Estimating the state of the Covid-19 epidemic in France using a non-Markovian model

Raphaël Forien[a], Guodong Pang[b], and Étienne Pardoux[c]

[a]INRAE, BioSP, Centre INRAE PACA, Domaine St-Paul, 84914 Avignon Cedex FRANCE, raphael.forien@inrae.fr

[b]The Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, College of Engineering, Pennsylvania State University, University Park, PA 16802 USA, gup3@psu.edu

[c]Aix–Marseille Université, CNRS, Centrale Marseille, I2M, UMR 7373 13453 Marseille, France, etienne.pardoux@univ.amu.fr

June 27, 2020

Abstract

In this paper, we use a deterministic non-Markovian epidemic model to estimate the state of the Covid-19 epidemic in France. This model allows us to consider realistic distributions for the exposed and infectious periods in a SEIR model, contrary to standard ODE models which only consider exponentially distributed exposed and infectious periods. We present theoretical results linking the (unobserved) parameters of the model to various quantities which are more easily measured during the early stages of an epidemic. We also stress the main quantitative differences between the non-Markovian and the Markovian (ODE) model. We then apply these results to estimate the state of the Covid-19 epidemic in France by analyzing three regions: the Paris region, the northeast regions and the rest of the country, based on current knowledge on the infection fatality ratio and the exposed and infectious periods distributions for Covid-19. Our analysis is based on the hospital data published daily by Santé Publique France (daily hospital admissions, intensive care unit admissions and hospital deaths).

Introduction

At the beginning of an epidemic outbreak, some quantities are easier to observe and report than others. For example the number of hospital deaths related to Covid-19 has been precisely and regularly reported in several countries. Another well documented quantity is the doubling time of the number of cases (which coincides with that of the number of deaths, as we shall explain). On the other hand, some quantities of interest are very hard to directly measure or to estimate: the actual number of infected individuals, the true death rate of the disease, and most notably the now famous reproduction number R_0, corresponding to the

average number of individuals infected by a single infectious individual in a fully susceptible population.

This paper presents a theoretical study of how the "observable" quantities relate to the "unobserved" ones, under a very general epidemic model recently developed in [PP20], and an application of these results to attempt to predict the state of the Covid-19 epidemic in France throughout its different stages.

The first basic idea is to estimate an exponential growth rate from the available data (the number of hospital admissions or the number of deaths occurring in hospitals), and then to relate it to the parameters of a model of the evolution of the Covid–19 epidemic. Our approach is modeled upon the approach of Tom Britton in two recent preprints [Bri20b], [Bri20a]. Note that the exponential growth rate can be negative, which was the case in particular in France during the lockdown period.

The exponential growth rate is the most natural parameter which can be extracted from the data. It equals log(2) divided by the doubling time, i.e. the number of days necessary for the number of cases (or deaths) to double, a notion that everyone listening to the news at the time of the rise of the epidemic has heard of, and can easily understand. The other important parameters of the model, including the famous basic reproduction number R_0 (the mean number of individuals whom an infectious individual infects before recovering, at the start of the epidemic – that is while essentially everyone that around him/her is susceptible), can be computed from the exponential growth rate, and some other parameters of the model.

Our approach has two specific features. First, we shall carefully make explicit which parameters are needed to compute the quantities of interest, and we shall describe as precisely as possible how the the uncertainty about their value influences our results. The second aspect is that we shall use nonconventional models. The classical SIR and SEIR ODE models are law of large numbers limits of stochastic Markov SIR or SEIR models [BP19]. For those models to be Markovian, it is necessary that the infectious periods in the case of the SIR model of the various individuals in the population be i.i.d. copies of an exponential random variable (resp. the pair exposed period, infectious period in the SEIR model be i.i.d. copies of a pair of independent exponential random variables). However, in the case of Covid–19, like in most infectious diseases, the assumption that those durations follow an exponential distribution is completely unrealistic.

Recently, the last two authors of this paper have described the law of large numbers limit model of non–Markovian stochastic epidemic models with arbitrary distribution for the infectious period (resp. for the pair exposed and infectious period), see [PP20]. This type of deterministic models is an integral equation of Volterra type, of the same dimension as the classical ODE model. In particular, it is not much more complicated to simulate and compute. As we shall see in this paper, our model fits the data better than Markovian models, especially around a major change in the contact pattern in the population (introduction of nationwide lockdown measures). The differences between our model and the usual "Markov" model are not so much in the large time behavior, but rather in the transient short term evolution, as was recently observed in [SRE+20], who study a related model. We note that since the start of the epidemic, knowledge about the Covid–19 disease has substantially increased. Our original approach allows us to choose distributions which reflect closely the current knowledge about the disease.

Note that at least one other work has used a similar extended model to analyze the

Covid–19 epidemic. In [ZFK20], the authors use ODEs with delays, which correspond to our model with deterministic exposed and infectious periods. On the other hand [GAA+20], following the approach initiated by Kermack and McKendrick to analyze the plague epidemic in Mumbai in 1905–6 [KM27], uses a transport PDE SEIR model, where the rate of infection by an infectious individual depends upon the time since infection, and the rate at which exposed (resp. infectious) individuals become infectious (resp. removed) also depends on the time since infection. The results of the present paper are comparable to those in [GAA+20], where the analysis and the treated data concern the various *départements* of Île de France, while we compare Île de France, the North–East of France, and the rest of the country.

After this paper had been written, we discovered that the non Markovian SEIR epidemic model which we are using is described in Section 4.5 of the book [BCCF], and our formula for R_0 is a particular case of a formula which appears on page 141 of that book.

The paper is organized as follows. In a first section called "Methods", we describe our model and the methodology to extract the parameters of our model from the available data. In a second section called "Results", we describe the results we obtained by applying our method to the Covid-19 epidemic in France. The last section is a discussion of the conclusions of our work. Finally, an appendix contains the mathematical proof of one crucial result, which relates the exponential growth rate to the various quantities in our model.

1 Methods

1.1 Our Covid–19 non–Markovian SEIR epidemic model

Assume that each newly infected individual in the population becomes infectious after a random time $\mathcal{E}$ during which he/she is "exposed" and stops being infectious after a random time $\mathcal{I}$, after which he or she does not infect anyone any more, and also cannot be infected (either as a result of acquired immunity, isolation or death). We assume that infectious individuals attempt to infect other individuals (chosen uniformly from the population) at rate $\lambda(t)$, where t is the current time (the dependence on t of the contact rate λ is used to reflect the effect of containment measures such as lockdown). As a result, the process of newly infected individuals up to time t can be represented as

$$P\left(N\int_0^t \lambda(s)\overline{S}^N(s)\overline{I}^N(s)ds\right),$$

where $P(t)$ is a standard Poisson process, N is the size of the population which is assumed to be fixed, $\overline{S}^N(t)$ (resp. $\overline{I}^N(t)$) is the proportion of susceptible (resp. infectious) individuals in the population at time t.

Assume that the initial proportions of susceptible, exposed, infectious and removed individuals in the population are given by $(\overline{S}(0), \overline{E}(0), \overline{I}(0), \overline{R}(0))$. We also assume that the individuals who are initially exposed or infectious have a different distribution of sojourn times in the intermediary states (reflecting the fact that they have been infected at some time in the past, before the initial time $t = 0$). Thus let $(\mathcal{E}_0, \mathcal{I}_0)$ denote a random variable distributed as the time an individual who is initially exposed stays exposed ($\mathcal{E}_0$) and then infectious ($\mathcal{I}_0$), and let $\mathcal{I}_1$ be distributed as the time an individual who is initially infectious

stays infectious. Note that usually $\mathcal{I}_0$ and $\mathcal{I}$ have the same distribution for those who are initially exposed, while the remaining infectious period $\mathcal{I}_1$ may have a different distribution. Similarly, the remaining exposed period $\mathcal{E}_0$ may have a different distribution from that of $\mathcal{E}$.

Define

$$G^c(t) = \mathbb{P}(t < \mathcal{E}), \qquad \Psi(t) = \mathbb{P}(\mathcal{E} \leq t < \mathcal{E} + \mathcal{I}), \qquad \Phi(t) = \mathbb{P}(\mathcal{E} + \mathcal{I} \leq t),$$

as well as

$$G_0^c(t) = \mathbb{P}(t < \mathcal{E}_0), \qquad \Psi_0(t) = \mathbb{P}(\mathcal{E}_0 \leq t < \mathcal{E}_0 + \mathcal{I}_0), \qquad \Phi_0(t) = \mathbb{P}(\mathcal{E}_0 + \mathcal{I}_0 \leq t),$$

and

$$F_1(t) = \mathbb{P}(\mathcal{I}_1 \leq t), \qquad F_1^c(t) = 1 - F_1(t) = \mathbb{P}(t < \mathcal{I}_1).$$

Let $(\overline{S}^N(t), \overline{E}^N(t), \overline{I}^N(t), \overline{R}^N(t))$ denote the proportion of susceptible, exposed, infectious and removed individuals at time t in the population, where N is the total size of the population, which is assumed to be fixed throughout the epidemic (the deaths due to the epidemic are counted among the "removed"). We assume that the durations $(\mathcal{E}_i, \mathcal{I}_i)$ of all the infected individuals infected during the epidemic are i.i.d. (for each i, $\mathcal{E}_i$ and $\mathcal{I}_i$ may be correlated).

