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.

CHAPTER 7

Parameter identification in epidemiological models

Ana Carpio[a] **and Emile Pierret**[b]

[a] *Universidad Complutense de Madrid, Departamento de Análisis Matemático y Matemática Aplicada, Madrid, Spain* [b] *ENS Paris-Saclay, CMLA, Gif-sur-Yvette, France*

7.1 Introduction

After the work of Kermack and McKendrick [18], SIR type models became a standard tool in epidemiology [10]. These models involve populations of susceptible S, infected I, and recovered R individuals, assuming immunity of the latter [2,9,27]. Many variants have been developed to include additional details. SEIR variants single out the individuals exposed to the virus E, which may also become infective [20,21]. Immunity of the recovered is dropped in SEIRS systems [19,23]. SEIJR and SIJR models enforce quarantine and isolation measures on diagnosed individuals [8,11]. Delay in responses is incorporated in [26]. In general, the specific structure of the selected models depends on the available information and on assumptions about the epidemic spread [1,15]. There is a vast mathematical literature on the qualitative behavior of such systems [10].

To achieve a predictive value, one must fit the model parameters to available data. Optimization and adjoint-based data assimilation techniques, for instance, seek to chose parameters in such a way that the difference between data and model predictions is minimized [11,27]. Nevertheless, official population counts for epidemiological studies are affected by different sources of noise and uncertainty. The diagnosed, recovered, and dead individuals notified in official reports are often defined in different ways in different regions or change as official criteria are modified. Uncertainty in the data spreads to any analyses based on them. Therefore, it is convenient to pursue approaches that quantify uncertainty [4,9,13,22] and provide information on relevant parameter ranges, rather than specific values. Here, we develop a general framework to infer coefficients for epidemiological models from population counts with quantified uncertainty. We illustrate the method on a SEIJR model including contention measures sequentially enforced or lifted, using data since the onset of current covid19 pandemic from the Madrid region, Spain, for which different periods are well established. In this way, we can relate variations in the model rates and in the population distribution with policies enforced as

Mathematical Analysis of Infectious Diseases
https://doi.org/10.1016/B978-0-32-390504-6.00012-7

time grows. This study assumes a closed system: the total population is conserved. Migration effects introduce new features and render more difficult epidemic stabilization, even when strong distancing measures are enforced [7]. We modify the SEIJR model to include population exchanges, discussing the potential of different parameters to control the system. Simple constrained optimization formulations allow us to explore possible parameter ranges which would prevent exponential growth.

The paper is organized as follows. Section 7.2 describes the SEIJR model under study. Section 7.3 introduces a Bayesian framework to identify the model parameters with quantified uncertainty in response to different contention policies. We illustrate the resulting predictions for coefficients, initial data, and final population distribution, including asymptomatic individuals in Section 7.4. Migration effects are incorporated in Section 7.5, whereas Section 7.6 considers a possible control formulation. Finally, Section 7.7 contains our conclusions.

7.2 SEIJR models for closed systems

SEIJR models divide the total population into several categories: susceptible (S), exposed (E), infective (I), diagnosed (J), and recovered (R) individuals. Exposed subjects E are asymptomatic, in principle, and possibly infectious. They can spread the virus at a rate q and develop symptoms at a rate k. Infective symptomatic individuals I are diagnosed at a rate α. Diagnosed subjects J are quarantined. Sick individuals who recover cumulate in the class R. Susceptible individuals can display different susceptibility for different reasons, such as age, sex or genetic characteristics [8], but also due to confinement/protective measures enforced. Standard SEIJR models [8] assume that 1) infection spread occurs in a closed system, 2) the death rate is the same for all individuals, neglecting death by other causes, 3) recovered subjects have immunity, 4) diagnosed individuals are quarantined, and 5) time delays in responses are ignored.

A formulation including two populations S_1 and S_2 of different susceptibility reads:

$$\begin{aligned}
\frac{dS_1}{dt} &= -\beta S_1(t)\frac{I(t)+qE(t)+\ell J(t)}{N},\\
\frac{dS_2}{dt} &= -\beta p S_2(t)\frac{I(t)+qE(t)+\ell J(t)}{N},\\
\frac{dE}{dt} &= \beta(S_1(t)+pS_2(t))\frac{I(t)+qE(t)+\ell J(t)}{N} - kE(t),\\
\frac{dI}{dt} &= kE(t) - (\alpha+\gamma_1+\delta)I(t),\\
\frac{dJ}{dt} &= \alpha I(t) - (\gamma_2+\delta)J(t),
\end{aligned} \tag{7.1}$$

$$\frac{dR}{dt} = \gamma_1 I(t) + \gamma_2 J(t),$$
$$\frac{dD}{dt} = \delta I(t) + \delta J(t).$$

Here $N = S_1 + S_2 + E + I + J + R + D$ is the total population number, which is a conserved quantity, and D quantifies the dead. Time is measured in days. The transmission rate β represents how susceptible individuals become virus spreaders. The risk of infection for S_2 is lower than the risk for S_1 by a factor p. The reduced impact of diagnosed individuals on transmission is represented through the parameter ℓ. Recovery rates are γ_1 for the infective and γ_2 for the diagnosed, while their mortality rates are denoted by δ. Two constraints are imposed on these rates $\alpha > \gamma_1$ and $\gamma_2^{-1} = \gamma_1^{-1} - \alpha^{-1}$ [8].

Introducing an additional parameter t_{in}, the time at which spread starts [11], we can set initial conditions. The values for all populations at t_{in} are zero, except for $E(t_{\text{in}}) = 1$ and $S_1(t_{\text{in}}) = N - 1$. Alternatively, the initial population can be partitioned as $S_1 = (1-\rho)S$, $S_2 = \rho S$, ρ being the fraction of the susceptible population S at a lower risk of infection.

The reproduction number for this model is given by [8]

$$\mathcal{R}_0 = \beta(\rho + p(1-\rho))\left(\frac{q}{k} + \frac{1}{\alpha + \gamma_1 + \delta} + \frac{\alpha\ell}{(\alpha + \gamma_1 + \delta)(\gamma_2 + \delta)}\right).$$

It represents the expected number of cases directly caused by a single case in a population in which all individuals are susceptible to infection (no other individuals are infected or immunized). In contrast, the effective reproduction number $\mathcal{R}_e$ is the number of cases produced in the current state. This type of models can account for some features observed in SARS or covid type epidemics, such as the emergence of asymptomatic and symptomatic individuals, unequal susceptibility, and superspread events.

These systems display a rich variety of behaviors, which depend on the specific choices of parameters. For predictive purposes, it is essential to select realistic parameter ranges. In the next section, we formulate a Bayesian framework to identify parameters from available population counts, with quantified uncertainty.

7.3 Uncertainty quantification by Bayesian techniques

Bayes' theorem quantifies the probability of an event using prior knowledge about it [17]. It states that the posterior probability of observing a finite number of parameters $\boldsymbol{\nu}$ given data $\mathbf{d}$ is

$$p(\boldsymbol{\nu}|\mathbf{d}) = \frac{p(\mathbf{d}|\boldsymbol{\nu})p(\boldsymbol{\nu})}{p(\mathbf{d})}, \tag{7.2}$$

where $p(\boldsymbol{\nu})$ represents our prior knowledge on the parameters $\boldsymbol{\nu}$ and $p(\mathbf{d}|\boldsymbol{\nu})$ is the conditional probability (or likelihood) of observing data $\mathbf{d}$ given parameters $\boldsymbol{\nu}$. The normalization factor $p(\mathbf{d})$ describes the probability of the data. It can be obtained integrating $p(\mathbf{d}|\boldsymbol{\nu})p(\boldsymbol{\nu})$ with respect to $\boldsymbol{\nu}$.

7.3.1 Bayesian formulation for SEIJR coefficients

In our framework, the parameters are the model parameters,

$$\boldsymbol{\nu} = (t_{\text{in}}, \beta, \gamma_2, \delta, \alpha, \ell, q, p, k), \quad \gamma_1^{-1} = \gamma_2^{-1} + \alpha^{-1}, \tag{7.3}$$

while the prior distribution and the likelihood are defined as follows [5].

