

# Multiple epidemic waves as the outcome of stochastic SIR epidemics with behavioral responses: a hybrid modeling approach

Magdalena Ochab<sup>1</sup>, Piero Manfredi<sup>2</sup>, Krzysztof Puzynski<sup>1</sup>, Alberto d’Onofrio<sup>3,4,5\*\*</sup>

February 5, 2022

<sup>1</sup> Department of Systems Biology and Engineering, Silesian University of Technology, Gliwice, Poland

<sup>2</sup> Department of Economics and Management, University of Pisa, Via Ridolfi 10, 5612 Pisa, Italy.

<sup>3</sup> Department of Mathematics and Statistics, Strathclyde University, Glasgow, Scotland (UK)

<sup>4</sup> International Prevention Research Institute<sup>1</sup>, 95 Cours Lafayette 69006 Lyon, France

<sup>5</sup> Institut Camille Jordan, Université Claude Bernard Lyon 1, 21 Avenue Claude Bernard, 69100 Villeurbanne (France).

## Abstract

In the behavioral epidemiology (BE) of infectious diseases, little theoretical effort seems to have been devoted to understand the possible effects of individuals’ behavioral responses during an epidemic outbreak in small populations. To fill this gap, here we first build general, behavior implicit, SIR epidemic models including behavioral responses and set them within the framework of nonlinear feedback control theory. Second, we provide a thorough investigation of the effects of different types of agents’ behavioral responses for the dynamics of **hybrid** stochastic SIR outbreak models.

**In the proposed model, the stochastic discrete dynamics of infection spread is combined with a continuous model describing the agents’ delayed behavioral response. The delay reflects the memory mechanisms with which individuals enact protective behavior based on past data on the epidemic course. This results in a stochastic hybrid system with time-varying transition probabilities.** To simulate such system, we extend Gillespie’s classic stochastic simulation algorithm by developing analytical formulas valid for our classes of models. **The algorithm is** used to simulate a number of stochastic behavioral models and to classify the effects of different types of agents’ behavioral responses. **In particular this work focuses on the effects of the structure of the response function and of the form of the temporal distribution of such response.** Among the various results, we stress the appearance of multiple, stochastic epidemic waves triggered by the delayed behavioral response of individuals.

**Keywords:** Multiple Epidemic Waves, Social Distancing, Hybrid Systems, Stochastic Epidemic Models, Human Behavior, Memory Effects, Gillespie Algorithm.

## 1 Introduction

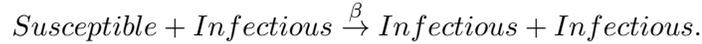
The principles of Mathematical Epidemiology (ME) were established more than one century ago with the milestone work by Sir Ronald Ross in 1916 and by Kermack and McKendrick [1, 2].

---

\*Corresponding Author. Email: alberto.d-onofrio@strath.ac.uk, adonofrio1967@gmail.com

<sup>1</sup>A. d’Onofrio is no more affiliated at IPRI since 15 June 2020

The Kermack and McKendrick's lied in the description of infection transmission by means of the mass action law of Chemistry and Statistical Mechanics. In their formulation, the individuals' social contact patterns relevant for transmission are represented as collisions between particles of a perfect gas, while transmission is modeled as a chemical reaction occurring with a certain probability upon a random encounter between two individuals of different types, namely a susceptible subject and an infective one [3]:



As a consequence, in their approaches, the social *contact rate* per individual, and the *transmission rate*  $\beta$  per social contact, were taken essentially as natural constants valid for appropriate combinations of spatial settings, human activities, cultural habits, institutions etc. However, it was acknowledged the possible dependence of these constants on the seasonality of social phenomena, e.g. the school calendar, or on climatic phenomena [4].

In the last four decades, mathematical models of infectious diseases have become key supporting tools for public health decisions [5–7]. This development was made possible by the increasing awareness of the need for better data and statistical tools allowing a finer and finer description of the transmission process. **Further dimensions were incorporated in the models: individual-level characteristics, as age and sex; meso-level fundamental structures, such as the community composition by household type, by space, by social activity; etc...** [5].

**The** impulse impressed by the fear for avian influenza and the 2009 H1N1 flu pandemic **induced** the development of highly sophisticated mathematical and computational models, **which** better integrate models with data [6, 8], **and which were** used to describe and predict the worldwide dynamics of influenza pandemics. Of course, the current COVID-19 pandemic, and related mitigation measures, are bringing further astonishing momentum to the discipline, with an endless list of contributions (**more than 9000 modeling papers and PhD Theses on COVID-19 are reported by Google Scholar at February 2 2022**).

However, in essentially all public health models, the individuals' social, transmission and vaccination behavior remain unaffected regardless of the trends in the risks of acquiring the infection they might perceive from available information.

Clearly, this is an unrealistic abstraction, increasingly less applicable in contemporary scenarios. **This became even more true** during the **current** COVID-19 pandemic, which represented a huge open-sky laboratory of humans' behaviors [9]. **Indeed**, individuals are frantic information-seekers and therefore hardly unaffected by the state of the disease and *related information*.

Attempts to remove this abstraction have led, in the last 20 years, to the birth of a new branch of ME that was termed the *behavioral epidemiology* (BE) of infectious diseases [7, 10–12].

**The** pioneering work in **BE** was due to Capasso and Serio [13, 14], who first assumed that the transmission intensity might also depend on the level of infection spread by modeling the contact rate  $\beta$  as a decreasing function of infection prevalence. This was the first instance of an epidemiological model incorporating individuals' *spontaneous social-distancing behavior* by a *prevalence-dependent* contact rate [7, 10–12]. A key postulate of **BE** is that infection spread is not anymore an independent process but it is instead the outcome of the interplay of an *infection-spread layer* with a number of other layers, first of all a *behavior layer*, in turn modulated by an *information layer*. In simple words, human behavior relevant for infection spread may be critically affected by the information about the spread and severity of the considered infection. And any change in human behavior will possibly affect infection trends, that might eventually feedback on behavior itself. This means that the spread of information is a critical component of infection transmission modeling and therefore requires careful modeling *per se*. All this led some of us to develop new classes of epidemiological models by using new variables, that we termed *information indexes*, in order to include not only individual perceptions about the current state of infection, but also the memory of past spread [7, 15, 16] and eventually also spatio-temporal memories. In particular, in [17] the Capasso-Serio model was extended by considering a deterministic *SIR* model for an endemic infection disease with a general *information-dependent* transmission rate allowing to also include past available information on infection spread.

In this article, we depart from the previously cited works [13, 17] along two main directions.