Definition 1. *The non-Markovian SEIR deterministic model is the solution of the set of integral equations:*

$$\begin{aligned}
\overline{S}(t) &= \overline{S}(0) - \int_0^t \lambda(s)\overline{S}(s)\overline{I}(s)ds, \\
\overline{E}(t) &= \overline{E}(0)G_0^c(t) + \int_0^t \lambda(s)G^c(t-s)\overline{S}(s)\overline{I}(s)ds, \\
\overline{I}(t) &= \overline{I}(0)F_1^c(t) + \overline{E}(0)\Psi_0(t) + \int_0^t \lambda(s)\Psi(t-s)\overline{S}(s)\overline{I}(s)ds, \\
\overline{R}(t) &= \overline{I}(0)F_1(t) + \overline{E}(0)\Phi_0(t) + \int_0^t \lambda(s)\Phi(t-s)\overline{S}(s)\overline{I}(s)ds.
\end{aligned}$$

Theorem 3.1 from [PP20] states that, under a very weak assumption on the joint distribution of $(\mathcal{E}, \mathcal{I})$, the unique solution of the above non-Markovian SEIR integral equations is the law of large numbers limit of the above described model. More precisely, as $N \to \infty$, $(\overline{S}^N(t), \overline{E}^N(t), \overline{I}^N(t), \overline{R}^N(t)) \to (\overline{S}(t), \overline{E}(t), \overline{I}(t), \overline{R}(t))$ in probability, locally uniformly in t. We call this model non–Markovian since the stochastic process $(\overline{S}^N(t), \overline{E}^N(t), \overline{I}^N(t), \overline{R}^N(t))$ is not a Markov process in general, because its future evolution depends not only upon its present value, but also upon how long the exposed and infectious individuals have already been in the corresponding compartment. For the model to be Markovian, it is necessary that the random variables $\mathcal{E}$ and $\mathcal{I}$ be independent and have exponential distributions. In that case, the integral equation SEIR model reduces to the following ODE model

$$\frac{d\overline{S}(t)}{dt} = -\lambda(t)\overline{S}(t)\overline{I}(t), \qquad \frac{d\overline{E}(t)}{dt} = \lambda(t)\overline{S}(t)\overline{I}(t) - \nu\overline{E}(t),$$

$$\frac{d\overline{I}(t)}{dt} = \nu\overline{E}(t) - \gamma\overline{I}(t), \qquad \frac{d\overline{R}(t)}{dt} = \gamma\overline{I}(t),$$

where ν (resp. γ) is the parameter of the exponential law of $\mathcal{E}$ (resp. $\mathcal{I}$), i.e., $\mathbb{E}(\mathcal{E}) = \nu^{-1}$, $\mathbb{E}(\mathcal{I}) = \gamma^{-1}$. Note that in the Markovian case, the convergence $(\overline{S}^N(t), \overline{E}^N(t), \overline{I}^N(t), \overline{R}^N(t)) \to (\overline{S}(t), \overline{E}(t), \overline{I}(t), \overline{R}(t))$ as $N \to \infty$ is a classical result, see, e.g., [BP19] for a recent account. The above ODE model is the one which is used by almost all papers dealing with epidemic models, in particular the Covid–19 models, with the exception, to our knowledge, of [ZFK20], where the case of fixed $\mathcal{E}$ and $\mathcal{I}$ is considered and [GAA+20, SRE+20]. It is fair to say that the exponential distributions are very poor models for the laws of $\mathcal{E}$ and $\mathcal{I}$, and it may seem strange that the above ODE model is so widely used, while it is based on unrealistic assumptions. While the large time behavior of the model (e.g., the endemic equilibrium in the case of a SIS or a SIRS model, see Remarks 2.8 and 3.5 in [PP20]) depend only on the expectation of the random vector $(\mathcal{E}, \mathcal{I})$, and are the same in a Markov and a non-Markov model with identical expected infectious periods, as we shall see below, the transient behavior of the model depends much more on the details of the law of $(\mathcal{E}, \mathcal{I})$.

Let us now explain the specific model (with two variants) which we will use for the Covid–19 epidemic. In our model, the two random variables $\mathcal{E}$ and $\mathcal{I}$ will be independent. The random variable $\mathcal{E}$ will be either fixed (3 days), or else will have a distribution with a density whose support is the interval $[2, 4]$. Concerning the random variable $\mathcal{I}$, we assume a bimodal law with support in $[3, 5] \cup [8, 12]$ (details in Subsection 1.3).

The idea behind that type of law is as follows. Our model is close to the SEIRU model of [LMSW20], which thanks to the flexibility of our class of models we are able to simplify. Indeed, in that paper the individuals are first exposed (state E), then infectious (state I) but without symptoms, and at the end of the I period, a fraction of the individuals are isolated quickly after the onset of symptoms (state R as "reported"), either because they are admitted to hospital or because they self-isolate at home, while the other individuals are not isolated (state U as "unreported"), either because they are unable to do so or because they have light or no symptoms. While in the R class, the individuals are isolated and they do not infect susceptibles any more. For that reason, we identify the "reported" class with part of the "recovered" class, thus reconciling the two meanings of R. Finally, some of the individuals are infectious while they are in the I class only, while others are infectious in both I and U classes. In our model, being in the U class is considered as staying longer in the I class (see Figure 1). This is the motivation for the bimodal distribution of the law of $\mathcal{I}$.

Remark 2. *Our model lacks one important feature which is widely believed to affect the dynamics of the Covid-19 epidemic, namely age structure. Our main motivation for neglecting age structure is to keep the model mathematically tractable and to reduce the number of parameters. Needless to say, more realistic predictions could be obtained by using a structured version of this model, provided all the parameters can be correctly estimated.*

1.2 The non–Markovian model during the early phase of the epidemic

During the early phase of the epidemic, the cumulative number of infected individuals remains small compared to the total size of the population. As a result, $\overline{S}(t) \approx 1$ during this phase. Letting $(E(t), I(t), R(t))$ denote the absolute numbers of exposed, infectious and

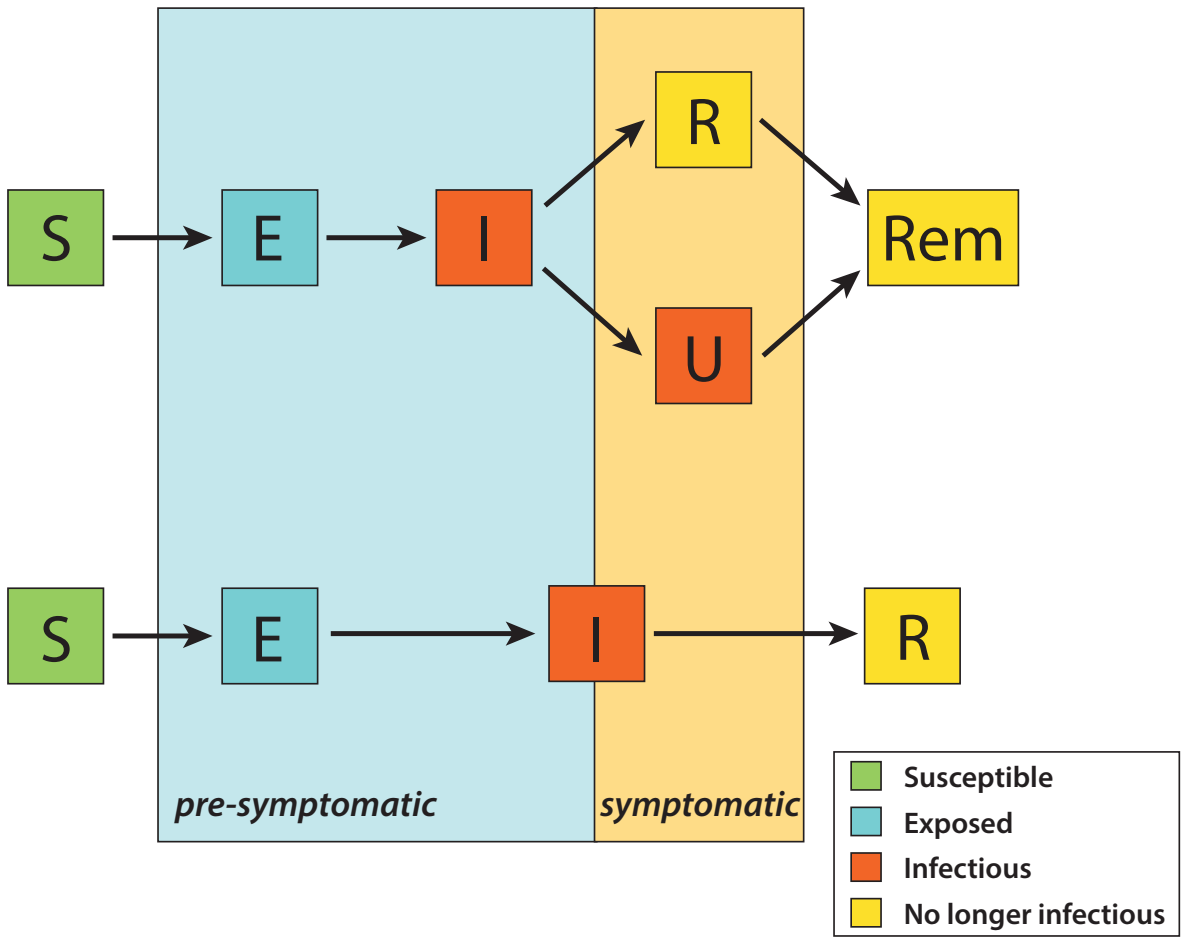

Figure 1: Flow chart of the SEIRU model of [LMSW20] and the equivalent SEIR non-Markovian model. In the latter, the two infectious compartments (I and U) are merged into one compartment, but the sojourn time $\mathcal{I}$ of the infectious compartment has a bimodal distribution corresponding to the two subpopulations of reported and unreported individuals. Note that R (reported) and Rem (removed) have been merged into a unique compartment.

removed individuals during this phase, *i.e.*,

$$(E(t), I(t), R(t)) = (N\overline{E}(t), N\overline{I}(t), N\overline{R}(t)),$$

and assuming that λ is constant during this phase, the non-Markovian SEIR model reduces to

$$\begin{aligned} E(t) &= E(0)G_0^c(t) + \lambda \int_0^t G^c(t-s)I(s)ds, \\ I(t) &= I(0)F_1^c(t) + E(0)\Psi_0(t) + \lambda \int_0^t \Psi(t-s)I(s)ds, \\ R(t) &= R(0) + I(0)F_1(t) + E(0)\Phi_0(t) + \lambda \int_0^t \Phi(t-s)I(s)ds. \end{aligned} \tag{1}$$

The initial state of an epidemic is seldom known, and the Covid-19 epidemic is no exception. However, it is always the case that, once sufficiently many individuals have been infected, the cumulative number of infected individuals grows exponentially with some rate $\rho > 0$. We thus look for solutions to

$$\begin{aligned} E(t) &= \lambda \int_{-\infty}^{t} G^c(t-s) I(s) ds, \\ I(t) &= \lambda \int_{-\infty}^{t} \Psi(t-s) I(s) ds, \\ R(t) &= \lambda \int_{-\infty}^{t} \Phi(t-s) I(s) ds, \end{aligned} \tag{2}$$

for $t \in \mathbb{R}$ which are of the form

$$E(t) = \boldsymbol{e}\, e^{\rho t}, \qquad I(t) = \boldsymbol{i}\, e^{\rho t}, \qquad R(t) = \boldsymbol{r}\, e^{\rho t}, \tag{3}$$

with $\boldsymbol{e} + \boldsymbol{i} + \boldsymbol{r} = 1$. These equations started from $-\infty$ can be seen as modeling an epidemic which has been growing for a very long time from an infinitesimal proportion of non-susceptible individuals.