To construct the prior distribution, we start with a multivaluate normal distribution of mean $\boldsymbol{\nu}_0$ and covariance matrix $\mathbf{G}_{\text{pr}}$

$$p(\boldsymbol{\nu}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\mathbf{G}_{\text{pr}}|}} \exp(-\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)^t \mathbf{G}_{\text{pr}}^{-1}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)),$$

where n is the number of parameters. The mean $\boldsymbol{\nu}_0$ is a parameter guess, whereas $\mathbf{G}_{\text{pr}}$ is a diagonal matrix with elements σ_i^2, $i = 1, ..., n$. Since our parameters must be positive and a Gaussian distribution admits negative values, we set as prior distribution

$$p_{\text{pr}}(\boldsymbol{\nu}) = \begin{cases} \exp(-\frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)^t \mathbf{G}_{\text{pr}}^{-1}(\boldsymbol{\nu} - \boldsymbol{\nu}_0)), & \nu_j \geq 0,\ j = 1, ..., n, \\ 0, & \nu_j < 0, \text{ for some } j. \end{cases} \tag{7.4}$$

For this expression to be a probability we should calculate the normalizing factor so that its integral is one. However, our subsequent calculations do not need it. We will work with the unnormalized distribution.

We define the conditional probability density $p(\mathbf{d}|\boldsymbol{\nu})$ as

$$p(\mathbf{d}|\boldsymbol{\nu}) = \frac{1}{(2\pi)^{L/2}\sqrt{|\mathbf{G}_{\text{n}}|}} \exp\Big(-\frac{1}{2} \|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|^2_{\mathbf{G}_{\text{n}}^{-1}} \Big), \tag{7.5}$$

where $\|\mathbf{v}\|^2_{\mathbf{G}_{\text{n}}^{-1}} = \overline{\mathbf{v}}^t \mathbf{G}_{\text{n}}^{-1} \mathbf{v}$ and L the amount of data. We assume additive Gaussian noise, i.e., the data $\mathbf{d}$ and the true parameters are related by

$$\mathbf{d} = \mathbf{f}(\boldsymbol{\nu}_{\text{true}}) + \boldsymbol{\varepsilon}, \tag{7.6}$$

$\mathbf{f}(\boldsymbol{\nu})$ being the observation operator. Here, the noise $\boldsymbol{\varepsilon}$ is distributed as a multivariate Gaussian $\mathcal{N}(0, \mathbf{G}_{\text{n}})$ with mean zero and covariance matrix $\mathbf{G}_{\text{n}}$. We define the observation operator during M consecutive days as

$$\mathbf{f}(\boldsymbol{\nu}) = (J(1), ..., J(M), R_J(1), ..., R_J(M), D_J(1), ..., D_J(M)), \tag{7.7}$$

where the dynamics of the diagnosed recovered R_J and diagnosed dead D_J are governed by

$$\frac{dR_J}{dt} = \gamma_2 J(t), \quad \frac{dD_J}{dt} = \delta J(t). \tag{7.8}$$

This observation operator is consistent with the data available in practice: daily cumulative counts of diagnosed individuals $\tilde{j}_m$, diagnosed recovered individuals r_m, and diagnosed dead d_m individuals, for $m = 1, ..., M$, see [6]. Putting the three blocks of data together we have

$$\mathbf{d} = (j_1, ..., j_M, r_1, ..., r_M, d_1, ..., d_M), \tag{7.9}$$

where $j_m = \tilde{j}_m - r_m - d_m$ are the active diagnosed, those who are neither dead nor recovered. In (7.5), we compare observations to the data $\mathbf{d}$ using the distance $\frac{1}{2}\|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|^2_{\mathbf{G}_n^{-1}}$. Notice that we distinguish diagnosed individuals who are dead, recovered, and still sick, and compare with model predictions for them neglecting the contribution of the undiagnosed, unlike [9]. For simplicity, we assume that $\mathbf{G}_n$ is a real diagonal matrix, $\mathbf{G}_n = \text{diag}(\sigma_1^2, \ldots, \sigma_L^2)$, and set all the variances for the same magnitude equal to a constant σ_J^2, σ_R^2, σ_D^2. Thus, $\sqrt{|\mathbf{G}_n|} = \sigma_J^M \sigma_R^M \sigma_R^M$, where $L = 3M$ is the number of data considered. Simpler cost functionals in the literature use only total cumulative case counts [11].

Inserting (7.4) and (7.5) in (7.2) and discarding normalization constants, the posterior distribution becomes

$$p_{pt}(\boldsymbol{\nu}) \sim \exp\left(-\frac{1}{2}\|\mathbf{f}(\boldsymbol{\nu}) - \mathbf{d}\|^2_{\mathbf{G}_n^{-1}} - \frac{1}{2}\|\boldsymbol{\nu} - \boldsymbol{\nu}_0\|^2_{\mathbf{G}_{pr}^{-1}}\right). \tag{7.10}$$

By studying this distribution, we may infer the coefficients of the model with quantified uncertainty, depending on the selected prior and covariances. This framework can be modified to consider different sets of parameters. For instance, we could suppress t_{in} and consider the initial populations at a certain time also unknowns to be identified, which we will do later.

7.3.2 Prior selection

We will use a guess for the model coefficients obtained by constrained optimization to define the mean $\boldsymbol{\nu}_0$ in (7.4). To do so, we first simplify the SEIJR system. We assume that the exposed class E is neglected, susceptibility is not distinguished ($S_1 = S_2 = S$, $p = 1$), and the infected are a small fraction of the whole population, that is, $\frac{S}{N} \sim 1$. The SEIJR model

becomes a SIJR simplification

$$\begin{aligned}
\frac{dS}{dt} &= -\beta(I + \ell J),\\
\frac{dI}{dt} &= (\beta - (\alpha + \gamma_1 + \delta))I + \ell\beta J,\\
\frac{dJ}{dt} &= \alpha I - (\gamma_2 + \delta)J,\\
\frac{dR}{dt} &= \gamma_1 I + \gamma_2 J,\\
\frac{dD}{dt} &= \delta(I + J),
\end{aligned} \tag{7.11}$$

with initial conditions $S(t_{\text{in}}) = N - 1$, $I(t_{\text{in}}) = 1$, $J(t_{\text{in}}) = R(t_{\text{in}}) = 0 = D(t_{\text{in}})$. Here, $N = S + I + J + R + D$ is the total population number, which is a conserved quantity. This system admits the explicit solution

$$\begin{aligned}
I(t) &= c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t},\\
J(t) &= \frac{\alpha}{\lambda_1 - \lambda_2} e^{\lambda_1 t} + \frac{\alpha}{\lambda_2 - \lambda_1} e^{\lambda_2 t},
\end{aligned} \tag{7.12}$$

where

$$\begin{aligned}
\lambda_1 &= \frac{\beta - D_1 - D_2}{2} - \frac{1}{2}\sqrt{\beta^2 - 2\beta D_1 + 2\beta D_2 + 4\alpha\ell\beta + D_1^2 - 2D_1 D_2 + D_2^2},\\
\lambda_2 &= \frac{\beta - D_1 - D_2}{2} + \frac{1}{2}\sqrt{\beta^2 - 2\beta D_1 + 2\beta D_2 + 4\alpha\ell\beta + D_1^2 - 2D_1 D_2 + D_2^2},\\
c_1 &= \frac{\beta - D_1 - \lambda_2}{\lambda_1 - \lambda_2}, c_2 = \frac{\beta - D_1 - \lambda_1}{\lambda_2 - \lambda_1},
\end{aligned}$$

with $D_1 = \alpha + \gamma_1 + \delta$, $D_2 = \gamma_2 + \delta$. Using (7.12), we obtain explicit expressions for the diagnosed recovered R_J and the diagnosed dead D_J, which are solutions of

$$R_J' = \gamma_2 J, \quad D_J' = \delta J, \quad R_J(t_{\text{in}}) = D_J(t_{\text{in}}) = 0. \tag{7.13}$$

Tentative average values for the rates α, γ_1, γ_2, δ, and for ℓ are usually available from clinical information, see Table 7.1 for the current pandemic in Spain. The remaining two parameters can be approximated optimizing the cost

$$f(\beta, t_{\text{in}}) = \frac{1}{2}\sum_{m=1}^{M} (\tilde{J}(\beta, m + t_{\text{in}}) - \tilde{j}_m)^2, \tag{7.14}$$

subject to the constraint

$$\tilde{J}' = \alpha I, \quad \tilde{J}(0) = 0, \tag{7.15}$$

Table 7.1: SEIJR model parameters. Guesses from clinical observation when available [25].