The first one deals with the interpretation side: we will cast the aforementioned behavior implicit models including the effects of delayed information [17]) in the framework of the theory of nonlinear feedback control systems [18–21]. The underlying idea is that the individuals’ behavioral responses considered in the BE literature, whereby e.g., agents reduce their daily number of social contacts depending on the information they have on disease severity. **This** is nothing other than a (nonlinear) feedback control. In particular, unlike many feedback systems studied in mathematical biology [19, 22, 23], the one considered here is a genuine feedback. **Indeed**, it is the outcome of voluntary - though uncoordinated - actions of agents.

The second main departure point is that most BE studies are based on deterministic models and as such they are suitable to describe epidemic or endemic scenarios for large populations only.

More in detail, we aim at studying the impact of human behavior during the spread of an epidemic outbreak in small-medium size populations, by taking into account that human decisions might also depend on past information. This has a number of implications. First, the involved nonlinear feedback control systems are stochastic, and therefore depart from the classical asymptotic deterministic description based on the qualitative theory of differential equations [18–21]. Moreover, as we are dealing here with a system having a finite life-span, it is necessary to assess the impact of the control action on a mainly transient phenomenon.

In particular, we depart from classical stochastic models of mathematical epidemiology, generically belonging (as the models of chemical kinetics, of which they can be considered as a subcase [7]) to the class of nonlinear birth and death Markov processes [24, 25]. The latter processes can be simulated by means of the well known Gillespie algorithm [26, 27]. However, the presence of the information memory index (denoted as  $M(t)$  in this paper) makes, from the theoretical viewpoint, the resulting model a *hybrid* stochastic system [28–30], where a birth-death model of epidemic spread is coupled with a piecewise deterministic model for the information index  $M(t)$ . The main implications are the following: i) the state variables of the stochastic epidemic sub-system impact on the - otherwise deterministic - dynamics of the memory; ii) in its turn, the model of the memory impacts on the birth-death epidemic component by making the key *occurrence probabilities*, namely the transmission rate  $\beta$ , heterogeneous in time. From the computational viewpoint, the latter issue requires to modify Gillespie’s algorithm following the lines of [28–31].

Based on this background, we then use selected simulation experiments to analyse and classify the effects of different individuals’ delayed behavioral responses during the course of a stochastic SIR-type epidemic in a fully susceptible population. In particular, the agents’ responses to the epidemic threat are modulated by: (i) the structure of the behavioral response function, (ii) the form and amplitude of the response function temporal kernels. Consistently, for each hypothesis on the behavioral response function, the resulting system is simulated to adequately reconstruct a number of key features of the stochastic epidemic, namely the simulated distributions of (i) the final attack rate, (ii) the epidemic extinction time, (iii) the epidemic maximum prevalence etc.

**Finally, we explicitly stress that our work did not intend to refer to COVID-19, which is an infection with a more complex dynamical structure than the one considered here [32–34]. Instead, our work had a quite general purpose namely to develop a general stochastic behavioral epidemiology framework including delayed responses by individuals. However, we will insert – only where appropriate – observations that could be related to (or of interest for) investigations on the current pandemics.**

The manuscript is structured as follows. Section two (i) presents the behavioral deterministic models and the related information indices that motivated this study, (ii) it plugs them in the formalism of nonlinear control theory, and (iii) it presents the models main features. Section three introduces the stochastic counterparts of the given models with special focus on the simulation issues that arise when lagged behavioral components are considered, thereby leading to hybrid stochastic systems. Section four explains the experiments. The subsequent sections illustrate the results for different forms of the information index. Concluding remarks follow.

## 2 Deterministic epidemic models with implicit behavior

In this section we present the deterministic counterparts of the stochastic behavioral models used in this paper. We depart from two basic, behavior-implicit, models with a prevalence-dependent response. Then, we discuss the structure of the information index  $M$  and present some main instances of delaying kernels tuning the behavioral response namely, the exponentially fading kernel and a generalisation that we term the acquisition fading kernel. **Then we** set the proposed systems in the framework of nonlinear control theory. Finally, we discuss the main qualitative features of the proposed class of models with special focus on the interplay between the strength of the behavioral response and the **related** time delay.

### 2.1 Prologue: the Capasso-Serio deterministic SIR epidemic model with social distancing

The model by Capasso and Serio [13] extended the classical SIR deterministic epidemic model by Kermack and McKendrick [2] to include a *prevalence-dependent* contact (or transmission) rate, taken as a decreasing function of infective prevalence to mimick individual behavior change as a response to the epidemic threat. The model reads as follows:

$$S' = -\beta(I) \frac{I}{N} S \quad (1)$$

$$I' = \beta(I) \frac{I}{N} S - \gamma I \quad (2)$$

where  $S = S(t)$ ,  $I = I(t)$  represent the number of susceptible and infective individuals at time  $t$ ,  $\gamma > 0$  is the recovery rate, and  $\beta(I)$  is the prevalence-dependent transmission rate which was assumed to obey

$$\beta'(I) < 0$$

$$\beta(0) = \beta_0$$

where  $\beta_0$  represents the transmission rate in the epidemic early phase where no individuals' responses are in place yet. The **key hypothesis** is that the larger the infective prevalence, the stronger will be the agents' response to the epidemic threat by correspondingly reducing their transmission - or contact - rates, that is by expanding their social distancing.

Among the main properties of the model (1)-(2) we mention that: i) an epidemic outbreak will occur only if  $\beta(0) (S(0)/N) > \gamma$  i.e., if the *basic reproduction number* (BRN) relevant for this model exceeds one; ii) for  $t \rightarrow \infty$  it is  $I(t) \rightarrow 0$ , that is, due to the deterministic continuous nature of the model the outbreak only ends for infinitely times. Moreover, the total number of infected subjects obeys the biologically meaningful relationship

$$T = \int_0^{+\infty} \beta(I(t)) \frac{I(t)}{N} S(t) dt = S(+\infty) - S(0)$$

. In substantive terms, the working of the model is rather simple: the inclusion of a simple agents' response increasing with prevalence has a protective role on the community. Compared to the baseline Kermack-McKendrick model, this response can mitigate the epidemic in the event of an outbreak and can - if very intense - even prevent the outbreak occurrence. This simplicity will obviously disappear in presence of more articulated behavioral response.