Proposition 3. *(i) If $\rho > 0$, the system (2) admits solutions of the form (3) for all $t \in \mathbb{R}$ if*

$$\lambda = \frac{\rho}{\mathbb{E}\left[e^{-\rho \mathcal{E}}(1 - e^{-\rho \mathcal{I}})\right]} \tag{4}$$

and with

$$\boldsymbol{e} = \mathbb{E}\left[1 - e^{-\rho \mathcal{E}}\right], \qquad \boldsymbol{i} = \mathbb{E}\left[e^{-\rho \mathcal{E}}(1 - e^{-\rho \mathcal{I}})\right], \qquad \boldsymbol{r} = \mathbb{E}\left[e^{-\rho(\mathcal{E}+\mathcal{I})}\right]. \tag{5}$$

Moreover, if Θ is an independent exponential variable with parameter ρ, then (3) also solves (1) for all $t \geq 0$ with

$$\begin{aligned} G_0^c(t) &= \mathbb{P}(\mathcal{E} > t + \Theta \,|\, \Theta < \mathcal{E}), & \Psi_0(t) &= \mathbb{P}(\mathcal{E} \leq t + \Theta < \mathcal{E} + \mathcal{I} \,|\, \Theta < \mathcal{E}), \\ \Phi_0(t) &= \mathbb{P}(t + \Theta \geq \mathcal{E} + \mathcal{I} \,|\, \Theta < \mathcal{E}), & F_1(t) &= \mathbb{P}(t + \Theta \geq \mathcal{E} + \mathcal{I} \,|\, \mathcal{E} \leq \Theta < \mathcal{E} + \mathcal{I}). \end{aligned} \tag{6}$$

(ii) If $\rho < 0$, then $(E(t), I(t)) = (\mathfrak{e} e^{\rho t}, \mathfrak{i} e^{\rho t})$ solves the first two lines of (2) for all $t \in \mathbb{R}$ if λ and ρ satisfy (4) and if

$$\mathfrak{e} = \mathbb{E}\left[e^{-\rho \mathcal{E}} - 1\right], \qquad \mathfrak{i} = \mathbb{E}\left[e^{-\rho \mathcal{E}}(e^{-\rho \mathcal{I}} - 1)\right].$$

The fact that a solution of the form $R(t) = \boldsymbol{r}\, e^{\rho t}$ only exists for positive ρ should not come as a surprise, since $t \mapsto R(t)$ is non-decreasing. Also note that we can rewrite (5) with the help of the variable Θ as follows,

$$\boldsymbol{e} = \mathbb{P}(\Theta < \mathcal{E}), \qquad \boldsymbol{i} = \mathbb{P}(\mathcal{E} \leq \Theta < \mathcal{E} + \mathcal{I}), \qquad \boldsymbol{r} = \mathbb{P}(\Theta \geq \mathcal{E} + \mathcal{I}). \tag{7}$$

The proof of Proposition 3 is postponed to Appendix A.

Corollary 4. *The basic reproduction number R_0, defined as the mean number of secondary infections caused by a single infectious individual in a fully susceptible population, is linked to the initial growth rate of the cumulative number of infected individuals ρ by the relation*

$$R_0 = \lambda\,\mathbb{E}[\mathcal{I}] = \frac{\rho\,\mathbb{E}[\mathcal{I}]}{\mathbb{E}\left[e^{-\rho\mathcal{E}}(1-e^{-\rho\mathcal{I}})\right]}.$$

This formula is valid in both cases $\rho > 0$ and $\rho < 0$.

Remark 5. *Note that if $(\mathcal{E},\mathcal{I})$ are two independent exponential random variables with parameters $\nu > 0$ and $\gamma > 0$, the equations of Definition 1 coincide with the law of large numbers limit of the Markovian SEIR epidemic model [BP19]. In this case, Proposition 3 agrees with what is already known about Markovian epidemic models. In particular, in the case $\rho > 0$,*

$$\lambda = (\rho+\gamma)\left(1+\frac{\rho}{\nu}\right), \qquad \boldsymbol{e} = \frac{\rho}{\nu+\rho}, \qquad \boldsymbol{i} = \frac{\rho\nu}{(\rho+\gamma)\,(\nu+\rho)}, \qquad \boldsymbol{r} = \frac{\nu\gamma}{(\rho+\nu)(\rho+\gamma)}.$$

The equivalent formulas for the SIR model (i.e., with $\mathcal{E} = 0$) can be obtained by letting $\nu \to \infty$, in which case we see that $\rho = \gamma(R_0 - 1)$, as in formula (1) in [Bri20b].

If, however, we assume that the variables $\mathcal{E}$ and $\mathcal{I}$ are constant and equal to (t_e, t_i), the equations of Definition 1 can be seen as delay equations. In this case, Proposition 3 still applies, and the expectations in (4) and (5) can be omitted, leading to the relation

$$R_0 = \frac{\rho\, t_i}{e^{-\rho t_e}(1-e^{-\rho t_i})}.$$

As a result, we see that, for the same growth rate ρ, the contact rate and the relative proportions of exposed, infectious and removed individuals vary depending on the distribution of the sojourn times $(\mathcal{E},\mathcal{I})$. In particular, the growth rate ρ cannot be used to accurately determine the value of $R_0 = \lambda\,\mathbb{E}[\mathcal{I}]$ if one lacks information about the distribution of $(\mathcal{E},\mathcal{I})$ (see Figure 2).

1.3 Distribution of the sojourn times for Covid-19

The evolution of infectiousness for Covid-19 from the time of infection remains uncertain, but some early studies already provide constraints on the distribution of the sojourn times $(\mathcal{E},\mathcal{I})$ for this disease. In particular, [HLW+20] estimate that infectiousness starts as early as 2.3 days before the onset of symptoms and declines within 7-8 days of symptom onset. Assuming that symptoms start on average 5.2 days after infection (as reported in [HLW+20]), we conclude that $\mathcal{E}$ is between 2 and 4 days, and that $\mathcal{I}$ is at least 3 days (if the infected individual is isolated shortly after the onset of symptoms) and no more than 10 to 12 days.

For the exposed time $\mathcal{E}$, we shall take either a fixed value of 3 days or a linear combination of the form $2 + 2\,X_1$, where X_1 is a symmetric Beta random variable with support in $[0,1]$.

In this study, we assume that infected individuals are divided in two groups. Individuals in the first group, called the *reported* individuals, are isolated shortly after the onset of symptoms, either because they self-isolate at home or because they are admitted to a hospital,

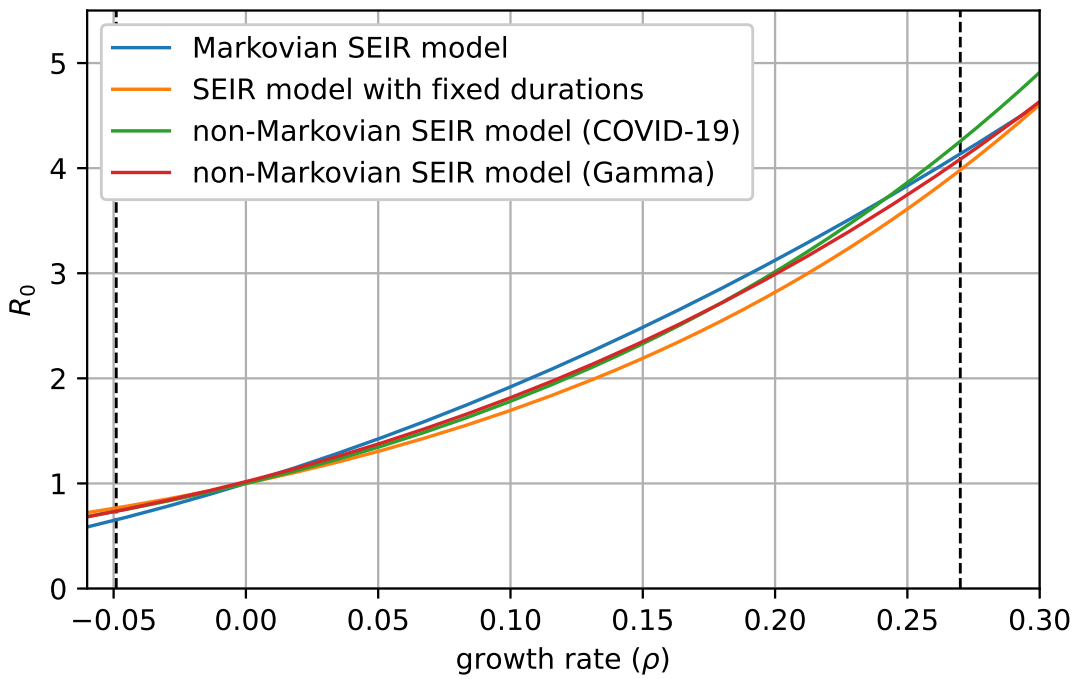

Figure 2: Value of R_0 as a function of the growth rate ρ for different distributions of the exposed and infectious periods $(\mathcal{E}, \mathcal{I})$. Four types of distributions are displayed: exponential (corresponding to the Markovian SEIR model), fixed, bimodal distribution mimicking Covid-19 (see Subsection 1.3) and Gamma distribution. All distributions have a mean exposed time of 3 days and a mean infectious time of 4.8 days, corresponding to a proportion of reported individuals of 0.8. The two dashed vertical lines show the growth rates of the Covid-19 epidemic in mainland France before and during lockdown.

where we assume that they do not infect anyone else. For these individuals, we assume that $\mathcal{I} = 3 + X_2$, where X_2 is a symmetric Beta random variable with support in $[0, 1]$.

