delta) J(t), \\
\frac{dR}{dt} &= \gamma_1 I(t) + \gamma_2 J(t), \\
\end{align*}
\]

where \( N = S_1 + S_2 + E + I + J + R + D \) is the total population number, which remains constant. \( D \) is the number of dead individuals. The exposed \( E \) are a class of asymptomatic and possibly infectious individuals. The possibility of transmission from exposed individuals \( E \) is represented by the parameter \( \gamma \). They may progress to the infective state \( I \) at a rate \( \gamma_1 \). The class \( I \) is composed of symptomatic, infectious, and undiagnosed individuals. Infectious individuals \( I \) become diagnosed \( J \) at a rate \( \gamma_2 \). The recovery rate of the infective \( I \) is \( \gamma_1 \), whereas the recovery rate of the diagnosed \( J \) is \( \gamma_2 \). The recovered individuals \( R \) keep track of the cumulative number of sick individuals who become healthy again. Diagnosed individuals \( J \) are isolated from the rest. Their reduced impact on transmission is represented through a parameter \( \epsilon \). Mortality of infected \( I \) and diagnosed \( J \) individuals caused by the virus is denoted by \( \delta \). Finally, \( \beta \) represents the transmission rate: how susceptible \( S \) individuals become virus spreaders. Time is measured in days.

The model has to be complemented with initial conditions. This fact introduces an additional parameter \( t_{in} \) to locate the time at which local spread started [10]. Other approaches assume the initial data unknown instead [8], in our case that choice would increase considerably the number of unknowns. Furthermore, we consider that the risk of infection for \( S_1 \) is lower than the risk for \( S_2 \) by a factor \( \rho \). The total population is initially partitioned as \( S_1 = (1 - \rho)S \) and \( S_2 = \rho S \), \( \rho \) being the fraction of the susceptible population \( S \) at a lower risk of infection. Risk might vary due to specific characteristics of the population (age, sex, genes) [26]. Here, variations will be due to confinement/protection measures enforced on part of the population.

Two constraints are usually imposed on the parameters: (1) \( a > \gamma_1 \) and (2) \( \gamma_2 \cdot \gamma_1 = a \cdot (1 - a) \) [26]. Moreover, the following expression for the basic reproduction number [26] holds

\[ R_0 = \beta(\rho + p(1 - \rho)) \left( \frac{q + 1}{a + \gamma_1 + \delta} + \frac{a \epsilon}{(a + \gamma_1 + \delta)(\gamma_2 + \delta)} \right) \]

The reproduction number represents the expected number of cases immediately originated by one case in a population where all individuals are susceptible to infection, that is, no other individuals are infected or immunized (naturally or through vaccination). Instead, the effective reproduction number \( R_e \) is just the number of cases produced in the current state of a population.
results in physics 35 (2022) 105375
3

\( S \) \( (7) \) the susceptibility degree is not distinguished.
\( E \) neglected. Assuming further that:

(a) SEIJR models we have introduced assume that (1) spread takes place in a closed system, (2) the death rate is the same for everybody served in SARS epidemics, such as the emergence of symptomatic and infected cases. In practice, only the diagnosed recovered diagnosed people who recover or die, thus

(7) \( S \) \( (7) \) the susceptibility degree is not distinguished.
\( E \) neglected. Assuming further that:

(b) SEIJR based Bayesian inference and predictions for the total number of diagnosed individuals using counts from Period 1 (red), Periods 1–2 (green), Periods 1–2–3 (blue) and Periods 1–2–3–4 (magenta). For each of them, top colored triangles separate the inference from the prediction part of the simulations. True data are marked by yellow circles. Solid curves correspond to best fits, dashed curves and dotted curves to different types of sample averages. Shaded areas and dotted curves define uncertainty regions, see Section “Incorporating the effect of contention measures” for a discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

| Par. | Definition | Guess |
|------|------------|-------|
| \( \beta \) | Transmission rate per day | 1/5–1/6 (stats) |
| \( \alpha \) | Rate of progression from infective to diagnosed per day | 1/10–1/11 (stats) |
| \( \gamma_1 \) | Recovery rate from infective to diagnosed | 1/14 (practice) |
| \( \delta \) | Covid-19 induced mortality per day | 1/10–1/11 (stats) |
| \( \ell \) | Relative measure of isolation of diagnosed cases | |
| \( \rho \) | Reduction in risk of Covid-19 infection for class \( S \) | |
| \( t_{in} \) | Time at which local spread starts | |
| \( p \) | Fraction of the population at a lower risk | |

This type of models reproduces crucially some characteristics observed in SARS epidemics, such as the emergence of symptomatic and asymptomatic individuals, superspread events and unequal susceptibility, for instance. We will use them here with data from the current Covid-19 epidemic. First guesses for some of the model parameters can be estimated from average observations, see Table 1. First guesses for two key parameters, \( t_{in} \) and \( \beta \) can be obtained from simplified SEIJR approximations, as we explain in the next section.

Fitting the initial stages of the outbreak

The SEIJR models we have introduced assume that (1) spread takes place in a closed system, (2) the death rate is the same for everybody served in SARS epidemics, such as the emergence of symptomatic and infected cases. In practice, only the diagnosed recovered diagnosed people who recover or die, thus

(a) Daily counts of diagnosed, recovered and dead individuals (PCR confirmed) in Spain since February 25th, 2020, until May 22th, 2020 [28]. After an initial period of uncontrolled spread (Period 1), borders were closed, while all the population being able to work online, or not working in basic activities, was confined at home in the whole country (Period 2): education, administration, tourism, shopping, leisure activities... Lockdown was later extended to all non essential activities (Period 3). Only food and medical supplies, healthcare, security, essential transport and essential production remained exposed. Confinement was then released by stages, first some workers (Period 4), then the rest, while introducing recommendations for the use of masks and social distancing.