## 2.2 A deterministic endemic SIR model with social distancing

In [17], the notion of prevalence-dependent social distancing first formulated by Capasso and Serio was extended in a fairly general manner by taking the transmission rate  $\beta$  as a decreasing function of a general *information index* denoted by  $M(t)$  i.e.,  $\beta = \beta(M)$ ,  $\beta'(M) < 0$ . The concept of information index was first introduced in [15], to summarize the entire amount of - current and past - available information about infection spread that agents can use to inform their response to the infection threat. In [17] the idea of social distancing was cast within a basic behavior implicit *information-dependent* SIR model for an endemic - rather than epidemic - infection (therefore also including the vital dynamics of the population). The model read as follows:

$$S' = \mu(N - S) - \beta(M(t))\frac{I}{N}S \quad (3)$$

$$I' = \beta(M(t))\frac{I}{N}S - \gamma I - \mu I, \quad (4)$$

where  $\mu$  represent both the birth and death rates, kept equal to ensure population stationarity over time. The corresponding epidemic variant is obtained by just setting  $\mu = 0$ .

A model using the  $M$  index in a behavior-explicit setting was proposed by [35] within their outbreak model with normal vs altered behavior.

As argued in [17], the information index is a flexible summary of the infection state and can therefore take a wide range of forms. A general form adequate for our present purposes is the following prevalence-dependent form

$$M(t) = \int_0^{+\infty} g(I(t-q))K(q)dq \quad (5)$$

where:  $g(I)$  is an increasing function of the prevalence, and  $K(q)$  is a delaying kernel (or 'memory function'), that is a probability density function that weighting the levels of past prevalence. For the sake of the simplicity, in our subsequent analyses we will use  $g(I) = kI$ .

In [17] it was shown that in the endemic case (i.e., for  $\mu > 0$ ), the model (3)–(4)–(5) has two equilibria: i) the disease-free equilibrium point  $DFE = (N, 0)$ , which is Locally Asymptotically Stable (LAS) and Globally Asymptotically Stable (GAS) provided that the model BRN is lower or at most equal to one, where

$$BRN = \frac{\beta(0)}{\mu + \nu}$$

and a unique endemic equilibrium

$$EE = (S_e, I_e)$$

where  $(S_e, I_e)$  is the unique solution of the following system

$$S_e = N - \left(1 + \frac{\nu}{\mu}\right) I_e$$

$$S_e = N \frac{\mu + \nu}{\beta(I_e)}.$$

The stability of the endemic state depends on the specific model of the information index (5). For example, if the kernel  $K$  obeys  $K(q) = \delta(q)$  where  $\delta(\cdot)$  is the Dirac function then  $M(t) = I(t)$  (i.e., the model is unlagged), and  $EE$  is GAS (see [17, 36] and appendix A).

### 2.3 Plugging social distancing within the framework of nonlinear control theory

Even on an intuitive ground, the information index  $M(t)$  can be read as a feedback control variable (traditionally denoted as  $u(t)$  in nonlinear control theory), since it represents a decision variable that - at the individual level - agents may tune to reduce their hazard of infection. By setting  $\beta_0 = \beta(0)$  and defining

$$\psi(M) = 1 - \frac{\beta(M)}{\beta_0}$$

one can rewrite system (3)–(4)–(5) in the classical form of nonlinear control theory [18]

$$x' = f(x) + \psi(M)g(x) \quad (6)$$

where

$$\begin{aligned} x &= (x_1, x_2) = (S, I) \\ f(x) &= \left( \mu(N - x_1) - \beta_0 \frac{x_2}{N} x_1, \beta_0 \frac{x_2}{N} x_1 - (\nu + \mu)x_2 \right) \\ g(x) &= \left( \beta_0 \frac{x_2}{N} x_1, -\beta_0 \frac{x_2}{N} x_1 \right). \end{aligned}$$

In particular, the case where the information index is assumed to be simply given by the infection prevalence,  $M = I = x_2$ , has a nice control–theory interpretation: it is the deviation of the state  $x_2(t)$  (i.e.  $I$ ) with respect to the ideal reference case of absence of disease, i.e.  $x_2^{ref}(t) = 0$ :  $M = x_2 - x_2^{ref}(t)$ .

### 2.4 Modelling the information index and its temporal kernels

In this section, we state our main assumptions on the transmission rate and the delaying kernel, and provide some relevant control-theoretic interpretations.

As regards the functional form of the transmission rate resulting from the above definition (2.3) i.e.,  $\beta = \beta_0(1 - \psi(M))$ , this could allow to design the control by applying techniques of differential geometry [18,20,21]. However, this approach would not keep under control the key constraint on function  $\psi(M)$  namely:

$$\begin{aligned} 0 &< \psi(M) < 1 \\ \psi(0) &= 0 \\ \psi'(M) &> 0. \end{aligned}$$

Therefore, in our numerical simulations we will use the following phenomenological family of functions for the information-dependent transmission rate  $\beta(M)$ :

$$\beta(M) = \beta_0 \frac{M_{50}^p}{M_{50}^p + M^p} \quad (7)$$

where  $p$  is an integer such that  $p \geq 1$ , and  $M_{50}$  is such that  $\beta(M_{50}) = 0.5\beta(0)$ .

Note that, in the particular case where the information index only includes current information (that is,  $M(t) = I(t)$ ) and for  $p = 1$ , the corresponding force of infection (FOI)  $h(I) = \beta(I)I/N$  (the instantaneous hazard of infection faced by a susceptible individual per time unit) is increasing, whereas for  $p \geq 2$  the FOI is non–monotone with a maximum at  $I = M_{50}$ .

Note moreover that, the fact that the DFE is LAS (GAS, actually) only if  $BRN < 1$  independently of the memory kernel, i.e. only if in absence of the behavioral effect the disease cannot remain endemic, implies that one cannot design a  $\psi(M)$  such **that the** DFE is stabilized.

As regards the memory kernel, we will consider three main cases:

- The memoryless case  $K(q) = \delta(q)$ . In this case the information index  $M$  only includes current information.
- The exponentially fading memory kernel (EFK):

$$K(q) = a \text{Exp}(-aq).$$

This represents the often realistic case where past information receives a decreasing weight. In this case, the integro–differential system can be reduced to ordinary differential equation (ODEs) by adding the further ODE:

$$M'(t) = a(g(I) - M) \quad (8)$$

- The acquisition–fading memory kernel (AFK):

$$K(q) = \frac{a_1 a_2}{a_2 - a_1} (\text{Exp}(-a_1 q) - \text{Exp}(-a_2 q)).$$

This represents another possibly realistic scenario where no information (about the infection state) is available at current time, then it is gradually acquired up to a maximum before starting fading out exponentially, as in the EFK case.

Even in this case the integro–differential model (3)-(4)-(5) can be reduced to ODEs by the additional equations

$$Z'(t) = a_1(g(I) - Z) \quad (9)$$

$$M'(t) = a_2(Z - M) \quad (10)$$

Note that if the two sub-processes of information acquisition **and memory** fading occur at the same relative speed  $a_1 = a_2 = a$ , then our 2-dimensional acquisition–fading kernel reduces (via the limit  $a_2 \rightarrow a_1$ ) to the so-called *Erlang*(2,  $a$ ) =  $ate^{-at}$  kernel [37].