By contrast, individuals in the second group remain infectious for much longer, either because they show very mild symptoms or no symptoms at all or because they fail to be isolated. For these *unreported* individuals, we assume that $\mathcal{I} = 8 + 4\,X_3$, where X_3 is a symmetric Beta random variable with support in $[0, 1]$.

Let p_R be the proportion of reported individuals and let Y be a Bernoulli random variable with parameter p_R, independent of X_2 and X_3. Then we can write

$$\mathcal{I} = Y(3 + X_2) + (1 - Y)(8 + 4\,X_3).$$

We thus see that the distribution of $\mathcal{I}$ is bimodal, with a first peak at around 3.5 days and a second one around 10 days.

It seems that asymptomatic individuals are less infectious than symptomatic ones (see [CSC+20]). We could have included in the model different values of infectiosity depending upon the duration of the infectious period of each individual, using the theory developed in [FPP20], but we lack quantitative information about the various levels of infectiosity

to make serious predictions, while the available data does not allow us to estimate many parameters. Note also that, as shown in [FPP20], during the early phase of the epidemic, the product of the number of asymptomatic patients and their infectiosity determines the evolution of the epidemic. Overestimating the infectiosity leads to an underestimation of the number of cases, which has no significant impact on the dynamic of the epidemic during the early phase, but may lead to underestimation of the proportion of the population which has acquired immunity at the end of this early phase.

1.4 Estimating the state of the epidemic prior to lockdown measures

In France, as of June 2020, the Covid-19 epidemic underwent three main stages. First, a rapid exponential growth (of the number of cases, number of hospitalized patients, number of deaths), followed by a slowing down of the epidemic due to the lockdown measures put into place around 16 March, settling after a few weeks to an exponential decrease of the number of newly hospitalized patients and new deaths. The third phase started on 11 May, with the progressive lifting of lockdown restrictions.

Since testing has been limited to a subset of symptomatic individuals and tests were performed at various intervals following symptom onset, the number of reported positive cases might not be exactly proportional to the true number of infected individuals throughout the different stages of the epidemic (and more importantly the ratio of infected individuals to tested individuals may vary significantly between the different phases). Hospital deaths, however, have been reported daily from 15 February. Moreover, from 18 March onwards, Santé Publique France has published daily reports of newly hospitalised patients, newly admissions in intensive care units (ICU) and new deaths in each administrative *département*, [San20].

Assuming that the distributions of the intervals between infection and hospital admission, admission in ICU and death do not vary over time, the growth rate of these quantities is necessarily identical to that of the cumulative number of infected individuals. As a result, if the distribution of the exposed and infectious periods $(\mathcal{E}, \mathcal{I})$ is known, we can infer the contact rate λ corresponding to a given growth rate ρ using (4).

Note that, since Santé Publique France only started publishing regional data on 18 March, the growth rate is inferred from the slope of the cumulative deaths during the first week of lockdown. Since the delay between infection and death in hospital is believed to be at least 10 days (*e.g.* [SKL+20]), it is safe to assume that all the individuals who died during this period were infected before lockdown measures took effect (however, see Subsection 2.1 below for the particular case of the Grand Est and Hauts-de-France regions).

If moreover one knows the distribution of the interval $\mathcal{D}$ between infection and death, for example, as well as the infection fatality ratio f (the probability that an individual eventually dies, given that he or she has been infected), it is possible to infer the state of the epidemic. Indeed, the cumulative number of deaths at time $t \geq 0$, denoted by $\Lambda_D(t)$, is then given by

$$\Lambda_D(t) = f N \, \mathbb{E}\left[1 - \overline{S}(t - \mathcal{D})\right],$$

where N is the total population size (the expectation is taken with respect to the random

variable $\mathcal{D}$). During the early phase of the epidemic,

$$(1-\overline{S}(t)) \approx \frac{1}{N}(E(t)+I(t)+R(t)) = \frac{1}{N}e^{\rho t},$$

where t is the time elapsed since the (unknown) time of the start of the epidemic (*i.e.,* the theoretical time at which only one individual was infected in the population), and $E(t)$, $I(t)$ and $R(t)$ solve the linearized system (1). Thus, during this phase,

$$\Lambda_D(t) = f\,\mathbb{E}\left[\mathbf{1}_{\{\mathcal{D}\leq t\}}\,e^{-\rho\mathcal{D}}\right]e^{\rho t}. \tag{8}$$

Hence, if one knows the infection fatality ratio and the distribution of $\mathcal{D}$, one can infer both ρ and t from the observed cumulative number of deaths prior to lockdown measures. Using (4), one can deduce the value of the contact rate before lockdown measures, λ_0, (given that the distribution of $(\mathcal{E},\mathcal{I})$ is known) and thus simulate the epidemic up to this time.

Note that if p_H is the probability that a given infected individual is admitted to a hospital at some point, and if $\mathcal{D}_H$ is the interval between the time of infection and hospital admission, then, the cumulative number of hospitalized patients at time $t \geq 0$, denoted by $\Lambda_H(t)$, is given by

$$\Lambda_H(t) = p_H N\mathbb{E}\left[1-\overline{S}(t-\mathcal{D}_H)\right].$$

Hence, during the early phase of the epidemic,

$$\Lambda_H(t) = p_H\mathbb{E}\left[\mathbf{1}_{\{\mathcal{D}_H\leq t\}}e^{-\rho\mathcal{D}_H}\right]e^{\rho t}.$$

As a result, one can obtain the probability of being admitted to hospital if one knows f and the distributions of $\mathcal{D}$ and $\mathcal{D}_H$ through the relation,

$$p_H = f\frac{\Lambda_H(t)\mathbb{E}\left[e^{-\rho\mathcal{D}}\right]}{\Lambda_D(t)\mathbb{E}\left[e^{-\rho\mathcal{D}_H}\right]}, \tag{9}$$

for any time $t \in [0, t_L]$ for which $\mathcal{D} < t$ and $\mathcal{D}_H < t$ almost surely, where t_L is the time at which lockdown measures are put into place.

In this paper, we shall assume that $\mathcal{D}_H$ and $\mathcal{D}$ are of the form

$$\mathcal{D}_H = m_H(x_H + (1-x_H)2X_H), \qquad \mathcal{D} = \mathcal{D}_H + m_D(x_D + (1-x_D)2X_D), \tag{10}$$

where X_H and X_D are two independent Beta random variables with support in $[0,1]$. (The parameters m_H and m_D then correspond to the mean infection to hospital admission and admission to death intervals.) The parameters m_H, x_H, m_D and x_D are then numerically fitted to the data, using an MCMC algorithm.

1.5 Estimating the contact rate during lockdown measures

At the start of lockdown measures, the contact rate in the population dropped sharply over the course of a few days. As a result, the daily number of deaths in hospitals started to slow down, before decreasing at a steady rate after approximately three weeks. In addition, the

number of new hospital admissions and ICU admissions also decreased at a steady rate after a (shorter) transition period. Assuming that, at this point, the proportion of susceptible individuals in the overall population remains close to 1, this corresponds to the situation where

$$E(t) \approx E(s)e^{\rho_L(t-s)}, \qquad I(t) \approx I(s)e^{\rho_L(t-s)},$$

for $t \geq s$ and with $\rho_L < 0$ (ρ_L is the growth rate of the epidemic during lockdown). Thus, according to Proposition 3, we can use (4) to estimate the contact rate during lockdown λ_L from the rate of decrease of the daily number of hospital deaths (or hospital admissions).

In what follows, we shall assume that the distribution of the sojourn times $(\mathcal{E}, \mathcal{I})$ is unaffected by the various containment measures put into place during the course of the epidemic. This may not be true, especially if contact tracing and massive testing are implemented at a sufficient scale in the population, but the extent of such measures in France, at least during the early stages of the Covid-19 epidemic, remained very limited.

1.6 Estimating the contact rate after the easing of lockdown measures

The easing of lockdown restrictions was organized in different phases. On 11 May, people were allowed to leave their homes, shops started to reopen and schools progressively welcomed pupils again. On 2 June, bars and restaurants reopened in most of the country, and on 22 June, all schools reopened, along with cinemas, and most activities resumed, although sanitary measures continued to be enforced (*e.g.* wearing a mask remains mandatory in mainy public places including trains and public transports).

After the easing of lockdown restrictions, on May 11, we can assume that the contact rate in the population gradually shifted to a new value, λ_E, corresponding to another growth rate ρ_E. As we shall see below, at the end of the lockdown period, the proportion of susceptible individuals in the population had already dropped by 5 to 10%, at least in Île de France and the Grand Est and Hauts-de-France regions. As a result, the approximation $\overline{S}(t) \approx 1$ may not be valid at this point, preventing us from directly applying (4). Since the epidemic keeps decreasing, however, the value of $\overline{S}(t)$ does not vary much during the period over which we estimate the new growth rate ρ_E. It follows that we can replace λ by $\overline{S}(t_E)\lambda$ in (4), where $\overline{S}(t_E)$ is the proportion of susceptible individuals at the end of the lockdown period, and hence deduce the corrected value of the contact rate.

The same correction should also be applied in the future when estimating the contact rate in later stages of the epidemic.

1.7 Estimating the state of the epidemic

Once we have estimated the contact rates prior to lockdown (λ_0), during lockdown (λ_L) and after lockdown (λ_E), as well as the interval between the (theoretical) index case and the start of lockdown t_L, we are in a position to compute the state of the epidemic. To do this

we numerically solve the equations of Definition 1 with

$$\lambda(s) = \begin{cases} \lambda_0 & \text{if } s < t_L, \\ \lambda_L & \text{if } t_L \leq s < t_E, \\ \lambda_E & \text{if } s \geq t_E, \end{cases}$$

where t_L is time at which lockdown measures are implemented and t_E is the time at which these measures are eased. For the initial condition, using Proposition 3, we choose

$$\overline{S}(0) = 1 - \frac{1}{N}, \qquad \overline{E}(0) = \frac{\boldsymbol{e}}{N}, \qquad \overline{I}(0) = \frac{\boldsymbol{i}}{N}, \qquad \overline{R}(0) = \frac{\boldsymbol{r}}{N},$$

with $\boldsymbol{e}$, $\boldsymbol{i}$, $\boldsymbol{r}$ as in (1) and G_0^c, Ψ_0, Φ_0, F_1 as in (6), choosing ρ equal to the growth rate prior to lockdown in (1) and as the parameter of the exponential random variable Θ in (6). In this way, by Proposition 3, $1 - \overline{S}(t) = \frac{1}{N}e^{\rho t}$ for all $t \leq t_L$, as expected.