(b) SEIJR based Bayesian inference and predictions for the total number of diagnosed individuals using counts from Period 1 (red), Periods 1–2 (green), Periods 1–2–3 (blue) and Periods 1–2–3–4 (magenta). For each of them, top colored triangles separate the inference from the prediction part of the simulations. True data are marked by yellow circles. Solid curves correspond to best fits, dashed curves and dotted curves to different types of sample averages. Shaded areas and dotted curves define uncertainty regions, see Section “Incorporating the effect of contention measures” for a discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

| Par. | Definition | Guess |
|------|------------|-------|
| \( \beta \) | Transmission rate per day | 1/5–1/6 (stats) |
| \( \alpha \) | Rate of progression from infective to diagnosed per day | 1/10–1/11 (stats) |
| \( \gamma_1 \) | Recovery rate from infective to diagnosed | 1/14 (practice) |
| \( \delta \) | Covid-19 induced mortality per day | 1/10–1/11 (stats) |
| \( \ell \) | Relative measure of isolation of diagnosed cases | |
| \( \rho \) | Reduction in risk of Covid-19 infection for class \( S \) | |
| \( t_{in} \) | Time at which local spread starts | |
| \( p \) | Fraction of the population at a lower risk | |

This type of models reproduces crucially some characteristics observed in SARS epidemics, such as the emergence of symptomatic and asymptomatic individuals, superspread events and unequal susceptibility, for instance. We will use them here with data from the current Covid-19 epidemic. First guesses for some of the model parameters can be estimated from average observations, see Table 1. First guesses for two key parameters, \( t_{in} \) and \( \beta \) can be obtained from simplified SEIJR approximations, as we explain in the next section.

Fitting the initial stages of the outbreak

The SEIJR models we have introduced assume that (1) spread takes place in a closed system, (2) the death rate is the same for everybody served in SARS epidemics, such as the emergence of symptomatic and infected cases. In practice, only the diagnosed recovered diagnosed people who recover or die, thus

(a) SEIJR models we have introduced assume that (1) spread takes place in a closed system, (2) the death rate is the same for everybody served in SARS epidemics, such as the emergence of symptomatic and infected cases. In practice, only the diagnosed recovered diagnosed people who recover or die, thus

(b) SEIJR based Bayesian inference and predictions for the total number of diagnosed individuals using counts from Period 1 (red), Periods 1–2 (green), Periods 1–2–3 (blue) and Periods 1–2–3–4 (magenta). For each of them, top colored triangles separate the inference from the prediction part of the simulations. True data are marked by yellow circles. Solid curves correspond to best fits, dashed curves and dotted curves to different types of sample averages. Shaded areas and dotted curves define uncertainty regions, see Section “Incorporating the effect of contention measures” for a discussion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

| Par. | Definition | Guess |
|------|------------|-------|
| \( \beta \) | Transmission rate per day | 1/5–1/6 (stats) |
| \( \alpha \) | Rate of progression from infective to diagnosed per day | 1/10–1/11 (stats) |
| \( \gamma_1 \) | Recovery rate from infective to diagnosed | 1/14 (practice) |
| \( \delta \) | Covid-19 induced mortality per day | 1/10–1/11 (stats) |
| \( \ell \) | Relative measure of isolation of diagnosed cases | |
| \( \rho \) | Reduction in risk of Covid-19 infection for class \( S \) | |
| \( t_{in} \) | Time at which local spread starts | |
| \( p \) | Fraction of the population at a lower risk | |

This type of models reproduces crucially some characteristics observed in SARS epidemics, such as the emergence of symptomatic and asymptomatic individuals, superspread events and unequal susceptibility, for instance. We will use them here with data from the current Covid-19 epidemic. First guesses for some of the model parameters can be estimated from average observations, see Table 1. First guesses for two key parameters, \( t_{in} \) and \( \beta \) can be obtained from simplified SEIJR approximations, as we explain in the next section.

Fitting the initial stages of the outbreak

The SEIJR models we have introduced assume that (1) spread takes place in a closed system, (2) the death rate is the same for everybody served in SARS epidemics, such as the emergence of symptomatic and infected cases. In practice, only the diagnosed recovered diagnosed people who recover or die, thus

\[ \frac{dS}{dt} = -\beta(I + \ell J), \]  
\[ \frac{dI}{dt} = (\beta - (\alpha + \gamma_1 + \delta)) I + \ell \beta J, \]  
\[ \frac{dJ}{dt} = \alpha I - (\gamma_2 + \delta) J, \]  
\[ \frac{dR}{dt} = (\gamma_1 + \gamma_2) J. \]  

This can be done starting from educated guesses and optimizing a cost functional with respect to them. The clinical information collected during the current pandemic [29] yields tentative average values for the rates \( \alpha, \gamma_1, \gamma_2, \delta, \) and for \( \ell, \) collected in Table 1. We then seek to fit the remaining parameters by optimizing a cost. A popular choice is

\[ f(\beta, t_{in}) = \frac{1}{2} \sum_{j=1}^{L} (J(\beta, j + t_{in}) - \tilde{y}_j)^2, \]  

where \( \tilde{y}_j, j = 1, \ldots, L, \) are cumulative numbers of diagnosed people for \( L \) days and the cumulative variable \( J \) solves \( J = aI, J(0) = 0, \) with \( I \) given by (3). This variable \( J \) is in fact the total cumulative number of diagnosed individuals, obtained adding to \( J \) the diagnosed recovered \( R_j \) and the diagnosed dead \( D_j, \) solutions of

\[ R'_j = \gamma_2 J, \quad D'_j = \delta J, \quad R_j(t_{in}) = D_j(t_{in}) = 0. \]  

This is an important distinction. Note that Eq. (4) discounts the diagnosed people who recover or die, thus \( J \) tracks only the active diagnosed cases. In practice, only the diagnosed recovered \( R_j, \) the diagnosed dead \( D_j, \) and the diagnosed active \( J \) or total \( J \) are recorded by Health Care Systems, since the contribution coming from undiagnosed infected cases is unknown.