It is immediate to verify that:

- In the case of the exponentially fading kernel, the memory index/control variable  $M(t)$  can be read as the output of a low–pass filter with characteristic cut–off frequency  $a$  that is applied to the ‘signal’  $g(I(t))$ . Under our work assumption  $g(I) = kI$  the transfer function  $H(\lambda)$  between the input  $I$  and the output  $M$  reads as follows

$$H(\lambda) = \frac{\widehat{M}(\lambda)}{\widehat{I}(\lambda)} = \frac{k\lambda}{a + \lambda} = \widehat{K}(\lambda),$$

where  $\lambda \in \mathbb{C}$  and  $\widehat{M}(\lambda)$ ,  $\widehat{I}(\lambda)$  and  $\widehat{K}(\lambda)$  are the Laplace transforms of, respectively,  $M(t)$ ,  $I(t)$  and  $K(q)$ ;

- In the case of the acquisition–fading kernel, the memory index  $M(t)$  can be read as the output of the application of two low–pass filters in series (with characteristic frequencies  $a_1$  and  $a_2$ ) that is applied to the ‘signal’  $g(I(t))$ . In particular, in the case  $g(I) = kI$ , the transfer function between the input  $I$  and the output  $M$  reads as follows

$$H(\lambda) = \frac{\widehat{M}(\lambda)}{\widehat{I}(\lambda)} = \frac{ka_1 a_2}{(a_1 + \lambda)(a_2 + \lambda)} = \widehat{K}(\lambda). \quad (11)$$

Therefore, in our problems, the processes of fading and (if it is the case) acquisition of the memory are analogous to the error with memory connected to the measure of a feedback signal by means of a mechanic or electronic device [19]. This is a typical scenario in control theory, where the controller often acquires the information on the variables to be controlled by means of measure devices. The latter can have their own internal dynamics and moreover, are also aimed at reducing extrinsic noises that are characterized by large frequencies.

Adopting again the control-theory formalism, the final structure of the considered family of models (6) is as follows:

- In case of EFK:

$$x = (x_1, x_2, x_3) = (S, I, M)$$

$$f(x) = \left( \mu(N - x_1) - \beta_0 \frac{x_2}{N} x_1, \beta_0 \frac{x_2}{N} x_1 - (\nu + \mu)x_2, a(g(x_2) - x_3) \right)$$

$$g(x) = \left( +\beta_0 \frac{x_2}{N} x_1, -\beta_0 \frac{x_2}{N} x_1 \right), 0$$

- In case of AFK:

$$x = (x_1, x_2, x_3, x_4) = (S, I, Z, M)$$

$$f(x) = \left( \mu(N - x_1) - \beta_0 \frac{x_2}{N} x_1, \beta_0 \frac{x_2}{N} x_1 - (\nu + \mu)x_2, a_1(g(x_2) - x_3), a_2(x_3 - x_4) \right)$$

$$g(x) = \left( +\beta_0 \frac{x_2}{N} x_1, -\beta_0 \frac{x_2}{N} x_1 \right), 0$$

## 2.5 Properties of the previous general deterministic endemic model with social distancing

We discuss here the main stability properties of the endemic state of model (6) under the different hypotheses considered on the form of the delaying kernel.

**First, in the case of EFK, it was proven in [17] that the Endemic Equilibrium is LAS. Moreover, in [38] by means of an appropriate Liapunov function it was shown that under the additional condition  $(M\beta(M))' > 0$  the Endemic Equilibrium is globally stable.** Additionally, in [17], the case of an *Erlang*<sub>2</sub> type memory was considered. As already pointed out, this memory represents a special case of the acquisition-fading kernel where the two sub-processes of acquisition and fading occur at the same rate  $a_1 = a_2 = a$ . The corresponding transfer function reads as follows

$$\frac{\widehat{M}(\lambda)}{\widehat{I}(\lambda)} = \frac{ka^2}{(a + \lambda)^2},$$

In this case, the EE can be either LAS or can be destabilized by a Hopf bifurcation (see [17]).

Let us now consider model (6) for a general delaying kernel  $K(\tau)$ . Denoting as  $\widehat{K}(\lambda)$  the corresponding Laplace transform and linearizing at EE one gets:

$$\lambda^2 + c_1\lambda + c_0 + d(\lambda + \mu) = \widehat{K}(\lambda) = 0,$$

where:

$$c_1 = \mu + \beta(I_e)I_e > 0$$

$$c_0 = \beta^2(I_e)I_e S_e > 0$$

$$d = -\beta'(I_e)I_e S_e > 0$$

Under an AFK,  $\widehat{K}(\lambda)$  is given by formula (11) and it can easily be rewritten as follows:

$$K(\lambda) = \frac{a_1 a_2}{(\lambda + a_1)(\lambda + a_2)} = \frac{p}{\lambda^2 + \sigma\lambda + p}$$

where  $\sigma = a_1 + a_2$  and  $p = a_1 a_2$ .

This finally yields the following characteristic polynomial:

$$\lambda^4 + q_1\lambda^3 + q_2\lambda^2 + q_3\lambda + q_4 = 0$$

where

$$q_3 = \sigma + c_1 > 0$$

$$\begin{aligned}
q_2 &= c_1\sigma + p + c_0 > 0 \\
q_1 &= c_0\sigma + (c_1 + d)p > 0 \\
q_0 &= (d\mu + c_0)p > 0
\end{aligned}$$

Since all the coefficients are positive, applying the Routh–Hurwitz criterion we get the following local stability conditions of the EE:

$$H(p, \sigma) = q_1 q_2 q_3 - q_1^2 - q_3^2 q_0 > 0, \quad (12)$$

In particular, if

$$H(p, \sigma) < 0, \quad (13)$$

then EE becomes unstable through Hopf a bifurcation at the locus

$$H(p, \sigma) = 0.$$

As well known, the Hopf Theorem only provides local information in the parametric space. In other words, it only states that in a parametric neighborhood of the bifurcation point the solution of the system is a periodic solution of small amplitude. Therefore, in our case this will hold for values of  $(p, \sigma)$  in the region  $H(p, \sigma) < 0$  that are sufficiently close to points where  $H(p, \sigma) = 0$ . In order to draw more general information about points in the interior of the region  $H(p, \sigma) < 0$  one can take advantage of the Yakubovitch–Efimov–Fradkov theorem [39].