It might seem counter-intuitive to start the epidemic with some fraction of removed individuals ($\overline{R}(0) > 0$), even more so as we assume that only one individual is not susceptible at this time. This, however, is only an artefact of our model. We are not trying to estimate the *true* initial state of the epidemic; we merely find a suitable initial condition so that the observed exponential growth prior to lockdown measures fits the observed data.

2 Results

For our estimation of the state of the Covid-19 epidemic in France, we split the country in three regions, or patches: the Grand Est and Hauts-de-France regions (the northeastern part of the country where the epidemic took off earliest), Île de France (*i.e.* the densely populated Paris region) and the rest of the country, excluding Corsica and the overseas territories.

The exclusion of Corsica was motivated by its singular behaviour. Indeed, in this island, the epidemic took off very quickly in March, followed by a very sharp plateau, much more marked than in the rest of the country. In any case, including Corsica in our analysis should not significantly affect our results.

2.1 Growth rates of the epidemic

We measured the growth rate of the cumulative number of infected individuals during the early phase of the epidemic by fitting an exponential curve to the cumulative number of deaths in each patch, between 19 March and 26 March (earlier data was only available at the national level). Despite the fact that very strict lockdown measures were already effective at the time, hospital deaths continued to increase exponentially until at least 26 March, mainly due to the fact that infected individuals who died of Covid-19 during this period had been infected before national lockdown, hence during the exponential growth phase. The growth rates varied between 0.20 in the Grand Est and Hauts-de-France regions and 0.26, 0.28 in the rest of the country and in Île de France, corresponding to a doubling time of the cumulative number of infected individuals of around 2.5 days (see Table 1).

Region	Île de France	Grand Est and Hauts-de-France	Rest of France (excluding Corsica and overseas)	Mainland France
Growth rate during the early phase (from hospital deaths)	0.28	0.20	0.26	0.27
Growth rate under lockdown	-0.060	-0.053	-0.051	-0.055
Growth rate after 11 May	-0.046	-0.056	-0.054	-0.052
Growth rate after 2 June	-0.0076	-0.033	-0.047	-0.019

Table 1: Growth rates measured prior to lockdown measures, in the steady phase of the lockdown period and after the easing of lockdown restrictions in each of the three patches used for the estimation (Île de France, Grand Est and Hauts-de-France, and the rest of mainland France, excluding Corsica). Growth rates prior to lockdown are measured using hospital deaths between 19 March and 26 March (except for the whole country for which we used data from 1st March).

The relatively slow growth rate in the Grand-Est and Hauts-de-France regions might seem surprising, given that this is where the epidemic first took off at a very quick pace. We interpret this as the result of emergency measures which were put into place from 7 March in the most affected areas: school closures, banning of public events and gatherings, *etc.*[AFP20]. Indeed, if the epidemic had been growing at this rate of 0.20 from the beginning, our model shows that it should have started much earlier than it actually did, and the number of hospital deaths in the whole country would have grown more slowly (but for a longer time) at this time, as shown on Figure S1. If instead we assume that the growth rate in this region was 0.27 (*i.e.* the growth rate at the national level) up until 7 March, then 0.20 between 7 March and 16 March, the SEIR model predicts a cumulative number of hospital deaths much closer to what was actually observed at the national level (see Figure S1).

After a few weeks of lockdown, daily hospital admissions and deaths started to decline, again exponentially, at a steady rate in each patch, as shown on Figure 3. We observed a drop in the number of daily admissions and deaths every week-end, which was systematically compensated at the start of the following week, indicating reporting delays. We thus fitted simultaneously three exponential curves to the *cumulative* number of hospital admissions, hospital deaths and ICU admissions between 3 April and 11 May (starting only on 13 April for ICU admissions and deaths). The obtained values are displayed in the second line of Table 1. Overall, the growth rate of the epidemic in each patch stood between -0.051 and -0.06, corresponding to a halving time of the number of infected individuals of around 12 days.

Using the same method, we can estimate the new growth rate of the cumulative number of infected individuals after the easing of lockdown measures, on 11 May. Overall, the epidemic appears to be decreasing but the decrease appears to slow down after the easing of

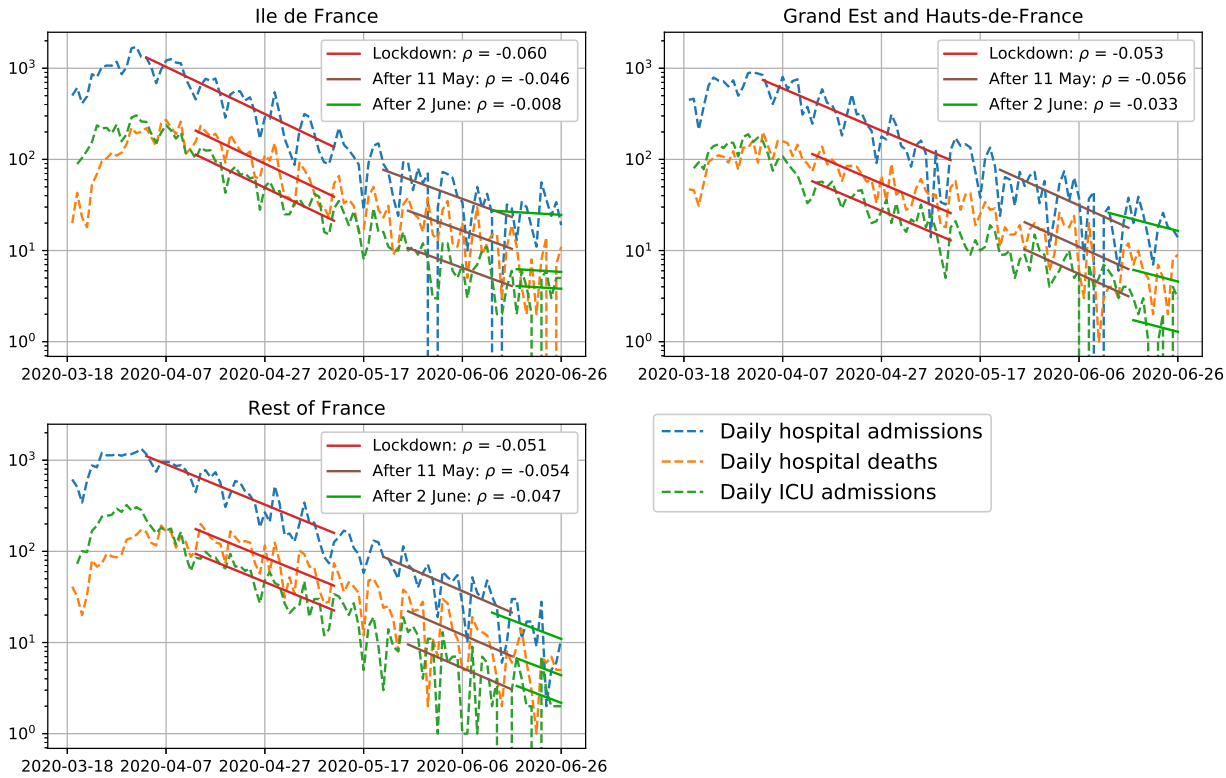

Figure 3: Growth rates of the cumulative number of infected individuals during (ρ_L) and after (ρ_E) lockdown in France, deduced from the number of hospital admissions, ICU admissions and hospital deaths in each patch. In each instance, an exponential curve was fitted simulateneously to each cumulative number of hospital admissions, ICU admissions and deaths. The last value in each patch is deduced from the cumulative number of hospital admissions, ICU admissions and hospital deaths during the last two weeks of June.

lockdown restrictions in Île de France (-0.046 instead of -0.06 during lockdown), reflecting the increased contact rate in this region. This seems to be confirmed by the most recent data, which suggests an increased propagation of the virus in Île de France and in the Grand Est and Hauts-de-France regions (third set of solid lines on Figure 3). This raises fears that further lifting of restrictions might cause this growth rate to increase even further, eventually triggering a second epidemic wave. Indeed, not enough time has elapsed since the last easing of restrictions (on 2 June and even later on 22 June) in order to accurately estimate the new contact rate in the population.

Note that the SEIR model predicts a small but non-negligible (of the order of 5 to 10%) level of acquired immunity, at least in Île de France and the Grand Est and Hauts-de-France regions (see Subsection 2.3). We would thus expect to observe a progressive acceleration of the exponential decrease in the number of new hospital admissions and deaths. The fact that this does not take place suggests that the easing of lockdown restrictions somewhat compensates the effect of this acquired immunity.

So far, hospital admissions, ICU admissions and hospital deaths seem to continue to decrease, suggesting that the new contact rate remains small enough to avoid a second

epidemic wave. Note however that, because of the delay between infection and hospital admission, our estimate of the contact rate is always at least two to three weeks behind the current one. As a result, the new contact rate after the latest easing of restrictions might already be too high, and we would only notice three weeks later.

2.2 Reproduction numbers and the effect of lockdown measures

Using the relation (4) and our assumptions on the distribution of $(\mathcal{E}, \mathcal{I})$ from Subsection 1.3, we can estimate the values of the reproduction number R_0 in each patch during each phase of the epidemic. Since the growth rates prior to lockdown and during lockdown are relatively uniform across the different patches (respectively, around 0.27 and -0.049, taking into account our interpretation of the lower growth rate in the Grand Est and Hauts-de-France regions), we use the same value for all patches.

The proportion of reported individuals affects both the expectation of the infectious period $\mathbb{E}[\mathcal{I}]$ and the whole distribution of $(\mathcal{E}, \mathcal{I})$. Considering Corollary 4, we thus observe that the estimated value of the reproduction number depends on the proportion of reported individuals, and more generally on the whole distribution of $(\mathcal{E}, \mathcal{I})$.

If we assume that all infected individuals are reported ($p_R = 1$), the reproduction number in France is 3.4 prior to lockdown measures and 0.79 during lockdown. If, however, we assume that all individuals are unreported ($p_R = 0$), the reproduction number in France is 6.3 prior to lockdown and 0.66 during lockdown. The estimate of the reproduction number then varies continuously between these values for intermediate values of the proportion of reported individuals. For instance, if the proportion of reported individuals is close to 0.8, the reproduction number is estimated to be 4.2 during the early phase of the epidemic and 0.73 during lockdown, whereas if this proportion is 0.2, $R_0 = 6$ before lockdown and 0.67 during lockdown. Note that our estimate of R_0 fits well the results in [SLX+20].