SEIJR models are adequate for these fittings because solutions admit explicit expressions which reduce numerical errors when dealing with exponentially growing solutions, see Appendix “Solutions of the SEIJR
model”. We will resort to the Levenberg–Marquardt–Fletcher algorithm [30] to optimize the costs. The final values we obtain for \( t_0 \) and \( \beta \) are 12 days and 0.6262, starting the optimization from initial guesses 10 and 0.6. This is consistent with the fact that deaths occurred as early as February 13 in Spain were proven to be caused by Covid-19. Notice that we are fitting a cumulative magnitude \( J = J + R_i + D_j \). Even if the fitting for \( \bar{J} \) is accurate, as Fig. 2 shows, the results worsen noticeably when we use these parameters to calculate \( J, R_i, D_j \) and compare with the data recorded for each of them.

We could improve the overall guess using these values as starting point for an algorithm optimizing the cost

\[
f(\alpha, \gamma, \ell, \epsilon, t_{\text{in}}) = \frac{1}{2} \sum_{j=1}^{L} (J(\alpha, \gamma, \ell, \epsilon, j + t_{\text{in}}) - \bar{J})^2, \tag{11}
\]

with respect to all of the parameters, or resorting to more detailed cost functionals. However, our goal here is to quantify uncertainty in usual rough fits and predictions obtained with them. Therefore, we will use them as priors for the subsequent Bayesian studies.

Uncertainty quantification by Bayesian techniques

Bayes’ theorem describes the probability of an event, based on prior knowledge about it [31]. According to it, the posterior probability of observing a finite number of parameters \( v \) given data \( d \) would be

\[
\rho(v|d) = \frac{\rho(d|v)\rho(v)}{\rho(d)}
\]

where \( \rho(d|v) \) is a conditional probability (the likelihood of observing data \( d \) given parameters \( v \) ), and \( \rho(v) \) represents our prior knowledge on the parameters \( v \). The normalization factor \( \rho(d) \) represents the probability of the data. It is also a marginal probability, which can be obtained integrating \( \rho(d|v)\rho(v) \) with respect to \( v \).

Let us fit our problem in this framework. The parameters are the model parameters, that is,

\[
v = (t_0, \beta, \gamma, \ell, \epsilon, \alpha), \quad \gamma_i^{-1} = \gamma_0^{-1} + a_i^{-1}, \tag{12}
\]

for the SJR model or, for SEIJR,

\[
v = (t_0, \beta, \gamma, \ell, \epsilon, q, p, k), \quad \gamma_i^{-1} = \gamma_0^{-1} + a_i^{-1}. \tag{13}
\]

Then, the prior distribution, the likelihood and the posterior distribution are defined as follows.

Prior distribution

For the prior distribution, we use a parameter guess \( v_0 \) as the mean of a multivariate normal distribution with a covariance matrix \( G_v \) constructed from the deviations of each variable

\[
\rho(v) = \frac{1}{(2\pi)^{n/2}|G_v|^{1/2}} \exp\left( -\frac{1}{2} (v - v_0)^T G_v^{-1} (v - v_0) \right), \tag{14}
\]

where \( n \) is the number of parameters. We choose a diagonal covariance matrix \( G_v \) with elements \( a_i^2, i = 1, \ldots, n \). In practice, we have to modify this proposal because our parameters are always positive and gaussians may produce negative values. Thus, we set

\[
p_{\rho}(v) = \left\{ \begin{array}{ll}
\exp\left( -\frac{1}{2} (v - v_0)^T G_{v_i}^{-1} (v - v_0) \right), & v_j \geq 0, j = 1, \ldots, n, \\
0, & v_j < 0, \text{ for some } j.
\end{array} \right. \tag{15}
\]

This will be our choice of prior distribution \( p_{\rho}(v) \). We do not need to calculate the normalization factor for later use, since our sampling techniques do not require it.

Likelihood

For the conditional probability density \( \rho(d|v) \) we set

\[
\rho(d|v) = \frac{1}{(2\pi)^{n/2}|G_v|^{1/2}} \exp\left( -\frac{1}{2} ||f(v) - d||^2_{G_v^{-1}} \right), \tag{16}
\]

where \( ||v||^2_{G_v} = \bar{v}^T G_v^{-1} \bar{v}, G_v \) being the covariance matrix representing the noise in the data \( d, \) and \( f(v) \) the observation operator. We assume additive Gaussian noise, i.e., the observations and true parameters would be related by

\[
d = f(v_{\text{true}}) + \epsilon. \tag{17}
\]

Here, the noise \( \epsilon \) is distributed as a multivariate Gaussian \( \mathcal{N}(0, G_\epsilon) \) with mean zero and covariance matrix \( G_\epsilon \).