To apply the theorem, note preliminarily that the system is bounded and has two unstable equilibrium points, and the DFE has as stable manifold the segment  $l = (x_1, 0, 0, 0)$  with  $x_1 \in [0, N]$  that corresponds to the trivial case of absence of disease in the target population. The biological and geometrical nature of the stable manifold of DFE implies that there are not heteroclinic orbits connecting EE to DFE.

As a consequence, from the Yakubovitch–Efimov–Fradkov theorem (see corollary 1 of [39]) if  $H(p, \sigma) < 0$  then system undergoes self-sustained oscillations, not necessarily periodic, called *Yakubovitch oscillations* or *Y-oscillations* [40]. We report below two useful remarks.

**REMARK** *The above results can easily be extended to delays in the form of  $n \geq 3$  first order low-pass filters in series.*

**REMARK** *Note that, in our case (and in the generalization) the Y-oscillations are fully induced by the behavior i.e., by the control. In control language, this is the consequence of the fact that the control is based on a 'measure' coming from a double filtering of the prevalence, i.e from the indirect nature of the control. Note however that the choice of the weight  $K(q)$  is primarily based on behavioral arguments.*

As regards the analytical form of the condition (12), this is cumbersome and gives no particular hints, so that also (13) become of scarce practical relevance. Therefore, we numerically computed the Jacobian and its 4 associated eigenvalues at the endemic equilibrium. This allowed us to plot (see figure 1) the stability and instability regions in function of the two key parameters: i)  $T_1 = 1/a_1$  which is the average acquisition time of information; and  $T_2 = 1/a_2$  which is the average fading time. In particular, figure 1 reports the contour plot of the following function:

$$z(T_1, T_2) = \text{Max} \left( 0, \text{Max}_{k \in \{1, 2, 3, 4\}} (\text{Re}(\lambda_k(T_1, T_2))) \right)$$

## 2.6 Deterministic modeling of the epidemic case ( $\mu = 0$ )

In the rest of the present work we will mostly focus on the purely epidemic case ( $\mu = 0$ ). This case was not previously analyzed with the exception of the case  $M = I$ , i.e. the Capasso–Serio model.

For the general model with a generic information-dependent  $M$  index it is easy to see that it exists a disease-free

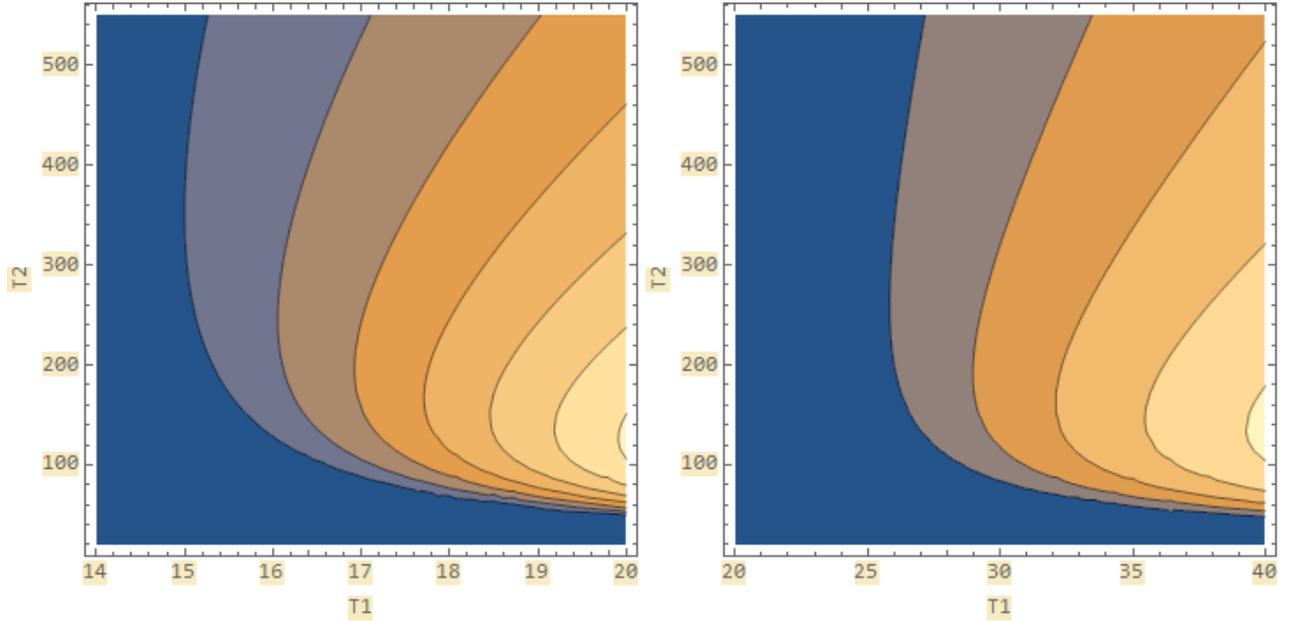


Figure 1: Model (6) with acquisition–fading kernel: contour plot of the stability and instability regions in terms of parameters  $T_1$  and  $T_2$  as determined by the function  $z(T_1, T_2)$ . Dark Blue: LAS region. Lighter blue and other colors: instability region. In the instability region the different colors indicates the amplitude of  $z(T_1, T_2)$ , i.e. the real part of the eigenvalue (of the Jacobi Matrix) having the largest real part. The Basic Reproduction Number is set to either  $BRN = 20$  (left panel) or to  $BRN = 2$  (right panel). Other parameters:  $\mu = 1/(365.25 * 75)$ ,  $\nu = 1/7$ .

equilibrium  $x_2 = 0$  and that  $x_1(t)$  tends to an equilibrium value that depends on the function  $\psi(M(t))$  and on the dynamics of  $M(t)$ , which is in turn determined by the delay kernel. A fact not previously pointed out in the literature is that under appropriate behavioral responses multiple behavior-induced epidemic waves can arise even in the course of a single epidemic. **We will numerically illustrate these deterministic results in the Appendix.**

### 3 Stochastic modelling of social distancing under generic $M$ -indices

In this section we introduce the stochastic counterparts of the previous deterministic models with social distancing and discuss the simulation issues that arise when the behavioral component is involved. We start from the simpler case where no memory effects are considered, and then extend to the case of hybrid systems arising under memory effects.

#### 3.1 No memory effects: stochastic version of the Capasso-Serio model

As is well-known, the deterministic models of infectious diseases can be considered as a good approximation of the true underlying stochastic systems **only if** the population size is large enough so that stochastic oscillations **around** the mean pattern become negligible. **In the absence of behavioral effects, if the population is small then the system is modelled by a birth-death Markov process** [24].