Par.	Definition	Guess
β	Transmission rate per day	
k	Rate of progression to the infectious state per day	
α	Rate of progression from infective to diagnosed per day	1/5–1/6 (stats)
γ_1	Rate at which infectious individuals recover per day	$\gamma_1^{-1} = \gamma_2^{-1} + \alpha^{-1}$
γ_2	Rate at which diagnosed individuals recover per day	1/10–1/11 (stats)
δ	covid-19 induced mortality per day	1/10–1/11 (stats)
ℓ	Relative measure of isolation of diagnosed cases	1/14 (practice)
q	Relative measure of infectiousness for the exposed	
p	Reduction in risk of covid-19 infection for class S_2	
t_{in}	Time at which local spread starts	
ρ	Fraction of the population at a lower risk	

with I given by (7.12). This cumulative function $\tilde{J}(t + t_{\text{in}}) = (J + R_J + D_J)(t)$ admits an explicit expression too. The data $\tilde{j}_m$, $m = 1, ..., M$, represent the cumulative numbers of diagnosed people during M days. This optimization problem can be solved employing a Levenberg-Marquardt-Fletcher algorithm [14]. Once a first guess for all this subset of parameters is available, we find a first guess for all the coefficients in the SEIJR model optimizing the cost

$$f(\beta, t_{\text{in}}, \alpha, \gamma_2, \delta, \ell, q, k, p) = \frac{1}{2} \sum_{m=1}^{M} (\tilde{J}(\beta, m + t_{\text{in}}, \alpha, \gamma_2, \delta, \ell, q, k, p) - \tilde{j}_m)^2, \tag{7.16}$$

with $\tilde{J}(t + t_{\text{in}}) = (J + R_J + D_J)(t)$, J, R_J, and D_J being solutions of (7.1) and (7.13). However, Levenberg-Marquardt-Fletcher optimization procedures may yield some negative coefficients. To avoid the technical complications associated to positivity restrictions, we can also use judicious choices for q, k, p instead.

7.3.3 Markov Chain Monte Carlo sampling

Once a prior has been defined, we sample the posterior distribution (7.10). Markov Chain Monte Carlo (MCMC) techniques have the potential of sampling unnormalized posterior distributions, as we discuss next. A Markov chain is a sequential stochastic process $X^0 \longrightarrow X^1 ... \longrightarrow X^n ...$, which evolves from one state to the next within an allowed set of states. To define a Markov chain we need:

- The set of states X the chain is allowed to reach, that is, the state space.

- A transition operator $q(X^{n+1}|X^n)$ defining the probability of moving from state X^n to state X^{n+1}.
- The initial distribution π_0 establishing the initial probability for the possible states.

Chains are assumed to be 'memoryless', that is, they depend only on the current state, and not on the previous path: $q(X^{n+1}|X^n, X^{n-1}, ..., X^0) = q(X^{n+1}|X^n)$. To generate a Markov chain, we first sample an initial state from π_0, and then move from one state to another as dictated by the transition operator $q(X^{n+1}|X^n)$.

There are many MCMC variants well adapted to sampling posterior distributions π, with different transition operators constructed from it [3]. Metropolis-Hastings algorithm, for instance, is a standard choice to sample an unnormalized distribution π. One drawback is that not all the proposed samples are accepted, wasting computational effort, specially in high dimensions. Hamiltonian Monte Carlo methods [12] reduce the correlation between consecutive sampled states by proposing moves to distant states which keep a high probability of acceptance. We will resort here to an algorithm that mixes several chains, allowing for parallelization and faster computation, which is also able to handle multimodal distributions. The Goodman-Weare [16] algorithm generates different chains for different walkers $X^1, \ldots, X^W$ which are mixed up. It proceeds as follows:

- Initialization: Generate the initial positions of the walkers $X_0^w \in \mathbb{R}^d$, $w = 1, ..., W$ sampling π_0. Set the distribution π to be sampled, the parameter a (typically 2), the number of samples N.
- Iteration: for $n = 0, ..., N-1$
  - for $w = 1, ..., W$
    - Draw a walker X_n^q at random from the set of walkers $\{X_n^j\}_{j \neq w}$.
    - Choose a random value z_w from the distribution $g(z)$ defined by

$$g(z) = \begin{cases} 1/\sqrt{z}, & \text{if} \quad z \in [1/a, a], \\ 0, & \text{otherwise.} \end{cases}$$

    - Calculate proposition $X_{\text{prop}}^w = X_n^q + z_w(X_n^w - X_n^q)$.
    - Calculate $s = z_w^{d-1} \frac{\pi(X_{\text{prop}}^w)}{\pi(X_n^w)}$.
    - Calculate $s = \text{Min}\,(1, s)$.
    - Draw r with uniform probability $\mathcal{U}(0, 1)$.
    - If $r \leq s$ set $X_{n+1}^w = X_{\text{prop}}^w$.
    - If $r > s$ set $X_{n+1}^w = X_n^w$.
- Final result: The Markov chains $\{X_0^w, ..., X_N^w\}$, for all the walkers $w = 1, ..., W$.

We set π_0 equal to the prior distribution (7.4) and π equal to the posterior distribution (7.10). By sampling this posterior distribution, we can find the Maximum a posteriori approximation

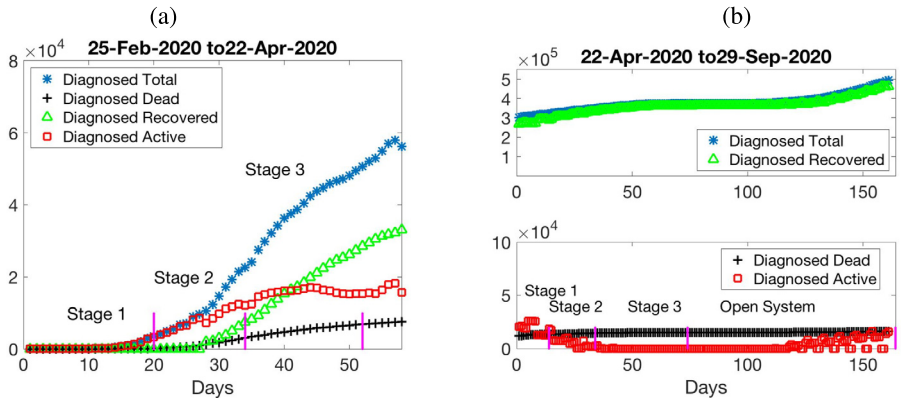

Figure 7.1: (a) Data from positive PCR tests until April 22th (only done to severe cases in hospitals). (b) Data from hospitals admittance and home surveillance from April 22th until September 29th. During the last period, no data were reported during the weekends (misplaced red squares at zero mark that fact).

(MAP) for the parameters, that is, the most likely coefficients for the SEIJR model. It is the sample with highest probability. We can also visualize the uncertainty in the choice of model coefficients for a given data set by means of histograms constructed with samples. The next section exemplifies the procedure with real data.

7.4 Effect of nonpharmaceutical actions

We illustrate the performance of the parameter identification procedure we have developed calculating coefficients for the SEIJR model with data for the current covid-19 pandemic from the Madrid region (Spain), available from [6] until April 22th, 2020, and, in more detail, from [7] after that date. Fig. 7.1 represents the data until the end of September 2020:

- Period 1 (lockdown): From Feb 25 until April 22, daily counts of diagnosed (positive PCR at hospital), dead (at hospital) and recovered (released from hospital) are available. Three stages can be distinguished according to contention measures implemented: until March 16 (free growth), until March 30 (basic lockdown), and until April 13 (severe lockdown).
- Period 2 (release): From April 22 until day July 2, daily counts were published for diagnosed (PCR or hospitalized or under surveillance at home), dead (at hospitals, care homes, and homes), and recovered (released from hospital, released from home surveillance) individuals. To quantify the diagnosed, we will discard the PCR positives which usually overlap with the rest and include repeated positives from the same individuals. We will consider hospital and home surveillance data. Three stages can be distinguished according to contention measures implemented: until May 4 (basic lockdown), until May 25

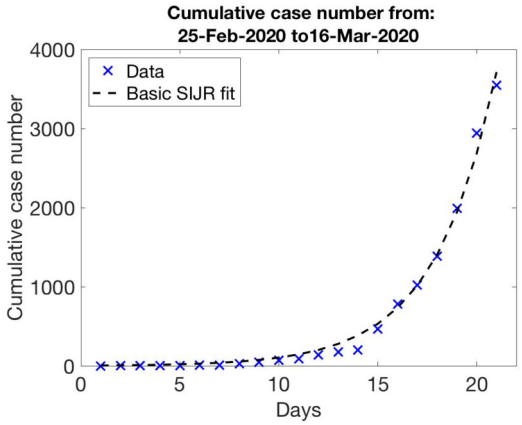

Figure 7.2: Cumulative number of cases versus fitting (7.15) for the parameter guess $t_{\text{in}} = 5.9$, $\beta = 0.6720$, $\alpha = 1/5$, $\gamma_1 = 1/15$, $\gamma_2 = 1/10$, $\delta = 1/10$, $\ell = 1/14$.