Country	R_0 prior to lockdown	R_0 during lockdown	Accounting for exposed period	Reference
France	2.9 - 3.35	0.55 - 0.72	implicit	[SRE+20]
France	3.1 - 3.3	0.47	no	[RKP+20]
France	3.3	0.5	yes	[SKL+20]
France	3.4 ± 0.1	0.65 ± 0.04	yes	[Mam20]
China	4.7 - 6.6	/	yes	[SLX+20]

Table 2: Estimates of the reproduction number in the literature

The relatively large values (around 6) obtained for small proportions of reported individuals suggest that this proportion is closer to 0.8 than to 0.2. Moreover, the values we obtain, even for large proportions of reported individuals, are sometimes larger than other estimates in the literature (Table 2). We can attribute this to two things. First, as shown in Figure 2, the Markovian SEIR model tends to slightly underestimate R_0 in the two regimes we are interested in (growth in the early phase and decrease under lockdown). Second, some studies do not take into account the exposed period $\mathcal{E}$ in their model. However, Corollary 4 shows that neglecting the exposed period leads to underestimating of the reproduction number

R_0, here by a factor of 0.76 for each day of exposed period ($e^{-\rho} = 0.76$ with $\rho = 0.27$). Finally, uncertainty about the infectious period $\mathcal{I}$ also affects the estimates of the reproduction number.

2.3 State of the epidemic and acquired immunity at the population level

Using (8) and Subsection 1.7, we can estimate the state of the epidemic in each patch, given the infection fatality ratio f and the distribution of the delay $\mathcal{D}$. Unfortunately, these two quantities are notoriously hard to measure during the early stage of the epidemic.

The infection fatality ratio for the Covid-19 epidemic in France has been estimated in at least two studies [RKP+20, SKL+20]. Both studies found a fatality ratio of 0.5%, with significant variation across age classes in [SKL+20]. These studies only account for hospital deaths, as do we, even though a significant number of deaths take place outside hospitals, mainly in nursing homes. Hence this ratio has to be corrected (roughly by a factor 1.6 [RKP+20]) to obtain the true infection fatality ratio of Covid-19 (and to obtain correct predictions for the expected number of deaths). Nevertheless, since we use hospital deaths to calibrate our SEIR model, we shall use the infection fatality ratio estimated by [SKL+20, RKP+20] when using (8).

We can at least bound this ratio from below using the observed excess mortality in some regions. For example, in Lombardy (Italy), 13,575 people died of Covid-19, for a total population of 10 million, showing that this ratio is at least 0.14% (or 0.088% for hospital deaths). Another estimate of this ratio was obtained in Germany, were around 1,000 individuals were tested, out of which 15.5% tested positive and 7 people died, yielding an infection fatality ratio of 0.37% [SSK+20] (corresponding to 0.23% for hospital deaths).

On the other hand, the infection fatality ratio can be bounded from above by the apparent death rate, that is to say, the ratio of Covid-19 related deaths and declared positive cases, at least while the epidemic seems to be receding, as is the case in France in June 2020, with less than 50 deaths per day. Taking only hospital deaths into account, this suggests that f is no more than 12%. Other countries have much lower apparent death rates, *e.g.* South Korea (2.3%), Germany (4.6%) and even Italy (14%). These discrepancies can mostly be attributed to differences in testing capacities and to a lesser extent to differences in hospital capacities.

The numerical fit to the hospital admissions and deaths data using (10) then yields the following:

$$\begin{aligned} m_H &= 14.8 \text{ (C.I. [11.5, 18])}, \quad x_H = 0.18 \text{ (C.I. [0.04, 0.43])}, \\ m_D &= 4.7 \text{ (C.I. [3.85, 7.17])}, \quad x_D = 0.96 \text{ (C.I. [0.11, 0.99])}, \end{aligned} \tag{11}$$

corresponding to a mean infection to hospital admission interval of 14.8 days and a mean infection to death interval of 19.5 days. The confidence intervals are computed from the trajectory of the MCMC run. One notes that the data contain more information on the mean intervals than on the ratios x_H ans x_D.

We also tried to estimate the proportion of reported individuals p_R at the same time as the delay parameters (m_H, x_H, m_D, x_D), but the fit was on average equally good for very

different values of this proportion (confidence interval: [0.06, 1]). As a result we are not able to give a meaningful estimate of this proportion using the available data.

We estimated the probabilities of being admitted to hospital given infection. These obviously depend on the infection fatality ratio, but, on the other hand, the admission to death ratio (*i.e.* the probability of dying given that one is admitted to a hospital) is constant. We estimated that the admission to death ratio was 0.13 in Île de France, 0.24 in the Grand Est and Hauts-de-France regions and 0.14 elsewhere. Note that we consider that these do not evolve over time, which may not be accurate, especially in the Grand Est and Hauts-de-France regions. Indeed, the relatively high death rate in hospitals in the Grand Est and Hauts-de-France regions during the early phase of the epidemic reflects the fact that hospitals in this part of the country were saturated at the time (army medics were eventually dispatched to help the local health system).

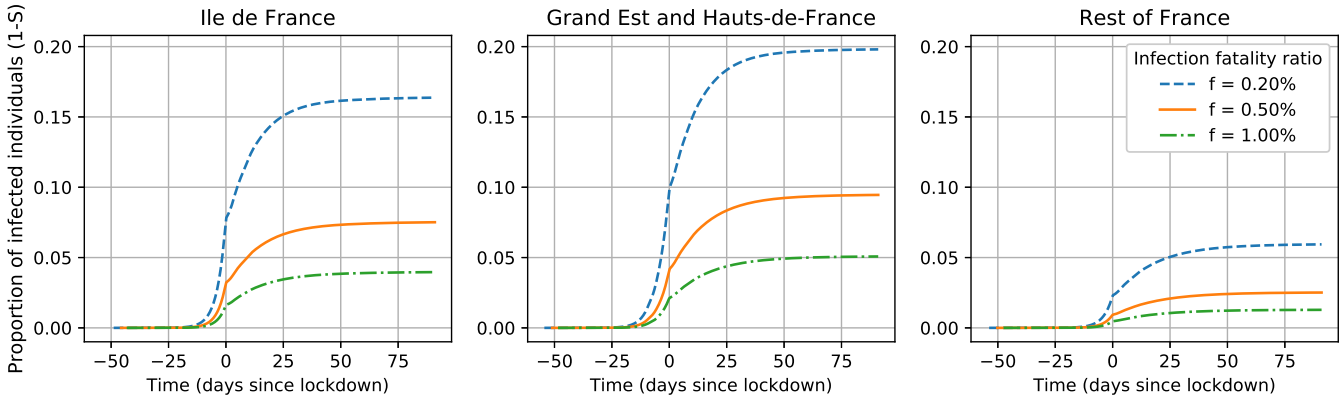

Figure 4: Predicted level of immunity in each patch as a function of the infection fatality ratio f. All the simulations use $p_R = 0.6$ and assume that death occurs on average 19.5 days after being infected. The rightmost point of each curve corresponds to the value on 15 June

Figure 4 shows the predicted levels of immunity (*i.e.,* the proportion of infected individuals) up to 15 June in each patch for three values of the infection fatality ratio f, using the delay distributions in (10) with the estimated parameters (11). Unsurprisingly, higher values of this ratio lead to lower predicted levels of immunity. Note that underestimating $\mathcal{D}$ has the same effect as overestimating the infection fatality ratio, as can be noted from (8). As a result it is equally important to estimate f and the distribution of $\mathcal{D}$, a fact which is too easily overlooked.

For the fatality ratio estimated by [RKP+20, SKL+20] ($f = 0.5\%$), we estimate that between 8 and 10% of the population has been infected by mid June in the most affected regions (Île de France, Grand Est and Hauts-de-France), and around 2 to 3% of the population in the rest of mainland France. As shown by Figure 4, these estimates vary significantly as a function of f and the distribution of the delay between infection and death ($\mathcal{D}$), highlighting the need for more accurate measures of these quantities.

3 Discussion

Our claim in this article is that using non-Markovian epidemic models to assess the state of an ongoing epidemic leads to more accurate estimates of the main quantities of interest and more pertinent predictions concerning the future of the epidemic. This has already been advocated in [ZFK20] and [SRE+20], and our goal was to lay out the main arguments using both theoretical tools and simulations. We also wished to stress which quantities need to be more precisely evaluated in order to better predict the state and the future evolution of the epidemic.

As exposed in Remark 5, knowing the distribution of the sojourn times in each compartment $(\mathcal{E}, \mathcal{I})$ is crucial to estimate the value of R_0 from the initial growth rate of the epidemic. Without knowledge on these sojourn times, the reproduction number can only be estimated using contact tracing, which is both difficult and costly. As shown in Figure 2, Markovian epidemic models (but not only) can in some regime underestimate the value of R_0 for a given initial growth rate of the epidemic. If this is the case, then they will underestimate the final proportion of infected individuals in the population in the absence of containment measures, and thus the predicted death toll of the epidemic.

Aside from this, the Markovian and non-Markovian models differ in their transitory behavior, even if the initial growth rate of the epidemic is kept fixed. This can most easily be noted just after strict lockdown measures were quickly put into place across the whole country around 16 March. Indeed, Figure 5 shows that the non-Markovian model has a greater inertia than the Markovian model. That is to say that, after lockdown measures are implemented, however strict, the cumulative number of infected individuals continues to grow, much more than predicted by the Markovian model. This is apparent in the slow decrease in hospital admissions and deaths even after 3 weeks of strict lockdown, something which the Markovian model would not have predicted (one way to force the Markovian model to display this kind of behavior would be to make the change in the contact rate more gradual, but we do not believe that this would reflect what took place in France).

A direct consequence is that the Markovian model will tend to underestimate the number of infectious individuals (*i.e.* who are still contagious) at the time of the easing of lockdown restrictions, and to underestimate the level of acquired immunity in the population.