In the absence of memory effects i.e., when considering the stochastic version of the Capasso-Serio model, the resulting stochastic model remains a birth–death stochastic processes [24, 26, 27, 41].

In this case, the model is simply the outcome of two transition processes and related events, namely:

i) transmission: a susceptible subject becomes infectious due to adequate social contacts with infectious subjects. The probability of a transmission event during the infinitesimal interval  $(t, t + dt)$  is:

$$Prob((S(t + dt), I(t + dt)) = (S(t) - 1, I(t) + 1)) = \beta(I) \frac{I}{N} S dt \quad (14)$$

However, unlike the basic stochastic SIR model, the probability in formula (14) is not bilinear.

ii) removal: an infectious individuals is removed from the infectious compartment due to recovery and immunity acquisition. The probability of a removal event during  $(t, t + dt)$  is:

$$Prob((S(t + dt), I(t + dt)) = (S(t), I(t) - 1)) = \gamma I dt \quad (15)$$

The stochastic simulation of birth and death Markov processes is elegantly done by the Gillespie algorithm [26, 27, 41] which in our case works as follows. Let us suppose that the  $n$ -th event occurred at time  $t_n$  so that the system after the event has state values

$$(S, I) = (S_n, I_n).$$

Therefore, denoting the time of the next event as

$$t_{n+1} = t_n + \tau$$

then  $\tau$  is such that

$$\gamma I_n \tau + \beta(I_n) \frac{I_n}{N} S_n \tau = r_n$$

where  $r_n$  is exponentially distributed. This yields:

$$\tau = \frac{r_n}{\gamma I_n + \beta(I_n) \frac{I_n}{N} S_n} \quad (16)$$

The type of the next event, which will either be the contagion of a susceptible or the removal of an infectious subject) is determined by randomization through a Bernoulli experiment based on the two complementary probabilities

$$\rho_{rec} = \frac{\gamma I_n}{\gamma I_n + \beta(I_n) \frac{I_n}{N} S_n}$$

$$\rho_{cont} = 1 - \rho_{rec}$$

### 3.2 Memory effects and simulation of hybrid stochastic systems

In the presence of a general memory effects the behavior of the system becomes more complicated. Indeed, the stochastic evolution of the discrete state variables  $(S, I)$  has to be coupled with an integral equation describing the dynamics of  $M(t)$ , which is a continuous one during each inter-events intervals  $(t_n, t_{n+1})$ . In particular, under our hypotheses on the memory **kernel**s, the dynamics of  $M$  is given by a system of ordinary differential equations depending on  $I_n$ .

As a consequence, the system is no longer a classical birth and death process, but it belongs to the class of the so-called hybrid models [28] where Stochastic Automata are coupled with deterministic processes by means of time varying propensities. Note however, that in our case the coupling is mutual. Indeed, a stochastic birth–death model (for infection spread) is coupled with an ODE (or a system of ODEs) model for the memory, yielding for  $M$  a continuous *piecewise deterministic model* that depends on the stochastic epidemic model. Indeed, in between two stochastic events, the dynamics of  $M(t)$  is fully deterministic i.e.,  $M(t)$  is a *piecewise deterministic process*. Overall, the above–described model can be formalized similarly to the class of Stochastic Hybrid Automata defined in [31].

Operatively, the infinitesimal probability of a contagion event in  $(t, t + dt)$ , given by:

$$Prob((S(t + dt), I(t + dt)) = (S(t) - 1, I(t) + 1)) = \beta(M(t)) \frac{I}{N} S dt \quad (17)$$

is time-dependent. This requires to resort to a version of the Gillespie algorithm with time-varying propensities, as in [28].

Namely, the time to the  $(n + 1)$ –th event is determined by the following equation:

$$\gamma I_n \tau + \frac{I_n}{N} S_n \int_{t_n}^{t_n + \tau} \beta(M(t)) dt = r_n \quad (18)$$

In the general case, the inequality

$$0 \leq \beta(M) \leq \beta_0,$$

implies that

$$0 \leq \int_{t_n}^{t_n + \tau} \beta(M(t)) dt \leq \beta_0 \tau.$$

As a consequence, from (18) and provided that  $I_n > 0$  (the case  $I_n = 0$  is trivial because  $I = 0$  is an adsorbing state for the system) we can infer that

$$\tau_L \leq \tau \leq \tau_R,$$

where:

$$\tau_L = \frac{r_n}{\gamma I_n + \beta_0 \frac{I_n}{N} S_n}$$

$$\tau_R = \frac{r_n}{\gamma I_n}$$

which is an useful bound for the numerical simulations.

**REMARK** *The bounds  $\tau_L$  and  $\tau_R$  are not constant: they change at each step of the simulation algorithm.*

To compute  $\tau$  it is necessary to know the model linking the continuous state variable  $M(t)$  to the discrete state variables  $(S, I)$ .

Once  $\tau$  has been determined, the type of the next event is chosen (as in the classical Gillespie algorithm) by the following probabilities:

$$\bar{\rho}_{rec} = \frac{\gamma I_n}{\gamma I_n + \beta(M(t_n + \tau)) \frac{I_n}{N} S_n}$$

$$\bar{\rho}_{cont} = 1 - \bar{\rho}_{rec}$$

### 3.2.1 Numerical determination of $\tau$

Let us define

$$\psi(\tau) = \int_{t_n}^{t_n + \tau} \beta(M(t)) dt$$

Then one has to numerically solve the following equation

$$\gamma I_n \tau + \frac{I_n}{N} S_n \psi(\tau) = r_n \quad (19)$$

in the interval

$$\tau_L \leq \tau \leq \tau_R,$$

by some numerical algorithms. We then have two main cases. The first one occurs when function  $\psi(\tau)$  is available in closed form. In the case where  $\beta(M(t))$  can be analytically integrated in  $(t_n, t_{n+1})$ , then the equation (19) can be solved by means of well known standard iterative algorithms, such as the Newton–Raphson and the bisection methods (see appendix).

For the sake of notational simplicity we define

$$f(\tau) = \gamma I_n \tau + \frac{I_n}{N} S_n \psi(\tau) - r_n$$

thus to determine  $\tau$  we must solve

$$f(\tau) = 0$$

Note also that:  $f(\tau_L) < 0$  and  $f(\tau_R) > 0$ .