(masks enforced in closed spaces), and until July 2 (generalized use of masks and distancing).

Fig. 7.1(b) contains also data after July 2nd. After that date, airports, railways, and roads were opened to foreign visitors, the system was no longer closed. Before that date, there was some kind of mobility to and from other Spanish regions. However, most regions were undergoing a similar (or even better) evolution, so we will keep the assumption of a closed system in Periods 1 and 2.

To construct a prior, we first obtain an initial guess for the model parameters considering only the free growth Stage of Period 1. To do so, we optimize the constrained cost (7.14) fixing the parameter values in Table 7.1. Using a Levenberg-Marquardt-Fletcher algorithm, we find $\beta = 0.6720$, and $t_{\text{in}} = 5.9905$. The resulting fit for the cumulative case number $\tilde{J}$ associated to the SIJR approximation is depicted in Fig. 7.2. These are key parameters because they govern exponential growth in the initial stage.

Complementing these values with $p = 0$, $q = 1/2$, $k = 1/2$, we define the mean $\boldsymbol{\nu}_0$ for the prior (7.4). The diagonal covariance matrix $\mathbf{G}_{\text{pr}}$ has entries 0.1^2 for all the model coefficients and 10^2 for t_{in}. We consider standard deviations $\sigma_J = 10^3$, $\sigma_R = 10^3$, $\sigma_D = 10^2$ for the noise. With these choices, we can study (7.10) for the SEIJR model coefficients during the free growth stage. To consider data from different stages, we allow the model coefficients to change from stage to stage, that is, we consider piecewise coefficients.

7.4.1 Piecewise SEIJR system: lockdown

We use the two variables S_1 and S_2 to represent populations with different exposure to illness due to the enforcement of contention measures (distancing, confinement...), and split time in three consecutive intervals $[T_{i-1}, T_i]$, $i = 1, 2, 3$, according to the measures enforced. The resulting system of equations takes the form (7.1) with parameters

$$\begin{aligned}
\boldsymbol{\nu}^{(1)} &= (t_{\text{in}}, \beta^{(1)}, \gamma_2^{(1)}, \delta^{(1)}, \alpha^{(1)}, \ell^{(1)}, q^{(1)}, p^{(1)}, k^{(1)}), \quad t \in [T_0, T_1], \\
\boldsymbol{\nu}^{(2)} &= (\rho^{(2)}, \beta^{(2)}, \gamma_2^{(2)}, \delta^{(2)}, \alpha^{(2)}, \ell^{(2)}, q^{(2)}, p^{(2)}, k^{(2)}), \quad t \in [T_1, T_2], \\
\boldsymbol{\nu}^{(3)} &= (\rho^{(3)}, \beta^{(3)}, \gamma_2^{(3)}, \delta^{(3)}, \alpha^{(3)}, \ell^{(3)}, q^{(3)}, p^{(3)}, k^{(3)}), \quad t \in [T_2, T_3],
\end{aligned} \tag{7.17}$$

and $1/\gamma_1^{(i)} = 1/\gamma_2^{(i)} + 1/\alpha^{(i)}$, $i = 1, 2, 3$. Initial conditions for each period are:

- All unknowns vanish at $T_0 = t_{\text{in}}$ except $E(T_0) = 1$ and $S_1(T_0) = N - 1$, N being the total population.
- All unknowns are continuous at T_1 except for an abrupt jump in S_1 and S_2 due to contention measures enforced: $S_1(T_1^+) = (1 - \rho^{(1)})S_1(T_1^-)$ and $S_2(T_1^+) = S_2(T_1^-) + \rho^{(1)} S_1(T_1^-)$, $\rho^{(1)}$ represents the fraction of population confined in basic lockdown.
- All unknowns are continuous at T_2 except for an abrupt jump in S_1 and S_2 due to contention measures enforced: $S_1(T_2^+) = (1 - \rho^{(2)})S_1(T_2^-)$ and $S_2(T_2^+) = S_2(T_2^-) + \rho^{(2)} S_1(T_2^-)$, $\rho^{(2)}$ represents the additional fraction of population confined in severe lockdown.

The Bayesian formulation for this piecewise model uses a prior distribution (7.4) with $\boldsymbol{\nu} = (\boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}, \boldsymbol{\nu}^{(3)})$, number of parameters $n = 3 \times 9$, and a mean $\boldsymbol{\nu}_0$ constructed repeating the mean proposed for the first stage, $\boldsymbol{\nu}_0 = (\boldsymbol{\nu}_0^{(1)}, \boldsymbol{\nu}_0^{(2)}, \boldsymbol{\nu}_0^{(3)})$, with $\boldsymbol{\nu}_0^{(3)} = \boldsymbol{\nu}_0^{(2)} = \boldsymbol{\nu}_0^{(1)}$ except for changes at $\boldsymbol{\nu}_0^{(2)}(1) = 3/4$ and $\boldsymbol{\nu}_0^{(3)}(1) = 1/4$ to account for the fractions of populations confined at each stage. Accordingly, the diagonal of the covariance matrix $\mathbf{G}_{\text{pr}}$ has entries 0.1^2 for all the model parameters except 10^2 for t_{in}. The posterior distribution to be studied is then given by (7.10) with a similar covariance matrix for the noise. We keep $\sigma_J = 10^3$, $\sigma_R = 10^3$, $\sigma_D = 10^2$.

Sampling the posterior distribution by the MCMC algorithm in Section 7.3.3 we obtain the results represented in Table 7.2, after discarding the first $B = S/4$ samples as burn-in period. The coefficient values yielding the maximum probability are denoted by $\boldsymbol{\nu}_{\text{max}}$. The mean of all samples, which conveys a statistical meaning, is denoted by $\boldsymbol{\nu}_{\text{mean}}$. Notice the increase of the diagnose rate α and the decrease in the mortality rate δ from stage to stage, as well as the decrease in the transmission rates β and βp in the last stage. The increase in β during the second stage may be the outcome of large gatherings at the end of the previous phase, as well as lack of protective material during it. Panels (a)–(b) in Fig. 7.3 compare the SEIJR solutions

Table 7.2: Values of $\boldsymbol{\nu}_{\text{mean}}$ and $\boldsymbol{\nu}_{\text{max}}$ for three stages of the first period using the SEIJR model, with $\log(\boldsymbol{\nu}_{\text{mean}}) = -5734$, $\log(\boldsymbol{\nu}_{\text{max}}) = -124$, $\log(\boldsymbol{\nu}_0) = -21{,}431$, respectively. In the first row, the first columns represent t_{in}, while the rest correspond to ρ.

	$\nu^{(1)}_{\text{mean}}$	$\nu^{(2)}_{\text{mean}}$	$\nu^{(3)}_{\text{mean}}$	$\nu^{(1)}_{\text{max}}$	$\nu^{(2)}_{\text{max}}$	$\nu^{(3)}_{\text{max}}$
t_{in}, ρ	8.4233	0.7346	0.2231	5.3002	0.7203	0.1068
β	0.6570	0.7072	0.6439	0.6613	0.8231	0.6720
γ_1	0.0671	0.0600	0.0640	0.0860	0.0556	0.0575
γ_2	0.1279	0.0867	0.0889	0.1501	0.0674	0.0739
δ	0.1173	0.0475	0.0197	0.0351	0.0309	0.0123
α	0.1410	0.1949	0.2284	0.2014	0.3175	0.2596
ℓ	0.1031	0.1160	0.1195	0.0608	0.0760	0.1016
q	0.4909	0.5204	0.4760	0.5447	0.4704	0.3950
p		0.0902	0.0615		0.1395	0.0136
k	0.5121	0.4974	0.4888	0.3930	0.5016	0.3583

for $\boldsymbol{\nu}_{\text{max}}$ and $\boldsymbol{\nu}_{\text{mean}}$ with the recorded data. Panel (c) quantifies uncertainty in the magnitudes of the different populations at the end of the last stage. Panels (d)–(f) quantify uncertainty in the estimates for the model coefficients.