In practice, the data available are daily cumulative counts of diagnosed individuals \( j_{\alpha} \), diagnosed recovered \( r_m \), and diagnosed dead \( d_m, \) \( m = 1, \ldots, M, \) see [28]. Putting the three blocks of data together we have

\[
d = (J_1, \ldots, J_M, R_1, \ldots, R_M, d_1, \ldots, d_M), \tag{18}
\]

where the dynamics of the diagnosed recovered \( R_j \) and diagnosed dead \( D_j \) are governed by (10) whereas the diagnosed individuals \( J \) in which the infection is active are governed by (4) for SJR (see Appendix “Solutions of the SJR model” for analytic expressions) or (1) for SEIJR. In (16), we compare these observations to the data \( d \) using the distance \( \frac{1}{2} ||f(v) - d||^2_{G_v^{-1}}. \) To simplify, we consider the noise level for all observations to be uncorrelated, so that \( G_v \) is a real diagonal matrix, \( \epsilon_j = \text{diag}(\sigma_1^2, \ldots, \sigma_j^2), \) and set all the variances for the same magnitude equal to a constant \( \sigma^2, \sigma_0^2, \sigma_j^2. \) Thus, \( \sqrt{G_v} = \sigma_j \epsilon_j G_v^{-1}, \) where \( L = 3M \) is the number of data considered. Note that these cost functionals require more information than those based on total case counts: we distinguish diagnosed individuals who are dead, recovered and still sick, and compare with model predictions for them discarding the contribution of the undiagnosed, unlike [8,10].

Posterior distribution

Combining (15) with (16) and neglecting normalization constants, the posterior density becomes, up to multiplicative constants,

\[
p_{\rho}(v) \sim \exp\left( -\frac{1}{2} ||f(v) - d||^2_{G_v^{-1}} - \frac{1}{2} ||v - v_0||^2_{G_v^{-1}} \right). \tag{20}
\]

By sampling this posterior distribution, we can visualize the uncertainty in the inference of parameters for a given data set. To do so, we will resort to Markov Chain Monte Carlo Sampling [33,34]. Once we have a large collection of samples, we can extract information from the model (2)–(7) with quantified uncertainty, such as the global number of people who have been affected by the virus the last day of the period we are considering. In the next sections we exemplify the procedure for the different stages of the epidemic as observed in Fig. 1.
On the other hand, \( \sigma \) is the statistical meaning, it keeps track of a possible best fit to the data. \( \sigma = A \) as the final number of affected people in Fig. 3(e). It has been calculated from the posterior distribution (20) by MCMC techniques [34]. Sampling is done with deviations \( \sigma = 10 \) days. We usually set \( \sigma = 10 \) days. We then sample \( S \) parameter \( W \) initialized with \( \sigma \) parameter \( W \). Discarding the first \( S \) samples to draw histograms representing the marginal probabilities of each parameter \( \sigma \). Otherwise, the initial stage of parameter \( \sigma \) will not be large. When the distributions under study are symmetric, it will be close to \( \sigma \). Otherwise, it may depart from it. In our case, slight asymmetry is caused by discarding negative values. In principle, we could try to improve our estimate of the parameter values that maximize the likelihood by optimization procedures [33]. In practice, enforcing the positivity constraint while doing it may be problematic, and the best samples provide reasonable approximations for our purposes.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

The number of people affected by the virus with a SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

When the distributions under study are symmetric, it will be close to \( \sigma \). Otherwise, it may depart from it. In our case, slight asymmetry is caused by discarding negative values. In practice, we could try to improve our estimate of the parameter values that maximize the likelihood by optimization procedures [33]. In practice, enforcing the positivity constraint while doing it may be problematic, and the best samples provide reasonable approximations for our purposes.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

The number of people affected by the virus with a SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

When the distributions under study are symmetric, it will be close to \( \sigma \). Otherwise, it may depart from it. In our case, slight asymmetry is caused by discarding negative values. In principle, we could try to improve our estimate of the parameter values that maximize the likelihood by optimization procedures [33]. In practice, enforcing the positivity constraint while doing it may be problematic, and the best samples provide reasonable approximations for our purposes.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.

Panels (a)–(b) in Fig. 3 compare the observations that would be obtained with \( \sigma \) and \( \sigma \) to the original data. If we solve the SIJR model for a longer time, for instance, 14 days more, we reach about \( 5 \times 10^3 \) inflected individuals, see panel (c), and about \( 2.5 \times 3.5 \times 10^3 \) affected people in the absence of contention measures.
except for $\mathcal{S}_T$ the final values from the previous period at populations as a result of confinement measures. In each $\mathcal{i}$ periods for the data shown in Fig. 1(a) are marked by variations in these the first period of free growth model with two populations on the subpopulations by means of the SEIJR model distinguishing two

$\mathcal{S}$ $\mathcal{S}$ $\mathcal{T}$ $\mathcal{T}$ $\mathcal{I}$ $\mathcal{J}$ $\mathcal{R}$ $\mathcal{D}$ $\max$ $\min$ $\mathcal{W}$ $\mathcal{Q}$ $\mathcal{a}$ $\mathcal{b}$ $\mathcal{c}$ $\mathcal{d}$ $\mathcal{e}$ $\mathcal{f}$ $\mathcal{g}$ $\mathcal{h}$

$\mathcal{S}$ $\mathcal{S}$ $\mathcal{T}$ $\mathcal{T}$ $\mathcal{I}$ $\mathcal{J}$ $\mathcal{R}$ $\mathcal{D}$ $\max$ $\min$ $\mathcal{W}$ $\mathcal{Q}$ $\mathcal{a}$ $\mathcal{b}$ $\mathcal{c}$ $\mathcal{d}$ $\mathcal{e}$ $\mathcal{f}$ $\mathcal{g}$ $\mathcal{h}$

$\mathcal{S}$ $\mathcal{S}$ $\mathcal{T}$ $\mathcal{T}$ $\mathcal{I}$ $\mathcal{J}$ $\mathcal{R}$ $\mathcal{D}$ $\max$ $\min$ $\mathcal{W}$ $\mathcal{Q}$ $\mathcal{a}$ $\mathcal{b}$ $\mathcal{c}$ $\mathcal{d}$ $\mathcal{e}$ $\mathcal{f}$ $\mathcal{g}$ $\mathcal{h}$

20465 and 25327 for SEIR, respectively. See Table 2 for a comparison of the parameter values for both models. Note the high transmission rate $\beta$ (about 0.6) and the low diagnosis rate $\alpha$ (about 0.2). Most infected individuals are not detected. If we solve the SEIR model for a longer time, for instance, 14 days more, we reach about $2.2\times3.2\times10^6$ diagnosed individuals and $3.6\times5.4\times10^6$ affected people for $\mathcal{S}\max$ and $\mathcal{S}\min$ in the absence of contention measures.