Instead, if  $\psi(\tau)$  is not analytically known, one has to numerically approximate the integral that appears in the time-varying Gillespie algorithm. For this case, we propose the algorithm described by the following steps:

- Split the interval  $[0, \tau_R]$  in  $L \gg 1$  points, and define

$$h = \frac{\tau_R}{V}$$

$$\theta_k = kh$$

where  $k = 0, 1, \dots, V$ . Note that :

$$f(\theta_0) = f(0) = -r_n$$

$$\theta_V = \tau_R$$

- Approximate  $\psi(\theta_{k+1})$  by some standard strategy, such as:

$$\psi(\theta_{k+1}) \approx \psi(\theta_k) + \frac{h}{2} (\beta(M(t_n + \theta_k)) + \beta(M(t_n + \theta_{k+1}))).$$

Note that this implies that

$$f(\theta_{k+1}) \approx f(\theta_k) + \gamma I_n h + \frac{I_n}{N} S_n \frac{h}{2} (\beta(M(t_n + \theta_k)) + \beta(M(t_n + \theta_{k+1}))) \quad (20)$$

- Since  $\tau > \tau_L$ , we have to iterate computations (without checking if we have found the solution  $\tau$ )  $f(\theta)$  at least until the step

$$k_L = \left\lceil \frac{\tau_L}{h} \right\rceil$$

where  $[x]$  is the integer part of  $x$ . E.g.:  $[3.99] = 3$ .

- Suppose that the step  $q \geq k_L$  had been reached and that

$$f(\theta_q) < 0.$$

Then go to the next step  $\theta_{q+1}$  and compute  $f(\theta_{q+1})$  by means of the formula (20). The following three cases are possible: i)  $|f(\theta_{q+1})| < Tol$ , where  $Tol$  is a sufficiently small tolerance, so you can stop; ii)  $f(\theta_{q+1}) < -Tol$ , thus one must continue; iii)  $f(\theta_{q+1}) > Tol$ , which implies that then the solution  $\tau$  is such that

$$\tau \in (\theta_q, \theta_{q+1})$$

Since  $h$  is sufficiently small, one can use the approximation

$$\tau \approx \frac{\theta_q + \theta_{q+1}}{2} = \theta_q + \frac{h}{2}$$

**REMARK** *The most sensitive point in the previous algorithm is therefore that of appropriately choosing  $V$ , which must be sufficiently large. In our simulations, we have chosen  $V$  in a way such that the time step  $h$  is always kept smaller than 0.1 days. Moreover, one must remember that since  $\tau_R$  changes at each step of the Gillespie algorithm, the same holds for  $V$  Once determined  $\tau$  the type of the next event is chosen by the following probability:*

$$\bar{\rho}_{rec} = \frac{\gamma I_n}{\gamma I_n + \beta(M(t_n + \tau)) \frac{I_n}{N} S_n}$$

$$\bar{\rho}_{cont} = 1 - \bar{\rho}_{rec}$$

### 3.3 The case of the exponentially fading memory

In the **cases of EFK and of AFK kernels**, the coupling between the discrete and the continuous parts of the hybrid model is relatively simpler than in the general case. Indeed, as we have seen, in our examples the model of  $M(t)$  reduces, respectively, to one or two linear ordinary differential equations.

In this section and in the following one, we will show how  $\beta(M(t))$  can be analytically defined in all event intervals  $(t_n, t_{n+1})$  for both the EFK and the AFK. For the exponentially fading memory kernel, we will also provide the analytical form of  $\psi(\tau)$ , which allows to significantly speeds up the computations.

In the case of an exponentially fading memory, one has that in each interval  $(t_n, t_{n+1})$  between two consecutive events,  $M(t)$  is given by the following linear differential equation

$$M'(t) = a(g(I_n) - M) \quad (21)$$

$$M(t_n) = M_n \quad (22)$$

$$M_0 = 0. \quad (23)$$

since we suppose that the epidemics starts at  $t = 0$ . The solution of the ODE problem (21)–(22) is:

$$M(t) = g(I_n) + (M_n - g(I_n))Exp(-a\theta) \quad (24)$$

where

$$\theta = t - t_n.$$

The function  $\psi(\tau)$  is given by the integral

$$\psi(\tau) = \beta_0 \int_0^\tau \frac{M_{50}^p}{M_{50}^p + (g(I_n) + (M_n - g(I_n))Exp(-a\theta))^p} d\theta$$

Defining

$$K = \frac{M_{50}}{|M_n - g(I_n)|}, B = \frac{g(I_n)}{|M_n - g(I_n)|}, \sigma = \text{sign}(M_n - g(I_n))$$

yields

$$\psi(\tau) = \beta_0 K^p \int_0^\tau \frac{1}{K^p + (R + \sigma Exp(-a\theta))^p} d\theta$$

Defining the new variable

$$y = R + \sigma Exp(-a\theta)$$

yields

$$\psi(\tau) = \beta_0 K^p \int_{R+\sigma}^{R+\sigma Exp(-a\tau)} \frac{-1}{a} \frac{1}{(K^p + y^p)} \frac{1}{y - R} dy.$$

As for the above integral note the following: i) if  $\sigma = 1$  then we have that the upper integration limit is smaller than the lower integration limit (due to  $a > 0$ ) but since  $y = R + exp(-a\theta) \Rightarrow y > R$  thus the function to be integrated is negative and the integral is positive; ii) if  $\sigma = -1$ , the upper integration limit is greater than the lower integration limit and since  $y = R - exp(-a\theta) \Rightarrow y < R$ , the integrand function turns out to be positive, so that the integral will be positive as well (and, in principle, analytically solvable).

Further setting  $y = Kz$  one gets:

$$\psi(\tau) = \beta_0 \int_{r+\rho}^{r+\rho Exp(-a\tau)} \frac{-1}{a} \frac{1}{(1+z^p)} \frac{1}{z-r} dz$$

where  $r = R/K$  and  $\rho = \sigma/K$ .

Namely, for the case  $p = 1$  one has:

$$\psi(\tau) = \beta_0 \frac{-1}{a(1+r)} \left| \log(z-r) - \log(1+z) \right|_{r+\rho}^{r+\rho Exp(-a\tau)}$$

yielding

$$\psi(\tau) = \beta_0 \frac{-1}{a(1+r)} \log \left( \frac{1+r+\rho}{Exp(a\tau)(1+r)+\rho} \right) > 0$$

or equivalently (the following version reduces numerical errors in the case  $\exp(-a\tau) \ll 1$ )

$$\psi(\tau) = \beta_0 \frac{-1}{a(1+r)} \left( -a\tau - \log \left( \frac{1+r+\rho \exp(-a\tau)}{1+r+\rho} \right) \right) \quad (25)$$

For the case  $p = 2$  one has:

$$\psi(\tau) = \beta_0 \frac{-1}{a} \frac{1}{1+r^2} \left| \log(z-r) - (1/2)\log(1+z^2) - r \operatorname{Arctan}(z) \right|_{r+\rho}^{r+\rho \exp(-a\tau)}$$

For  $p \geq 3$ , the expression of  $\psi(\tau)$  becomes untractable.