7.4.2 Piecewise SEIJR system: release

In the previous section, the initial values for the model at t_{in} were known. Also, the data used to fit coefficients keep the same quality and nature during that period. When the quality of the data suddenly improves, there is a gap between predictions before and after that time. If we restart the inference scheme to account for it, initial values are missing unless we infer them from the previous period. We obtain a new Bayesian formulation which reflects this fact including as parameters not the time t_{in} but the initial states at a fixed time for all the unknowns.

We will apply this strategy to switch from the data depicted in Fig. 7.1(a) to the data depicted in Fig. 7.1(b), ten times larger. We consider a set of parameters for the SEIJR model (7.1) in the interval $[\tau_0, \tau_1]$ constituted by the initial data at $\tau_0 = T_3$ plus the model coefficients

$$\boldsymbol{\nu} = (S_{1,0}, I_0, J_0, E_0, R_0, D_0, \rho, \beta, \gamma_2, \delta, \alpha, \ell, q, p, k), \tag{7.18}$$

with $1/\gamma_1 = 1/\gamma_2 + 1/\alpha$. The mean $\boldsymbol{\nu}_0$ for the prior is formed by the best approximation for the coefficients found in the final stage of the previous period and the final values of the variables at T_3. We consider the same variances for the coefficients as before, plus standard deviations 10^4 for the populations of alive subjects and 10^3 for the dead. We keep the same covariances as before for the data.

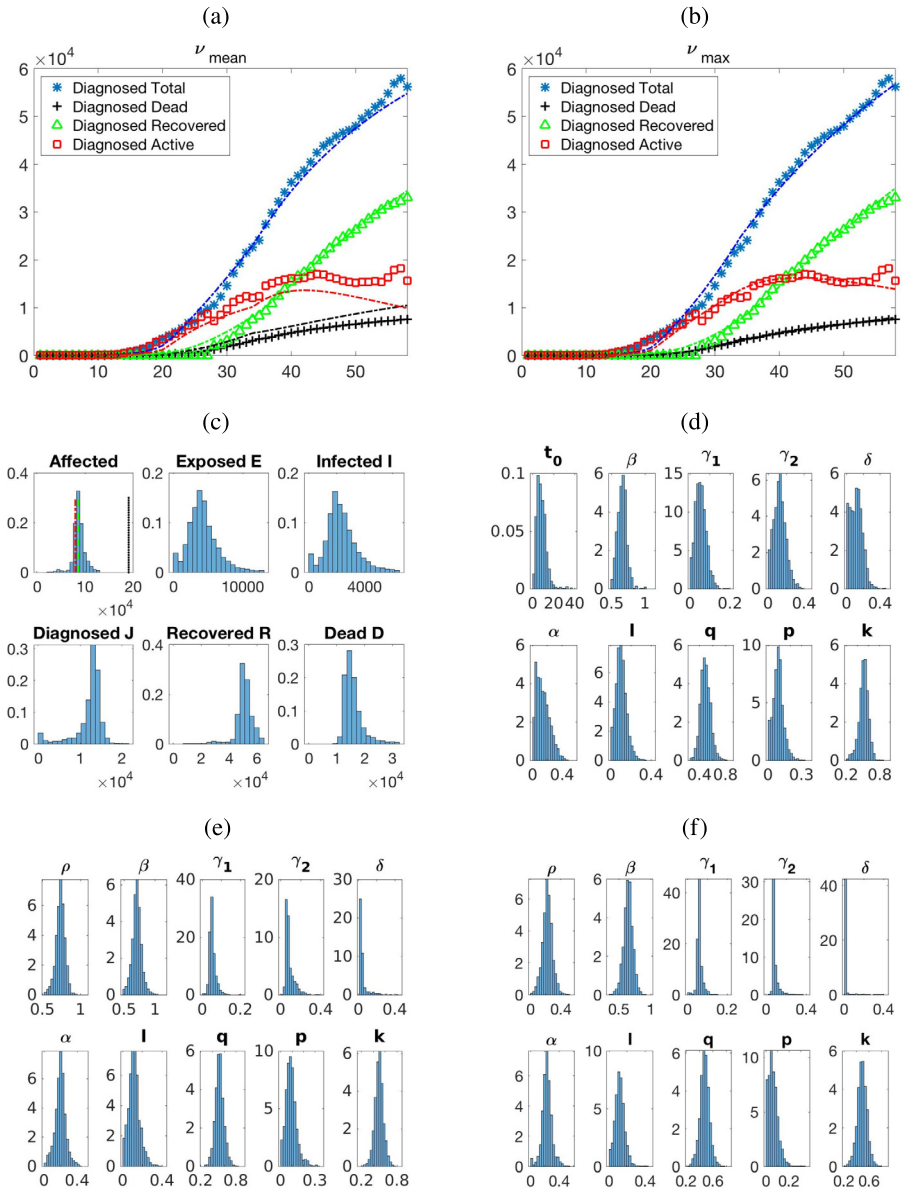

Figure 7.3: (a) Counts of diagnosed and dead cases for SEIJR solutions of (7.1) and (7.8) with $\boldsymbol{\nu}_{\text{mean}}$ (dashed lines), superimposed on data from Fig. 7.1(a). (b) Same for $\boldsymbol{\nu}_{\max}$. (c) Histograms representing probabilities for the total number of people affected by the virus at the end of the period, and for the different SEIJR categories. Superimposed lines represent the number of affected people for $\boldsymbol{\nu}_{\max}$ (red), $\boldsymbol{\nu}_{\text{mean}}$ (green), and the average of all samples (black). (d)–(f) Histograms representing a discrete approximation to the probability distribution of parameters during the three stages: (d) first, (e) second, (f) third. Sampling parameters $W = 500$, $S = 5 \times 10^6$, $B = S/4$, and acceptance parameter $a = 2$.

Table 7.3: Values of ν_{mean} and $\nu_{\max}$ for Stage 1 of the second period using the SEIJR model, with $\log(\nu_{\text{mean}}) = -12{,}980$, $\log(\nu_{\max}) = -2020$, $\log(\nu_{\text{in}}) = -406{,}460$, respectively. Populations are measured in a scale of 10^6 individuals.

	β	γ_1	γ_2	δ	α	ℓ	q	p	k
ν_{mean}	0.7646	0.2224	0.6878	0.0214	0.3287	0.2129	0.4565	0.2876	0.4455
$\nu_{\max}$	0.7081	0.1327	0.2931	0.0099	0.2425	0.1960	0.4387	0.2220	0.4060
	$S_{1,0}$	$S_{2,0}$	I_0	J_0	E_0	R_0	D_0		
ν_{mean}	0.0186	6.0576	0.0335	0.0998	0.0290	0.1219	0.0120		
$\nu_{\max}$	0.0137	5.9330	0.0009	0.0474	0.0282	0.2272	0.0120		

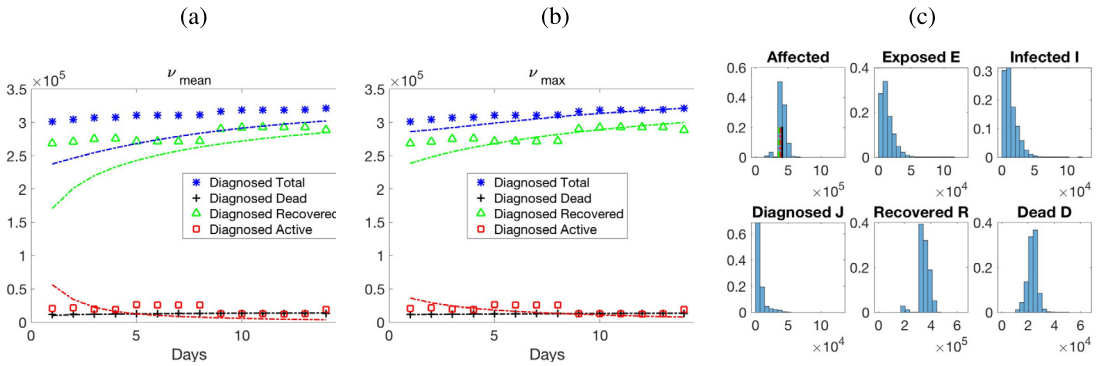

Figure 7.4: (a) Counts of diagnosed cases for SEIJR solutions of (7.1) and (7.8) with ν_{mean}, to be compared to Stage 1 in Fig. 7.1(b). (b) Same for $\nu_{\max}$. (c) Histograms representing probabilities for the total number of people affected by the virus at the end of the stage, and for the different SEIJR categories. Superimposed lines represent the number of affected people for $\nu_{\max}$ (red), ν_{mean} (green), and for the average of all samples (black). Sampling parameters $W = 100$, $S = 2 \times 10^6$, $B = S/4$, and acceptance parameter $a = 2$.