We have also shown that accurate estimates of both the infection fatality ratio and the distribution of the delay between infection and death (or hospital admission) are needed in order to reliably infer the state of the epidemic during its early stage. This can either be achieved through sophisticated statistical procedures, such as in [RKP+20, SKL+20], or by performing serological tests on large samples of a population which has been infected.

Note that an analysis similar to the one conducted in this paper can be done for a model where the infectiosity of each individual is a random function of the time since infection, those random functions for the various individuals being i.i.d., see [FPP20]. This would yield a model which could account for the variability, both in time and between individuals, of the infectiosity of each infected individual, something which is apparent for Covid-19 for example in [HLW+20].

Conflicts of interest The authors declare no conflict of interest.

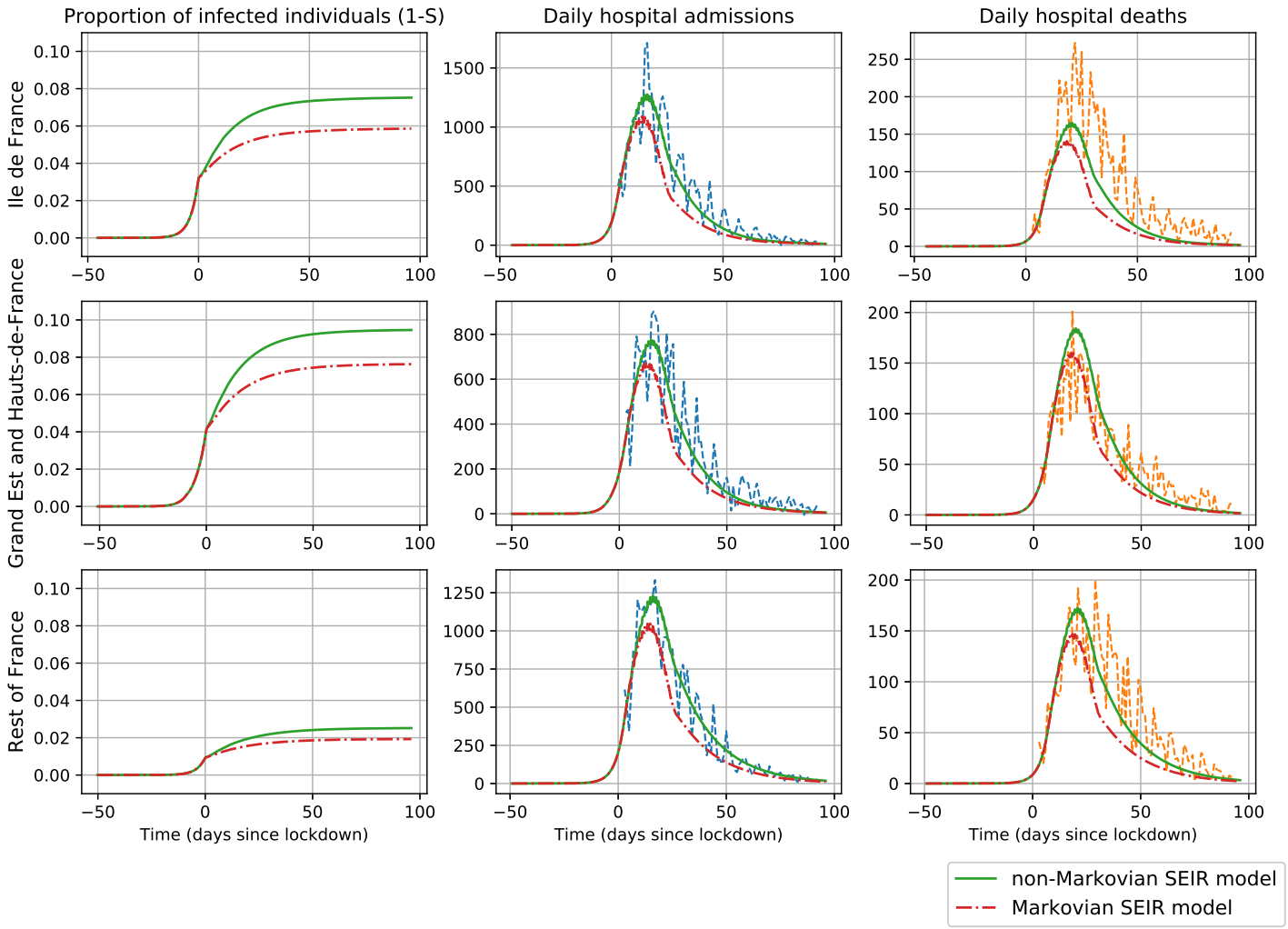

Figure 5: Predictions of the non-Markovian SEIR and Markovian SEIR models. The left column shows the cumulative number of infected individuals, the middle column shows daily hospital admissions (the dashed blue line corresponds to the data gathered from Santé Publique France) and the last column shows the daily hospital deaths (the dashed orange line shows the data from Santé Publique France). The non-Markovian model assumed a proportion of reported individuals of 0.6, and the Markovian model had the same expected exposed and infectious periods.

Code availability The code used to simulate the SEIR model and to analyse the data is made available at https://github.com/rforien/Fit_Covid19_nonMarkovian.git

A Proof of Proposition 3

We start with the case $\rho > 0$. Replacing $(E(t), I(t), R(t))$ in (2) by their expressions in (3) yields the following equations

$$\begin{aligned}
\boldsymbol{e}e^{\rho t} &= \lambda \boldsymbol{i} \int_{-\infty}^{t} G^c(t-s)e^{\rho s}ds \\
\boldsymbol{i}e^{\rho t} &= \lambda \boldsymbol{i} \int_{-\infty}^{t} \Psi(t-s)e^{\rho s}ds \\
\boldsymbol{r}e^{\rho t} &= \lambda \boldsymbol{i} \int_{-\infty}^{t} \Phi(t-s)e^{\rho s}ds.
\end{aligned}$$

The second equation translates into

$$\begin{aligned}
\lambda^{-1} &= \int_0^{\infty} \mathbb{P}(\mathcal{E} \leq s < \mathcal{E} + \mathcal{I})e^{-\rho s}ds \\
&= \mathbb{E}\left[\int_{\mathcal{E}}^{\mathcal{E}+\mathcal{I}} e^{-\rho s}ds\right] \\
&= \frac{\mathbb{E}\left[e^{-\rho\mathcal{E}}(1-e^{-\rho\mathcal{I}})\right]}{\rho}.
\end{aligned} \tag{12}$$

yielding (4). The remaining two equations give the relations between $\boldsymbol{e}$, $\boldsymbol{r}$ and $\boldsymbol{i}$. The first one yields

$$\begin{aligned}
\boldsymbol{e} &= \lambda \boldsymbol{i} \int_0^{\infty} \mathbb{P}(\mathcal{E} > s)e^{-\rho s}ds \\
&= \frac{\lambda \boldsymbol{i}}{\rho}\left(1 - \mathbb{E}\left[e^{-\rho\mathcal{E}}\right]\right),
\end{aligned} \tag{13}$$

and the third one gives, for $\rho > 0$,

$$\begin{aligned}
\boldsymbol{r} &= \lambda \boldsymbol{i} \int_0^{\infty} \mathbb{P}(\mathcal{E} + \mathcal{I} \leq s)e^{-\rho s}ds \\
&= \frac{\lambda \boldsymbol{i}}{\rho}\mathbb{E}\left[e^{-\rho(\mathcal{E}+\mathcal{I})}\right].
\end{aligned}$$

Combining these with the constraint $\boldsymbol{e} + \boldsymbol{i} + \boldsymbol{r} = 1$ yields (5). Note that $\lambda \boldsymbol{i} = \rho$.

To prove the second part of the statement, we only need to check that the quantities defined in (6) satisfy

$$\begin{aligned}
\boldsymbol{e}G_0^c(t) &= \lambda \boldsymbol{i} \int_{-\infty}^{0} G^c(t-s)e^{\rho s}ds \\
\boldsymbol{i}F_1^c(t) + \boldsymbol{e}\Psi_0(t) &= \lambda \boldsymbol{i} \int_{-\infty}^{0} \Psi(t-s)e^{\rho s}ds \\
\boldsymbol{r} + \boldsymbol{i}F_1(t) + \boldsymbol{e}\Phi_0(t) &= \lambda \boldsymbol{i} \int_{-\infty}^{0} \Phi(t-s)e^{\rho s}ds\,.
\end{aligned}$$

Indeed, combining these with (2) and the first part of the proposition, we obtain the fact that (3) solves (1). To check the first identity, write

$$\begin{aligned}\int_{-\infty}^{0} G^c(t-s)e^{\rho s}ds &= \int_{t}^{+\infty} \mathbb{P}(\mathcal{E}>s)e^{\rho(t-s)}ds \\ &= \mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}>t\}}\int_{t}^{\mathcal{E}} e^{\rho(t-s)}ds\right] \\ &= \frac{1}{\rho}\mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}>t\}}(1-e^{-\rho(\mathcal{E}-t)})\right] .\end{aligned}$$

Multiplying by $\lambda\boldsymbol{i}/\boldsymbol{e}$, we obtain

$$G_0^c(t) = \frac{\lambda\boldsymbol{i}}{\boldsymbol{e}}\int_{-\infty}^{0} G^c(t-s)e^{\rho s}ds = \frac{\mathbb{E}\left[\mathbf{1}_{\{\mathcal{E}>t\}}(1-e^{-\rho(\mathcal{E}-t)})\right]}{\mathbb{E}\left[1-e^{-\rho\mathcal{E}}\right]} .$$

We conclude by noting that the term on right-hand-side equals $\mathbb{P}(\mathcal{E}-\Theta>t\,|\,\Theta<\mathcal{E})$, where Θ is an independent exponential random variable with parameter ρ.