In the next section we study the influence of contention measures on the subpopulations by means of the SEIJR model distinguishing two populations, one of which is confined.

Incorporating the effect of contention measures

To incorporate the effect of confinement we consider the SEIR model with two populations $\mathcal{S}_1$ (unconfined) and $\mathcal{S}_2$ (confined). During the first period of free growth $[\mathcal{t}_{\text{in}}, \mathcal{T}_1]$, we have $\mathcal{S}_2 = 0$. The different periods for the data shown in Fig. 1(a) are marked by variations in these populations as a result of confinement measures. In each $\mathcal{i}$-th period $[\mathcal{T}_{\mathcal{i}-1}, \mathcal{T}_\mathcal{i}]$, $\mathcal{i} > 1$, we solve the SEIR model (1) using as initial values the final values from the previous period at $\mathcal{T}_{\mathcal{i}-1}$, for all the variables except for $\mathcal{S}_1$ and $\mathcal{S}_2$:

- Period 2: $(1-\rho)\mathcal{S}_1(T_1)$ and $\mathcal{S}_2(T_1) + \rho \mathcal{S}_1(T_1)$ are used as initial data for $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.
- Period 3: $(1-\rho)\mathcal{S}_1(T_2)$ and $\mathcal{S}_2(T_2) + \rho \mathcal{S}_1(T_2)$ are used as initial data for $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.
- Period 4: $\mathcal{S}_1(T_3) + (1-\rho)\mathcal{S}_2(T_3)$ and $\rho \mathcal{S}_1(T_3)$ are used as initial data for $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.

Recall that in the first period, the initial values for all the variables are zero, except $\mathcal{E}(\mathcal{t}_{\text{in}}) = 1$ and $\mathcal{S}_1(\mathcal{t}_{\text{in}}) = \mathcal{N} -1$. No parameter $\mathcal{t}_{\text{in}}$ appears in the next periods, we set it equal to zero. Instead, we introduce $\rho \in (0, 1)$ to quantify the abrupt changes in the fraction of people confined at the start of each period. We assume that the transmission rate for $\mathcal{S}_2$ is lower by a factor $\rho$, that is, $\rho \beta$ instead of $\beta$, due to the reduction of contacts with other people. Due to possible interaction with already sick people or people still working outside at home, we cannot set it equal to zero.

We adapt the framework presented in Section “Uncertainty quantification by Bayesian techniques” assembling these periods as we explain next. To consider stages $\mathcal{i}$, $\mathcal{i} = 1, \ldots, \mathcal{Q}$, we multiply the number of parameters by $\mathcal{Q}$. The first block of $\mathcal{Q}$ parameters is the standard one for the first period. The remaining blocks correspond each to one additional period, with $\mathcal{t}_{\text{in}}$ replaced by $\rho$. We keep the same initial guesses of the parameters used in Section "Uncertainty in the initial stage" as prior knowledge in all the periods, except for $\rho$, which is set equal to $3/4$, $4/5$, $15/16$ respectively, an approximation of the population switches at the different stages. The deviations are kept equal to 0.1 for all, except $\mathcal{t}_{\text{in}}$, for which we set it equal to 10. As for the data, we keep the same deviations as in Section "Uncertainty in the initial stage" in all the periods, in the absence of better information.

Let us consider first the initial confinement period. Fig. 5(a) compares data the evolution of the diagnosed subpopulations. Population dynamics is calculated solving the SEIR model in two sequential steps, in $[\mathcal{t}_{\text{in}}, \mathcal{T}_1]$ and $[\mathcal{T}_1, \mathcal{T}_2]$, using in each of them the parameter values obtained for that period and the initial data stipulated earlier. Panel (b) represents the solutions of the SEIR model including the contribution of undiagnosed and asymptomatic individuals. Panel (d) compares the distribution of some parameters in the two periods. The transmission rate $\beta$ increases slightly in the second period, while the diagnose rate remains low. These histograms are discretizations of the probability, so that the height of each bin is the number of samples in the bin divided by the total number and by the basis of the bins (which is the same for the histograms corresponding to the same parameters in this figure.
A. Carpio and E. Pierret

Fig. 5. First and second periods: (a) Data for diagnosed (asterisks), dead (crosses), recovered (triangles) and active (squares) cases, compared to solutions of (1) and (10) for $\nu_{\text{max}}$ (solid) and $\nu_{\text{mean}}$ (dashed), extended in (c) for a longer time. (b) SEIJR simulations of the numbers of exposed, infective, recovered and dead individuals for $\nu_{\text{mean}}$ including the undiagnosed and asymptomatic. (d) Histograms comparing the distribution of some parameters in the two periods. (e) Histograms representing the probability of the status of different populations at the end. Vertical lines mark the values for $\nu_{\text{max}}$ (red dot–dashed line), $\nu_{\text{mean}}$ (green dashed line), and the mean for all samples (black dotted line). Sampling parameters $W = 500$, $S = 4 \times 10^6$, $B = S/4$ and $a = 2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. Same as Fig. 5(e) and (c) for three periods of increasing confinement. Note the decrease in the numbers of exposed $E$ and infected $I$ cases. Sampling parameters $W = 500$, $S = 4 \times 10^6$, $B = S/4$ and $a = 2$. and the previous ones to allow for comparisons). Fig. 5(e) quantifies uncertainty in the total number of people affected by the virus after these two periods. If we keep the parameter values $\nu_{\text{max}}$ or $\nu_{\text{mean}}$ up to time $T_3 > T_2$, growth slows down, but it does not stabilize, see Fig. 5(c).