Sampling the resulting posterior distribution by the MCMC algorithm described in Section 7.3.3 we obtain the results represented in Table 7.3 for the first stage of data in Fig. 7.1(b). Again, the coefficient values yielding the maximum probability are denoted by $\nu_{\max}$, and the mean of all samples is denoted by ν_{mean}. Fig. 7.4 compares the SEIJR solutions for them with the recorded data. The agreement is worse than in Fig. 7.3 because we reduced the number of samples, it improves increasing the number of samples again.

We repeat the strategy for the next two stages adding a second block of coefficients. The resulting system of equations takes the form (7.1) with initial data at τ_1 and piecewise coef-

Table 7.4: Values of $\boldsymbol{\nu}_{\text{mean}}$ and $\boldsymbol{\nu}_{\text{max}}$ for the second and third stages of the second period using the SEIJR model, with $\log(\boldsymbol{\nu}_{\text{mean}}) = -626$, $\log(\boldsymbol{\nu}_{\text{max}}) = -398$, $\log(\boldsymbol{\nu}_{\text{in}}) = -32{,}489$, respectively. Populations are measured in a scale of 10^6 individuals.

	β	γ_1	γ_2	δ	α	ℓ	q	p	k
$\boldsymbol{\nu}^{(1)}_{\text{mean}}$	0.6925	0.0914	0.4028	0.0119	0.1183	0.1664	0.4245	0.1966	0.3405
$\boldsymbol{\nu}^{(1)}_{\text{max}}$	0.6514	0.0806	0.3668	0.0102	0.1032	0.0897	0.4202	0.1476	0.2591
$\boldsymbol{\nu}^{(2)}_{\text{mean}}$	0.7382	0.1753	0.5595	0.0151	0.2553	0.2196	0.4443	0.3036	0.4465
$\boldsymbol{\nu}^{(2)}_{\text{max}}$	0.6588	0.1909	0.8198	0.0182	0.2489	0.2956	0.5386	0.3683	0.4888
	$S_{1,0}$	$S_{2,0}$	I_0	J_0	E_0	R_0	D_0		
$\boldsymbol{\nu}_{\text{mean}}$	0.0171	5.8518	0.0048	0.0350	0.0206	0.2736	0.0135		
$\boldsymbol{\nu}_{\text{max}}$	0.0166	5.8371	0.0018	0.0274	0.0308	0.2801	0.0135		

ficients in $[\tau_1, \tau_2]$ and $[\tau_2, \tau_3]$. The total set of parameters is $\boldsymbol{\nu} = (\boldsymbol{\nu}^{(0)}, \boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)})$ with

$$\begin{aligned}
\boldsymbol{\nu}^{(0)} &= (S_{1,0}, I_0, J_0, E_0, R_0, D_0),\\
\boldsymbol{\nu}^{(1)} &= (\beta^{(1)}, \gamma_2^{(1)}, \delta^{(1)}, \alpha^{(1)}, \ell^{(1)}, q^{(1)}, p^{(1)}, k^{(1)}), \quad t \in [\tau_1, \tau_2],\\
\boldsymbol{\nu}^{(2)} &= (\rho^{(2)}, \beta^{(2)}, \gamma_2^{(2)}, \delta^{(2)}, \alpha^{(2)}, \ell^{(2)}, q^{(2)}, p^{(2)}, k^{(2)}), \quad t \in [\tau_2, \tau_3],
\end{aligned} \tag{7.19}$$

and $1/\gamma_1^{(i)} = 1/\gamma_2^{(i)} + 1/\alpha^{(i)}$, $i = 1, 2$, $S_{2,0} = N - (S_{1,0} + I_0 + J_0 + E_0 + R_0 + D_0)$. The mean $\boldsymbol{\nu}_0$ for the prior is formed by the best approximation for the coefficients found in the final stage of the previous period and the final values of the variables at τ_1. We consider the same variances for the coefficients and populations as before, as well as the same covariances as before for the data.

Sampling the resulting posterior distribution by the MCMC algorithm described in Section 7.3.3 we obtain the results represented in Table 7.4 for the second and third stages of data in Fig. 7.1(b). Again, the coefficient values yielding the maximum probability are denoted by $\boldsymbol{\nu}_{\text{max}}$, and the mean of all samples is denoted by $\boldsymbol{\nu}_{\text{mean}}$. Notice that most of the susceptible population enters the S_2 category, for which transmissibility is reduced by a factor p due to protective measures. Notice also that p and ℓ increase in the last stage, which suggests a generalized relaxation in protective and quarantine measures. This is consistent with the end of home confinement. However, the diagnose rate α and the recovery rates γ_1, γ_2 increase, and the epidemic remains stable. Fig. 7.5 compares the corresponding SEIJR solutions to data, and quantifies uncertainty in the final populations, as well as coefficients and initial populations.

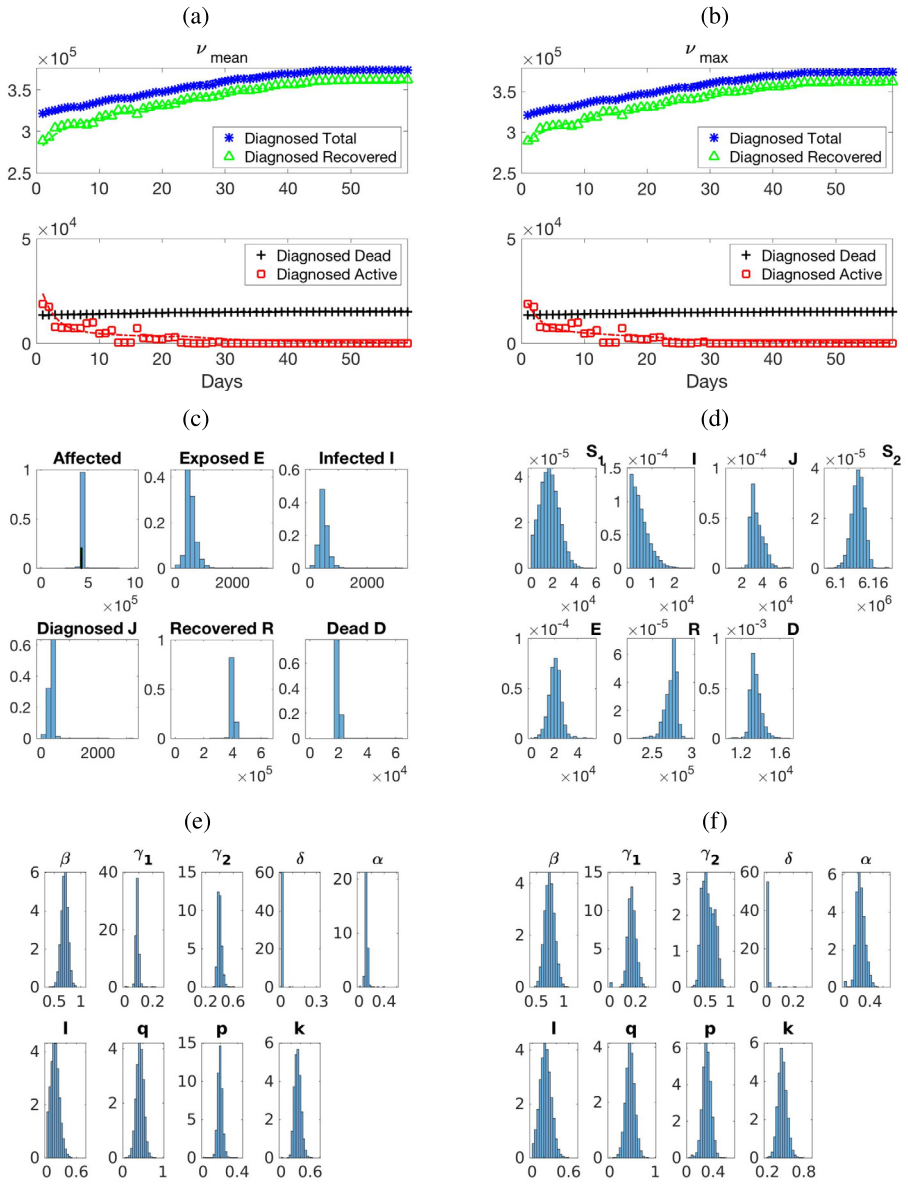

Figure 7.5: (a) Counts of diagnosed cases for SEIJR solutions of (7.1) and (7.8) with ν_{mean} (dashed lines), superimposed on data from Stages 2–3 in Fig. 7.1(b). (b) Same for $\nu_{\max}$. The differences between them are negligible. (c) Histograms of probabilities for the total number of people affected by the virus at the end of the period, and for the different SEIJR categories. The superimposed line represents the average of all samples. (d)–(f) Histograms for a discrete approximation to the probability distribution of parameters during the two stages: (e) first, (f) second, as well as the initial data for the period (d). Sampling parameters $W = 500$, $S = 5 \times 10^6$, $B = S/4$, and acceptance parameter $a = 2$.