Plugging (6) and (7) in the second equation, we obtain

$$\begin{aligned}\boldsymbol{i}F_1^c(t)+\boldsymbol{e}\Psi_0(t) = \mathbb{P}(\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I})\,\mathbb{P}(t+\Theta<\mathcal{E}+\mathcal{I}\,|\,\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I}) \\ +\mathbb{P}(\theta<\mathcal{E})\,\mathbb{P}(\mathcal{E}\le t+\Theta<\mathcal{E}+\mathcal{I}\,|\,\Theta<\mathcal{E}).\end{aligned}$$

Since, for $t\ge 0$,

$$\{t+\Theta<\mathcal{E}+\mathcal{I}\}\cap\{\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I}\} = \{\mathcal{E}\le t+\Theta<\mathcal{E}+\mathcal{I}\}\cap\{\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I}\},$$

and

$$\{\mathcal{E}\le t+\Theta<\mathcal{E}+\mathcal{I}\}\cap\{\mathcal{E}+\mathcal{I}\le\Theta\} = \emptyset,$$

by the law of total probability,

$$\boldsymbol{i}F_1^c(t)+\boldsymbol{e}\Psi_0(t) = \mathbb{P}(\mathcal{E}\le t+\Theta<\mathcal{E}+\mathcal{I}).$$

On the other hand, since $\lambda\boldsymbol{i}=\rho$,

$$\begin{aligned}\lambda\boldsymbol{i}\int_{-\infty}^{0}\Psi(t-s)e^{\rho s}ds &= \int_{0}^{+\infty}\mathbb{P}(\mathcal{E}\le t+s<\mathcal{E}+\mathcal{I})\rho e^{-\rho s}ds \\ &= \mathbb{P}(\mathcal{E}\le t+\Theta<\mathcal{E}+\mathcal{I}).\end{aligned}$$

For the third equation, we proceed in the same way, plugging (6) and (7), we obtain

$$\begin{aligned}\boldsymbol{r}+\boldsymbol{i}F_1(t)+\boldsymbol{e}\Phi_0(t) = \mathbb{P}(\mathcal{E}+\mathcal{I}\le\Theta)+\mathbb{P}(\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I})\,\mathbb{P}(\mathcal{E}+\mathcal{I}\le t+\Theta\,|\,\mathcal{E}\le\Theta<\mathcal{E}+\mathcal{I}) \\ +\mathbb{P}(\Theta<\mathcal{E})\,\mathbb{P}(\mathcal{E}+\mathcal{I}\le t+\Theta\,|\,\Theta<\mathcal{E}).\end{aligned}$$

Now, we note that, for $t\ge 0$,

$$\{\mathcal{E}+\mathcal{I}\le t+\Theta\}\cap\{\mathcal{E}+\mathcal{I}\le\Theta\} = \{\mathcal{E}+\mathcal{I}\le\Theta\}.$$

As a result, applying the law of total probability, we obtain

$$\boldsymbol{r} + \boldsymbol{i}F_1(t) + \boldsymbol{e}\Phi_0(t) = \mathbb{P}(\mathcal{E} + \mathcal{I} \leq t + \Theta).$$

On the other hand, since $\lambda \boldsymbol{i} = \rho$,

$$\begin{aligned}\lambda \boldsymbol{i} \int_{-\infty}^{0} \Phi(t-s)e^{\rho s}ds &= \int_{0}^{+\infty} \mathbb{P}(\mathcal{E} + \mathcal{I} \leq t + s)\rho e^{\rho s}ds \\ &= \mathbb{P}(\mathcal{E} + \mathcal{I} \leq t + \Theta).\end{aligned}$$

This concludes the proof of the first part of Proposition 3.

In the case $\rho < 0$, note that the computations in (12) and (13) remain valid, yielding the second part of the statement.

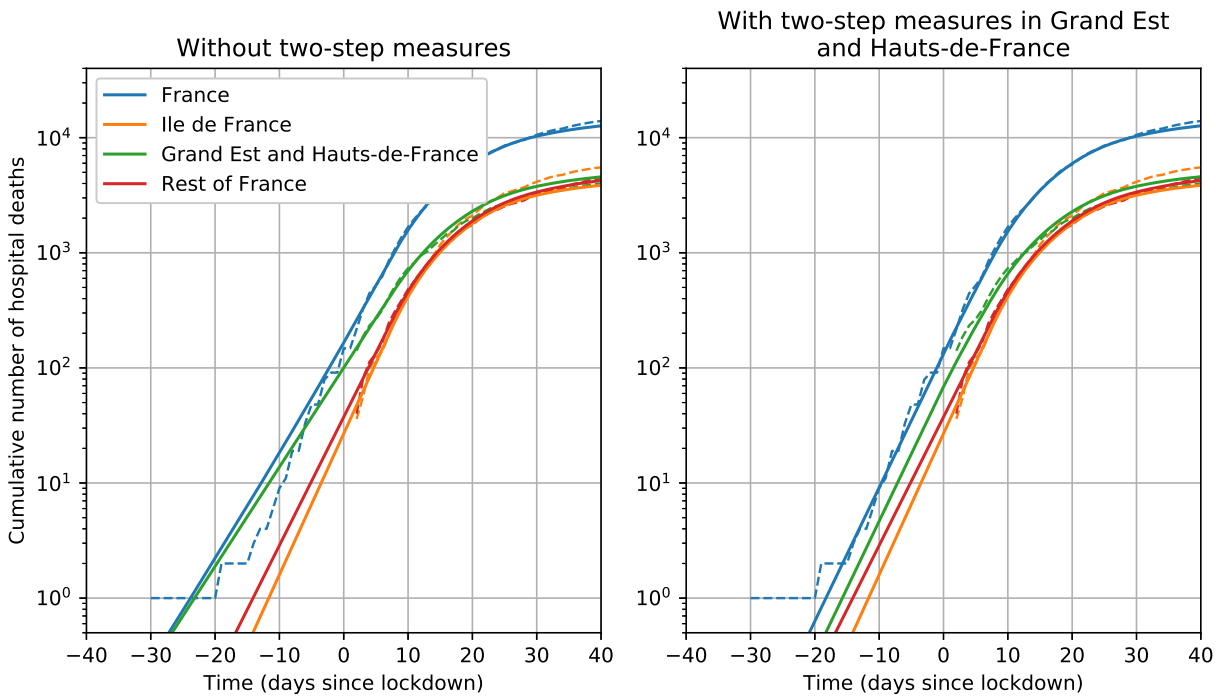

Figure S1: Predicted deaths during the early phase of the epidemic with and without two step measures in the Grand Est and Hauts-de-France regions. On the left, the growth rate inferred from the cumulative number of deaths in this patch is taken to be the growth rate of the epidemic in this region since the beginning, leading to a slow growth of the epidemic during the early phase, which does not fit the observations. On the right, the growth rate of the epidemic is 0.27 at first, and then 0.20 from 7 March, before changing again at the time of lockdown. The agreement with the observations is considerably improved.

References

[AFP20] AFP. Coronavirus. Fermeture des établissements scolaires dans l'Oise et le Haut-Rhin, annonce Édouard Philippe. https://www.ouest-

france.fr/sante/virus/coronavirus/coronavirus-fermeture-des-etablissements-scolaires-dans-l-oise-et-le-haut-rhin-annonce-edouard-6768736, March 2020.

[BCCF] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*.

[BP19] Tom Britton and Etienne Pardoux. Stochastic epidemics in a homogeneous community. In *Stochastic Epidemic Models with Inference*, pages 1–120. Springer, 2019.

[Bri20a] Tom Britton. Basic estimation-prediction techniques for Covid-19, and a prediction for Stockholm. *medRxiv preprint*, April 2020.

[Bri20b] Tom Britton. Basic prediction methodology for Covid-19: estimation and sensitivity considerations. *medRxiv preprint*, March 2020.

[CSC+20] Pascal Cossart, Olivier Schwartz, Patrick Couvreur, Frédéric Tangy, Dominique Costagliola, Olivier Faugeras, and Arnaud Fontanet. COVID-19 - Séance exceptionnelle de l'Académie des sciences. *www.youtube.com/watch?v=OCvDDlBD7yk*, 2020.

[FPP20] Raphaël Forien, Guodong Pang, and Étienne Pardoux. Epidemic models with varying infectiosity. 2020.

[GAA+20] Stéphane Gaubert, Marianne Akian, Xavier Allamigeon, Marin Boyet, Baptiste Colin, et al. Understanding and monitoring the evolution of the Covid-19 epidemic from medical emergency calls: the example of the Paris area. *medRxiv preprint*, March 2020.

[HLW+20] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine*, 26(5):672–675, 2020.

[KM27] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

[LMSW20] Zhihua Liu, Pierre Magal, Ousmane Seydi, and Glenn Webb. A COVID-19 epidemic model with latency period. *Infectious Disease Modelling*, April 2020.

[Mam20] Gary A Mamon. Fit of french covid-19 hospital data with different evolutionary models: regional measures of r0 before and during lockdown. *arXiv preprint arXiv:2005.06552*, 2020.

[PP20] Guodong Pang and Etienne Pardoux. Functional Limit Theorems for Non-Markovian Epidemic Models. *arXiv:2003.03249 [math]*, March 2020.

[RKP+20] Lionel Roques, Etienne K Klein, Julien Papaix, Antoine Sar, and Samuel Soubeyrand. Using early data to estimate the actual infection fatality ratio from COVID-19 in France. *Biology*, 9(5):97, 2020.

[San20] Santé Publique France. Données hospitalières relatives à l'épidémie de COVID-19, 2020.

[SKL+20] Henrik Salje, Cécile Tran Kiem, Noémie Lefrancq, Noémie Courtejoie, Paolo Bosetti, Juliette Paireau, Alessio Andronico, Nathanaël Hozé, Jehanne Richet, Claire-Lise Dubost, et al. Estimating the burden of SARS-CoV-2 in France. *Science*, 2020.

[SLX+20] Steven Sanche, Yen Ting Lin, Chonggang Xu, Ethan Romero-Severson, Nicolas W Hengartner, and Ruian Ke. The novel coronavirus, 2019-ncov, is highly contagious and more infectious than initially estimated. *arXiv preprint arXiv:2002.03268*, 2020.

[SRE+20] Mircea T Sofonea, Bastien Reyné, Baptiste Elie, Ramsès Djidjou-Demasse, Christian Selinger, Yannis Michalakis, and Samuel Alizon. Epidemiological monitoring and control perspectives: application of a parsimonious modelling framework to the covid-19 dynamics in France. 2020.

[SSK+20] Hendrik Streeck, Bianca Shculte, Beate Kuemmerer, et al. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *MedRxiv preprint doi: https://doi.org/10.1101/2020.05.04.20090076*, 2020.

[ZFK20] Sandor D. Katz Zoltan Fodor and Tamas G. Kovacs. Why integral equations should be used instead of differential equations to describe the dynamics of epidemics. *Arxiv*, April 2020.