We incorporate next the third additional period in which an even larger fraction of the population is confined at home. The results are reproduced in Fig. 6. Finally, the growth trend moderates, also in predictions for longer times, see Fig. 1(b). Table 3 reports the mean values $\nu_{\text{mean}}$ obtained after MCMC sampling, as well as the values corresponding to the best sample $\nu_{\text{max}}$. As mentioned earlier, $\nu_{\text{max}}$ has not a statistical meaning. It represents a best fit whose coefficients may fluctuate a bit with the number of samples. Instead, $\nu_{\text{mean}}$ conveys a statistical trend of the coefficients of the samples. Comparing the values of $\beta$ for $\nu_{\text{mean}}$, we remark an increase in $\beta$ in the second period. This fact is also observed in $\nu_{\text{max}}$ and the trend was already present in the histograms for $\beta$ in Fig. 5(d). According to the information available on the Spanish outbreak, and taking into account that infected people can take up to 14 days to show symptoms, this might be a delayed reflection of crowd gatherings occurred at the end of the first period, or also, a result of the lack of protective equipment for overwhelmed health care and security workers. We also observe a reduction in the mean recovery rates for $\gamma_1$ and $\gamma_2$ in the second period, which may be reflection of the saturation of the health care system and the scarceness of medical resources during the second period. Notice that the diagnose rate $\alpha$ is quite low. A large fraction of affected people remains undetected.

In a fourth period, a fraction of the population is released from confinement. The number of undiagnosed and exposed individuals is depleted and the spread of the epidemic is contained. Unlike before, the
triangles, we have the inference region, corresponding to the data we
from the 'prediction' regions for each of them. At the back of the
(separated by a bar) (dead, recovered and active). Colored triangles separate the 'inference'
Yellow circles represent the data: total counts of diagnosed people
$J$ samples. Thicker lines represent the total number of diagnosed cases
the number of quarantined infected and asymptomatic individuals.
usefulness of tests would be to increase the diagnose rate, augmenting
rate $\gamma$ second period, while $\beta$ using less data. However, the same trends persist: increase of
previous periods change slightly and we infer more moderate numbers
information in the analysis. The best coefficient values estimated for
SEIJR solutions close to the prior while most of them remain close to the
data as the model coefficients range through the sampled parameters.
This may be a consequence of fixing prior guesses for the model
parameters that worsen with time. Note that the predictions that would be
obtained using the prior are rather poor, compared to true counts, as
time grows. However, the predictions provided by $v_{\max}$ fit the data
quite well, even for later times, see Fig. 1(b).

Note that as we add data from new periods, we are including more
information in the analysis. The best coefficient values estimated for
previous periods change slightly and we infer more moderate numbers
of affected individuals, as compared with the previous studies done
using less data. However, the same trends persist: increase of $\beta$ in the
second period, while $\gamma_1$ and $\gamma_2$ decrease, decrease of $\delta$ and low diagnose rate $\alpha$. Very few tests were done during these periods. In fact, the
usefulness of tests would be to increase the diagnose rate, augmenting
the number of quarantined infected and asymptomatic individuals.

Fig. 1(b) provides a global view of our analysis. Shaded areas represent the total number of diagnosed cases $J + D_j + R_j$ obtained solving (1) and (10) for the last 1000 sampled parameters in each of
the four frameworks we have considered: red for Period 1, green for
Periods 1–2, blue for Periods 1–2–3, magenta for Periods 1–2–3–
4. Dotted lines represent the mean of the curves obtained for all the
samples. Thicker lines represent the total number of diagnosed cases
$J + D_j + R_j$ for $v_{\max}$ (solid), $v_{\text{mean}}$ (dashed) and $v_0$ (dash–dotted). Yellow circles represent the data: total counts of diagnosed people
(dead, recovered and active). Colored triangles separate the 'inference'
from the 'prediction' regions for each of them. At the back of the
triangles, we have the inference region, corresponding to the data we
use to infer the parameter values and the total number of affected
people. At the front of the triangles, we use model solutions to predict
the time evolution keeping the conditions of the last period considered
in the inference studies. Taking no measures leads to the evolution
represented in red. Confining people who are able to work online or
do not work in basic activities results in the dynamics marked in green.
Extending the confinement to all the population not working in strictly
essential activities leads to the forecast painted in blue. Releasing this
last fraction of the population results in the evolution represented in
magenta. Notice that the solid magenta curve corresponding to $v_{\max}$
agrees very well with the data past day 68 (last day used to calculate
it), whereas some magenta samples deviate considerably. This fact is
reflected in the dotted averages, which define somehow a confidence
region. After day 68 the population was released from confinement by
stages, and the use of masks was enforced, lowering the risk for the
users. The country remained closed. The different predictions associat-
ted to the four inference studies we carried out are not only due to
the confinement or the release of population fractions, but to the fact
that we allow for variations in the model coefficients in the different
periods to adapt them to additional amounts of data. The fact that the
transmission coefficient $\beta$ decreases with time due to improved
conditions is fundamental.

These studies are limited by the data quality. As mentioned earlier,
the order of magnitude of the population counts in official records
changes noticeably when only PCR confirmed cases are taken into
account or also probable cases are included. In the Spanish outbreak,
the number of probable cases may have been five times higher and
the number of dead individuals twice as much. Repeating our previous
studies scaling the data in that way, we find estimates about 2 million
people, consistent with the official conclusions inferred from selected
testing campaigns.