7.4.3 Equilibrium

These studies show that nonpharmaceutical actions such as lockdowns, protective measures and distancing can drive a closed SEIJR system from free growth to extinction. The key issue is to reach a regime in which the right hand side in the equation for the exposed E in (7.1) is negative. As a consequence, the right hand sides in the equations for the infected I and diagnosed J will also become negative and the system will reach an equilibrium in which the susceptible, recovered and dead individuals remain constant. This was achieved acting on the parameters α, ℓ, p in (7.1) and also ρ in (7.17) and (7.19):

- Increasing α we increase the number of diagnosed infected individuals, and decrease the number of infected people. This can be done by extensive or well targeted testing policies.
- Decreasing ℓ, we reduce the chances of diagnosed individuals to propagate the infection. This can be achieved by strict quarantine measures.
- Splitting the susceptible population in such a way that a fraction ρ follows strict distancing measures, we reduce spread.
- Reducing p, we decrease the susceptibility of that part of the population. It can be the result of confining them at home or imposing the use of masks outside home.

When the exposed population E is small, increasing α and decreasing ℓ by fostering early detection of infected individuals and strict quarantine of those diagnosed can render negative the right hand side of the equation for I in (7.1). However, unless measures to decrease the number of exposed E are taken too, by introducing ρ and p, the system will remain out of control. Critical increase will be delayed, but not stopped. By the structure of Eq. (7.1), ρ and p are the key parameters to reach an extinction regime in the closed system.

7.5 SEIJR model including migration

As we have seen in the previous sections, an epidemic governed by a SEIJR model can be driven to equilibrium by contention measures in a closed system. When the system is no longer closed and migration is allowed, this goal is much harder to achieve.

Migration can be incorporated in the model considering different degrees of complexity. For instance, [22] considers several cities (or states, or countries) in interaction by introducing a SEIJR system for each of them and additional terms coupling all of them. Often, there are so many local, national, and international interactions at different levels that it is hard to consider all individually. Here, we will represent incoming and outcoming populations by means of

sink and source terms:

$$
\begin{aligned}
\frac{dS_1}{dt} &= -\beta S_1(t)\frac{I(t)+qE(t)+\ell J(t)}{N(t)} + \theta_1\mu_{in}r_{in}(t) - \eta_1\mu_{out}r_{out}(t),\\
\frac{dS_2}{dt} &= -\beta p S_2(t)\frac{I(t)+qE(t)+\ell J(t)}{N(t)} + \theta_2\mu_{in}r_{in}(t) - \eta_2\mu_{out}r_{out}(t),\\
\frac{dE}{dt} &= \beta(S_1(t)+pS_2(t))\frac{I(t)+qE(t)+\ell J(t)}{N(t)} - kE(t)\\
&\quad + \theta_3\mu_{in}r_{in}(t) - \eta_3\mu_{out}r_{out}(t),\\
\frac{dI}{dt} &= kE(t) - (\alpha+\gamma_1+\delta)I(t) + \theta_4\mu_{in}r_{in}(t) - \eta_4\mu_{out}r_{out}(t),\\
\frac{dJ}{dt} &= \alpha I(t) - (\gamma_2+\delta)J(t),\\
\frac{dR}{dt} &= \gamma_1 I(t) + \gamma_2 J(t) + \theta_5\mu_{in}r_{in}(t) - \eta_5\mu_{out}r_{out}(t),\\
\frac{dD}{dt} &= \delta I(t) + \delta J(t),
\end{aligned}
\tag{7.20}
$$

where $\theta_1+\theta_2+\theta_3+\theta_4+\theta_5 = 1$, $\eta_1+\eta_2+\eta_3+\eta_4+\eta_5 = 1$, $\theta_i, \eta_i \geq 0$, $i = 1,2,3,4,5$ and $N(t) = N_{total} + \mu_{in}\int_0^t r_{in} - \mu_{out}\int_0^t r_{out}$ is the total population. Here, $\mu_{in}r_{in}$, $\mu_{out}r_{out}$ represent the individuals entering and exiting the system. Multiplying by θ_i we get the fractions of susceptible, exposed, infected, and recovered subjects. If a definite functional form for the time evolution of incoming and outgoing population $r_{in}(t)$ and $r_{out}(t)$ was known we could apply the parameter identification framework we have developed to infer the missing parameters.

Assuming the coefficients and initial populations are such that the epidemic spread would be on the brink of extinction in a closed system, incoming infected and exposed individuals contribute the terms which can destabilize the system. The requirement of negative tests before admitting outsiders would set θ_4 to zero and $\theta_3 \sim 0$. As long as the contribution $\theta_3\mu_{in}r_{in}(t)$ does not make positive the right hand side of the equation for E in (7.20), the epidemic would remain controlled. The system will become unstable when either $\theta_3\mu_{in}r_{in}(t)$ is large enough to change that sign, or p is large enough to have the same effect due of a relaxation of contention measures, see Fig. 7.1(b).

7.6 Optimization approach to control

As we have already observed, to keep the systems (7.1) and (7.20) out of the exponential growth regime, we need a negative right hand side in the equation for E. We discuss here a possible control strategy. Let us assume that the coefficients and initial populations for the

(7.1) model are such that the epidemic in the closed system is extinguishing, as it happens for the coefficients and populations obtained in Section 7.4.2.

We wish to identify the reduction factor p representing the protective measures enforced on the population S_2 who obeys the rules. Therefore, we will consider the cost $C(p)$ given by

$$\text{Min}_{p\in[0,1]}\left\{\frac{1}{2}\sum_{i=1}^{M}\left[\left(\beta(S_1(t_i)+pS_2(t_i))\frac{I(t_i)+qE(t_i)+\ell J(t_i)}{N}-kE(t_i)\right)^+\right]^2+\frac{c}{2}p^2\right\}, \tag{7.21}$$

where $\frac{c}{2}p^2$, $c>0$, represents the cost of enforcing such measures, subject to the differential constraint (7.1). A similar cost can be used for the open system (7.20) including there source and sinks representing migration, as we comment later. We can approximate the solution of this optimization problem by Newton techniques [24], which requires the knowledge of the first and second order derivatives of the cost (7.21).

Let us denote $F(i,p)=\beta(S_1(t_i)+pS_2(t_i))\frac{I(t_i)+qE(t_i)+\ell J(t_i)}{N}-kE(t_i)$. Then

$$\frac{dC}{dp}=\sum_{i=1}^{N}F(i,p)^+F_p(i,p)+cp. \tag{7.22}$$

We can apply gradient methods to optimize or exploit the characterization of minima in dimension one:

$$\frac{dC(p)}{dp}=0,\ \frac{d^2C(p)}{d^2p}>0. \tag{7.23}$$

This equation can be solved by standard methods for nonlinear equations, such as Newton-Raphson schemes [24]

$$p^{n+1}=p^n-\left(\frac{d^2C(p^n)}{d^2p}\right)^{-1}\frac{dC(p^n)}{dp}. \tag{7.24}$$