Finally, let us focus on the available data until the borders opened on
July 2, 2020, after which the country was no longer closed. As said
before, home confinement had been replaced by mask usage indoors and
outdoors. Fig. 8 compares the predictions obtained for the period
May 2 until July 15, 2020, with the parameters corresponding to $v_{\max}$
obtained while fitting the model to the four previous periods. The
number of total diagnosed people is well fitted. This suggests that
masks were an effective tool to contain the spread. The number of dead
people is overestimated, due to the fact that the official number of
dead individuals was reduced by about 2000 people on May 25. We
cannot compare our predictions with the official counts of recovered
individuals because there were no longer updated in this period.

Conclusions

The attempt to devise mathematical models to study the progress-
ion of a pandemic faces the need to handle large uncertainty in the

| $v_{\text{max}}$ 1st | $v_{\text{max}}$ 2nd | $v_{\text{max}}$ 3rd | $v_{\text{mean}}$ 1st | $v_{\text{mean}}$ 2nd | $v_{\text{mean}}$ 3rd |
|------------------|------------------|------------------|------------------|------------------|------------------|
| $l_\text{int}$ | 0.3479 | 0.7202 | 0.2236 | 7.9770 | 0.8064 | 2.0257 |
| $\beta$ | 0.6173 | 0.6902 | 0.0189 | 0.6894 | 0.0702 | 0.6796 |
| $\gamma_1$ | 0.0741 | 0.0363 | 0.0446 | 0.0410 | 0.0290 | 0.0318 |
| $\gamma_2$ | 0.1426 | 0.0437 | 0.0551 | 0.0532 | 0.0343 | 0.0461 |
| $\delta$ | 0.0696 | 0.0310 | 0.0131 | 0.0141 | 0.0180 | 0.0098 |
| $\alpha$ | 0.1541 | 0.2148 | 0.2343 | 0.1791 | 0.1851 | 0.1022 |
| $\gamma$ | 0.1056 | 0.1245 | 0.1031 | 0.1253 | 0.1800 | 0.0219 |
| $\theta$ | 0.4978 | 0.5301 | 0.5082 | 0.3872 | 0.7778 | 0.4234 |
| $\rho$ | 0.1139 | 0.1693 | 0.1643 | 0.1845 | 0.0001 | 0.0001 |
| $k$ | 0.4984 | 0.5106 | 0.5251 | 0.5802 | 0.5703 | 0.5339 |

**Fig. 7.** Same as Fig. 6(a)–(b) for four periods. Additional dotted lines in the lower part of panel (a) represent solutions of (3) and (10) for $v_0$. The numbers of exposed $E$, infected $I$ and active diagnosed $J$ individuals are depicted.
available data. We have developed a Bayesian framework to quantify uncertainty in the effects of lockdown measures through the coefficients of SEIJR and SIRJ models for human-to-human transmission. A key idea is the introduction of two populations, one of which has a lower risk of infection than the other. Lower risk may be due to confinement, as it happens for the data we consider here, or to preventive measures, such as the use of masks. Therefore, our methodology is not constrained to lockdown measures.

These techniques allow us to calibrate important magnitudes to forecast the evolution of the epidemic, such as the variation in the total number of affected people (including asymptomatic individuals), and could be adapted to infer coefficients from data from any country. We show how enforcing measures that deplete the number of undiagnosed and asymptomatic individuals, while reducing the transmission rate, we can stop the spread. We have focused on the data available for Spain during the first wave, which shows well differentiated data periods according to the measures taken. Moreover, the borders of the country remained closed, so that the system was indeed closed, as assumed by SIR type models. We see that the model coefficients in each period vary with the circumstances. For instance, transmission rates may augment as a result of increased interaction and lack of protective measures and recovery rates may decrease as a result of scarceness of resources. The diagnose rate is low, resulting in large number of undiagnosed individuals. Performing more PCR tests would increase the diagnose rate, allowing to quarantine more infected and asymptomatic individuals.

An additional difficulty when applying this inference framework for large periods of time (months) is the fact that uncertainty in the observed data accumulates over time when using cumulative data. This poses the problem of selecting adequate variances for the analysis. In the absence of reliable information in that respect, we have kept them fixed. Calculations with daily data do not show significative differences in the observed trends in our case. Moreover, we have used official data for PCR confirmed patients only. The effect of adding probable cases to the observed data accumulates over time when using cumulative data. This may not be the case here, thus additional studies taking this factor into account would be advisable [18,20]. Furthermore, standard SIR type models [12] are formulated for closed systems. Introducing spatial mobility [6,12] is an important issue that should be a subject for future work. Moreover, imperfect implementation of contention measures leads to delays, which might be better described by differential-delay models [35]. We have focused on human-to-human transmission here. Coronavirus originates in animals, such as bats, and arrive to humans through intermediate animal species which act as reservoirs for future waves [36], subject deserving further studies.

CRediT authorship contribution statement

Ana Carpio: Conceptualization, Supervision, Methodology, Analysis and Validation, Resources, Funding acquisition, Writing – original draft. Emile Pierret: Data curation, Visualization, Analysis and validation, Software development.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been partially supported by the FEDER /Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación grants No. MTM2017-84446-C2-1-R, PID2020-112796RB-C21 and ENS Paris Saclay program for student internships abroad. A. Carpio thanks G. Stadler for nice discussions.