These schemes involve the second order derivative

$$\frac{d^2C}{dp^2}=\sum_{i=1}^{N}[F^+(i,p)F_{p,p}(i,p)+\chi_{F>0}F_p(i,p)^2]+c \tag{7.25}$$

which fails to exist when $F(p)=0$, points at which, if encountered, the iteration should be modified switching to a gradient scheme. We can obtain all the required first and second order population derivatives with respect to p by simply differentiating twice the (7.1) system with respect to p and solving the resulting systems of differential equations. Setting

$$R(p)=\frac{I+qE+\ell J}{N},\ R_p(p)=\frac{I_p+qE_p+\ell J_p}{N},\ R_{pp}(p)=\frac{I_{pp}+qE_{pp}+\ell J_{pp}}{N},$$

we have

$$\begin{aligned}
\frac{dS_{1,p}}{dt} &= -\beta S_1(t)R_p(t) - \beta S_{1,p}(t)R(t),\\
\frac{dS_{2,p}}{dt} &= -\beta p S_2(t)R_p(t) - \beta p S_{2,p}(t)R(t) - \beta S_2(t)R(t),\\
\frac{dE_p}{dt} &= \beta(S_1(t) + pS_2(t))R_p(t) + \beta(S_{1,p}(t) + pS_{2,p}(t))R(t)\\
&\quad - kE_p(t) + \beta S_2(t)R(t) = F_p,\\
\frac{dI_p}{dt} &= kE_p(t) - (\alpha + \gamma_1 + \delta)I_p(t),\\
\frac{dJ_p}{dt} &= \alpha I_p(t) - (\gamma_2 + \delta)J_p(t),
\end{aligned} \tag{7.26}$$

$$\begin{aligned}
\frac{dS_{1,pp}}{dt} &= -\beta S_1(t)R_{pp}(t) - \beta S_{1,pp}(t)R(t) - 2\beta S_{1,p}(t)R_p(t),\\
\frac{dS_{2,pp}}{dt} &= -\beta p S_2(t)R_{pp}(t) - \beta p S_{2,pp}(t)R(t) - 2\beta p S_{2,p}(t)R_p(t)\\
&\quad - 2\beta S_{2,p}(t)R(t) - 2\beta S_2(t)R_p(t),\\
\frac{dE_{pp}}{dt} &= \beta(S_1(t) + pS_2(t))R_{pp}(t) + \beta(S_{1,pp}(t) + pS_{2,pp}(t))R(t) - kE_{pp}(t)\\
&\quad + 2\beta(S_{1,p}(t) + pS_{2,p}(t))R_p(t) + 2\beta S_{2,p}(t)R(t) + 2\beta S_2(t)R_p(t) = F_{pp},\\
\frac{dI_{pp}}{dt} &= kE_{pp}(t) - (\alpha + \gamma_1 + \delta)I_{pp}(t),\\
\frac{dJ_{pp}}{dt} &= \alpha I_{pp}(t) - (\gamma_2 + \delta)J_{pp}(t),
\end{aligned} \tag{7.27}$$

with zero initial data.

If the solution we find satisfies $F(p)^+ = 0$ in the end, our system remains out of the exponential growth region. Note that the additional term cp^2 suppresses a possible continuum of candidates. Nevertheless, values for c are difficult to estimate in practice.

When we are using initial data and coefficients obtained from Bayesian inference in a previous time period, we can quantify uncertainty in the range of possible p by solving the optimization problem for each sample of coefficients and initial data, and constructing histograms with the resulting solutions. However, some of the samples may not lead to $F(p)^+ = 0$ in the fixed time t_M, forcing an increase. This fact yields information on the uncertainty on the control time.

A similar strategy could be applied for the open system (7.20) modifying F to include the source:

$$F(\rho) = \beta(S_1(t) + pS_2(t))\frac{I(t) + qE(t) + \ell J(t)}{N(t)} - kE(t) + \theta_3[\mu_{in}r_{in}(t) - \mu_{out}r_{out}(t)],$$

and taking into account that the total population $N(t)$ is no longer constant.

7.7 Conclusions

We have developed a Bayesian inference framework to estimate unknown parameters in epidemiological models, like model coefficients or initial data, given population counts along the progression of the epidemic. This technique quantifies uncertainty in the fittings, providing parameter ranges to guide and focus analytical studies. We can obtain predictions for the final distribution of populations, including subjects difficult to track, such as asymptomatic individuals, and the total magnitude of the affected population. We have modified the model to include migration effects, discussing the potential role of key parameters to prevent exponential growth. Simple optimization formulations may allow us to establish control regimes, though how to estimate the cost of the contention measures necessary to implement them is an open question.

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 interships abroad.

References

[1] B. Ambikapathy, K. Krishnamurthy, Mathematical modelling to assess the impact of lockdown on Covid-19 transmission in India: model development and validation, JMIR Public Health Surveillance 6 (2020) e19368.
[2] R.M. Anderson, R.M. May, Population biology of infectious diseases: part I, Nature 280 (1979) 361–367.
[3] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[4] J.M. Brauner, S. Mindermann, M. Sharma, A.B. Stephenson, T. Gavenciak, D. Johnston, J. Salvatier, G. Leech, T. Besiroglu, G. Altman, H. Ge, V. Mikulik, M. Hartwick, Y.W. Teh, L. Chindelevitch, Y. Gal, J. Kulveit, The effectiveness and perceived burden of nonpharmaceutical interventions against Covid-19 transmission: a modelling study with 41 countries, medRxiv, https://dx.doi.org/10.1101/2020.05.28.20116129, 2020.
[5] A. Carpio, S. Iakunin, G. Stadler, Bayesian approach to inverse scattering with topological priors, Inverse Problems 36 (2020) 105001.
[6] Centro de Coordinación de Alertas y Emergencias Sanitarias, Enfermedad por el coronavirus (covid-19), Actualizaciones n^o 30 - n^o 135, Ministerio de Sanidad, Gobierno de España, 2020.

[7] Dirección General de Salud Pública - Servicio Madrileño de Salud y Hospitales privados, Datos Covid-19, Consejería de Sanidad, Comunidad de Madrid, 2020-21.
[8] G. Chowell, P.W. Fenimore, M.A. Castillo-Garsow, C. Castillo-Chavez, SARS outbreak in Ontario, Hong Kong and Singapore: the role of diagnosis and isolation as a control mechanism, Los Alamos Unclassified Report LA-UR-03-2653, 2003.
[9] J. Dehning, J. Zierenberg, F.P. Spitzner, M. Wibral, J.P. Neto, M. Wilczek, V. Priesemann, Inferring change points in the spread of Covid-19 reveals the effectiveness of interventions, Science 369 (2020) eabb9789.
[10] O. Diekmann, J.A.P. Heesterbeek, Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation, John Wiley and Sons, 2000.
[11] G. Ding, L. Chang, J. Gong, L. Wang, K. Cheng, D. Zhang, SARS epidemical forecast research in mathematical model, Chinese Science Bulletin 49 (2004) 2332–2338.
[12] S. Duane, A.D. Kennedy, B.J. Pendleton, R. Duncan, Hybrid Monte Carlo, Physics Letters B 195 (1987) 216–222.
[13] S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, M. Monod, Imperial College covid-19 Response Team, A.C. Ghani, C.A. Donnelly, S.M. Riley, M.A.C. Vollmer, N.M. Ferguson, L.C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe, Nature 584 (2020) 257–261.
[14] R. Fletcher, Modified Marquardt subroutine for non-linear least squares, Rpt. AERE-R 6799, Harwell, 1971.
[15] A. Huppert, G. Katriel, Mathematical modelling and prediction in infectious disease epidemiology, Clinical Microbiology and Infection 19 (2013) 999–1005.
[16] J. Goodman, J. Weare, Ensemble samplers with affine invariance, Communications in Applied Mathematics and Computational Science 5 (2010) 65–80.
[17] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Applied Mathematical Sciences, vol. 160, Springer, 2005.
[18] W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London Series A 115 (1927) 700–721.
[19] S.M. Kissler, C. Tedijanto, E. Goldstein, Y.H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period, Science 368 (2020) 860–868.
[20] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R.M. Eggo, Early dynamics of transmission and control of Covid-19: a mathematical modelling study, The Lancet 20 (2020) 553–558.
[21] J. Kuniya, Prediction of the epidemic peak of coronavirus disease in Japan, Journal of Clinical Medicine 9 (2020) 789.
[22] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2), Science 368 (2020) 489–493.
[23] K.Y. Ng, M.M. Gui, Covid-19: development of a robust mathematical model and simulation package with consideration for ageing population and time delay for control action and resusceptibility, Physica D. Nonlinear Phenomena 411 (2020) 132599.
[24] J. Nocedal, S. Wright, Numerical Optimization, Springer, New York, 1999.
[25] RENAVE, Análisis de los casos de Covid-19 notificados a la RENAVE hasta el 10 de mayo en España, Informe Covid-19 n^o 33, Instituto de Salud Carlos III, 2020.
[26] S. Ruschel, T. Pereira, S. Yanchuk, L.S. Young, An SIQ delay differential equations model for disease control via isolation, Journal of Mathematical Biology 79 (2019) 249–279.
[27] J.L. Sesterhenn, Adjoint-based data assimilation of an epidemiology model for the Covid-19 pandemic in 2020, arXiv preprint, arXiv:2003.13071, 2020.