Appendix: solutions of the SIRJ model

Let us obtain explicit expressions for the solution of the (2)-(8) model. Consider the Eqs. (3)-(4) for I and J. Set $D_1 = \alpha + \gamma_1 + \delta_1$, $D_2 = \gamma_2 + \delta$. The system matrix is

$$A = \begin{pmatrix} \beta - D_1 & \xi \beta \\ \alpha & -D_2 \end{pmatrix}$$

with eigenvalues

$$\lambda_1 = \frac{D_1 - D_2 - \alpha}{2} - \frac{1}{2} \sqrt{\beta^2 - 2\beta D_1 + 2\alpha \xi \beta + \alpha^2 - 2D_1 D_2 + D_2^2},$$

$$\lambda_2 = \frac{D_1 - D_2 - \alpha}{2} + \frac{1}{2} \sqrt{\beta^2 - 2\beta D_1 + 2\alpha \xi \beta + \alpha^2 - 2D_1 D_2 + D_2^2},$$

and eigenvectors:

$$\mathbf{v}_1 = (-\xi \beta, \beta - D_1 - \lambda_1), \quad \mathbf{v}_2 = (-\xi \beta, \beta - D_1 - \lambda_2).$$

The general solution is

$$\begin{cases} I(t) = z_1 \mathbf{v}_1 e^{\lambda_1 t} + z_2 \mathbf{v}_2 e^{\lambda_2 t}, \\
J(t) = z_1 \mathbf{v}_1 e^{\lambda_1 t} + z_2 \mathbf{v}_2 e^{\lambda_2 t}, \end{cases} \quad z_1, z_2 \in \mathbb{R}.$$
The cumulative number of diagnosed people $J_t$ is then the integral of this magnitude, $J'_t = J$ starting from zero $J(0) = 0$:

$$J'_t = \frac{a}{\lambda_1 - \lambda_2} \left( e^{\lambda_1 t} - 1 - e^{\lambda_2 t} - 1 \right). \tag{25}$$

We can now integrate the equations for $S$, $R$ and $D$:

$$S(t) = -\beta I(t) - \beta S J_t(t), \tag{26}$$
$$R(t) = \gamma S I(t) + \gamma J J_t(t), \tag{27}$$
$$D(t) = \delta I(t) + \delta J J_t(t). \tag{28}$$

If we work with the diagnosed recovered and the diagnosed dead, we get

$$R'_t = \gamma S J_t(t), \tag{29}$$
$$D'_t = \delta J J_t(t). \tag{30}$$

The formulas given here set $t_0 = 0$. To use them with initial data at a generic $t_0$ we just replace $t$ by $t + t_0$ in the formulas obtained here.

References

[1] Rothana HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109:105433.
[2] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A novel coronavirus from patients with pneumonia in China. N Engl J Med 2020;382(8):727–33.
[3] Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College London; 2020; Report 9, http://dx.doi.org/10.25561/77482.
[4] Khaliﬁe S, Mitra T, Bandyopadhyay A, Schips M, Mascheroni P, Vanella P, Lange B, Binder S, Meyer-Hermann M. Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for public health measures. BMC Med 2021;19:32.
[5] Ambikapathy B, Krishnamurthy K. Modelling applications to assess the impact of lockdown on covid-19 transmission in India: Model development and validation. JMRI Public Health Surveillance 2020;6(2). http://dx.doi.org/10.2196/19568.
[6] Bouchahti A, Jebran A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. Chaos Solitons Fractals 2020;138:109941.
[7] Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Magerl S, Gavenčiak T, Ambikapathy B, Krishnamurthy K. Mathematical modelling to assess the impact of NPIs to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College London; 2020; Report 9, http://dx.doi.org/10.25561/77482.
[8] Khaliﬁe S, Mitra T, Bandyopadhyay A, Schips M, Mascheroni P, Vanella P, Lange B, Binder S, Meyer-Hermann M. Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for public health measures. BMC Med 2021;19:32.
[9] Bouchahti A, Jebran A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. Chaos Solitons Fractals 2020;138:109941.
[10] Ambikapathy B, Krishnamurthy K. Modelling applications to assess the impact of lockdown on covid-19 transmission in India: Model development and validation. JMRI Public Health Surveillance 2020;6(2). http://dx.doi.org/10.2196/19568.
[11] Bouchahti A, Jebran A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. Chaos Solitons Fractals 2020;138:109941.
[12] Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Magerl S, Gavenčiak T, Ambikapathy B, Krishnamurthy K. Mathematical modelling to assess the impact of NPIs to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College London; 2020; Report 9, http://dx.doi.org/10.25561/77482.
[13] Khaliﬁe S, Mitra T, Bandyopadhyay A, Schips M, Mascheroni P, Vanella P, Lange B, Binder S, Meyer-Hermann M. Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for public health measures. BMC Med 2021;19:32.
[14] Bouchahti A, Jebran A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. Chaos Solitons Fractals 2020;138:109941.
[15] Ambikapathy B, Krishnamurthy K. Modelling applications to assess the impact of lockdown on covid-19 transmission in India: Model development and validation. JMRI Public Health Surveillance 2020;6(2). http://dx.doi.org/10.2196/19568.
[16] Bouchahti A, Jebran A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. Chaos Solitons Fractals 2020;138:109941.
[17] Ambikapathy B, Krishnamurthy K. Modelling applications to assess the impact of lockdown on covid-19 transmission in India: Model development and validation. JMRI Public Health Surveillance 2020;6(2). http://dx.doi.org/10.2196/19568.
[18] Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Magerl S, Gavenčiak T, Ambikapathy B, Krishnamurthy K. Mathematical modelling to assess the impact of NPIs to Reduce COVID-19 Mortality and Healthcare Demand. Imperial College London; 2020; Report 9, http://dx.doi.org/10.25561/77482.
[19] Khaliﬁe S, Mitra T, Bandyopadhyay A, Schips M, Mascheroni P, Vanella P, Lange B, Binder S, Meyer-Hermann M. Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for public health measures. BMC Med 2021;19:32.