Of course, once  $\tau$  has been computed, one must set

$$M_{n+1} = g(I_n) + (M_n - g(I_n)) \exp(-a\tau)$$

### 3.4 The case of the acquisition–fading kernel

In the case of the acquisition–fading kernel, one has that in each interval  $(t_n, t_{n+1})$  the solution  $(Z(t), M(t))$  to the pair of addition differential equations, can be computed as follows. First,  $Z(t)$  is defined as the solution of the following differential equation

$$Z'(t) = a_1(g(I_n) - Z) \quad (26)$$

$$Z(t_n) = Z_n \quad (27)$$

$$Z_0 = 0 \quad (28)$$

The solution of the above ODE problem is:

$$Z(t) = g(I_n) + (Z_n - g(I_n)) \exp(-a_1\theta) \quad (29)$$

where

$$\theta = t - t_n.$$

Next,  $M(t)$  is the solution of

$$M'(t) = a_2(Z(t) - M) \quad (30)$$

$$M(t_n) = M_n. \quad (31)$$

$$(32)$$

implying:

$$M(t) = g(I_n) + \left( \frac{a_2}{a_2 - a_1} (Z_n - g(I_n)) \right) \exp(-a_1\tau) + \left( M_n - \frac{a_2 Z_n}{a_2 - a_1} + \frac{a_1 g(I_n)}{a_2 - a_1} \right) \exp(-a_2\tau) \quad (33)$$

Note that in this case the integral defining  $\psi(\tau)$  cannot be analytically solved even in the case  $p = 1$ .

**REMARK** *In this case, once  $\tau$  has been numerically computed, one must set*

$$Z_{n+1} = g(I_n) + (Z_n - g(I_n)) \exp(-a_1\tau) \quad (34)$$

$$M_{n+1} = g(I_n) + \left( \frac{a_2}{a_2 - a_1} (Z_n - g(I_n)) \right) \exp(-a_1\tau) + \left( M_n - \frac{a_2 Z_n}{a_2 - a_1} + \frac{a_1 g(I_n)}{a_2 - a_1} \right) \exp(-a_2\tau) \quad (35)$$

Simulations of scenarios where the memory kernel is the AFK are shown in the appendix, for the sake of the readability.

### 3.5 Algorithm for the localization of multiple 'process' peaks in the simulated time series

In our work we used Matlab 2019b for model simulation and the analysis of results. Given our focus on multiple - stochastic - epidemic waves, we briefly present the approach for the localisation of (multiple) epidemic peaks. The purpose of the algorithm is that of allowing to separate true 'process' peaks i.e., peaks induced by the behavioral response, from spurious oscillations due to the stochastic nature of the problem. The localization of the first peak was done by finding the maximum number of infectious individuals, because the first peak always resulted the higher one. The latter (purely due to susceptible depletion) is a somewhat straightforward consequence of the simplicity of the model. The localization of the second peak (and subsequent ones) was more complicated due to the noise of stochastic simulations. Firstly, we filtered the time courses and analysed only cases with 3 and more points. Noise was filtered by the Gaussian-weighted moving average filter (we applied Matlab function: *smoothdata*) with window size equal to 100 points. Then we accepted as true peaks, maxima arising in the smoothed data course according to the following criteria (i) minimum height of peak equal to 100 infectious, (ii) minimum time distance between subsequent peaks of 5 days (Matlab function: *findpeaks*). As matter of fact, since the appearance of multiple peaks was much neater in the case  $BRN = 15$  (compared to  $BRN = 2$ ), later on we will mainly report our results on multiple peaks for this case.

## 4 Results. Stochastic epidemics and behavior. The memoryless case

In this section we report the main results of a range of stochastic simulation experiments, with special focus on the effects of different hypotheses on the shape of the social distancing response function  $\beta(M)$  on the system behavior.

### 4.1 Simulation strategy, assignment of input parameters and main simulation outputs

In all our experiments we simulated epidemic outbreaks in a fully susceptible population, under the following assumptions on input parameters: i) the population size is set to  $N = 10000$ ; ii) the infection average duration is set to one week, implying  $\gamma = 1/7\text{days}^{-1}$ ; iii)  $M_{50\%}$  is allowed to take three possible values:  $M_{50\%} \in \{10, 50, 200\}$ . As regards the basic reproduction number, we considered two widely different cases: 1)  $BRN = 15$  corresponding to a highly transmissible infection, such as measles; 2)  $BRN = 2$ , corresponding to a moderately transmissible infection as it may be the case e.g., of seasonal influenza. The corresponding baseline transmission rates (in the absence of any behavioral response) are  $\beta_0 = 15/7\text{days}^{(-1)}$ , and  $\beta_0 = 2/7\text{days}^{(-1)}$ , respectively. In the deterministic SIR epidemic model in a wholly susceptible population, the considered figures of the  $BRN$  would yield a total proportion of people eventually infected i.e., a *final attack rate* of the outbreak - well above 99 percent for  $BRN=15$  and around 78 percent for  $BRN=2$ . As for the main outputs of stochastic simulation we primarily looked at the simulated probability density functions (PDF) of (i) the final attack rate, (ii) the epidemic extinction time, (iii) the maximum prevalence, and to the temporal realizations (or realised paths) of the prevalence function. In particular, in order to adequately reproduce the relevant probability density functions, the model simulation is replicated  $N_{sim} = 100000$  times.

### 4.2 Results: behavioral responses without memory effects

#### 4.2.1 Effects of $M_{50\%}$ in the stochastic case

In this subsection, we considered the memoryless case with  $p = 1$ , in order to characterize the role of parameter  $M_{50\%}$ .

The stationary PDFs of the final attack rate and of the extinction time are shown in the upper and lower panels of figure 2 (left panels:  $BRN = 15$ , right panels:  $BRN = 2$ ), respectively. Each panel reports three different histograms corresponding to the three values considered of  $M_{50\%}$  and, for comparison purposes, also the one corresponding to the benchmark case of the classic stochastic SIR model without behavioral response.

Concerning the final attack rates (upper panels of 2), we note that for both  $BRN = 15$  and  $BRN = 2$ , the simulated PDFs are always bimodal. The mode for low values of the infected subjects reflects stochastic extinction occurring soon after infection introduction, when the behavioral effects are still negligible. Obviously stochastic extinction is rather unlikely for  $BRN = 15$ , whereas it occurs with a much higher probability percentages for the 'classical SIR'; instead for  $BRN = 2$  this lower mode is very large.

We may note that even for  $M_{50\%} = 200$  for both  $BRN = 15$  and  $BRN = 2$  the effect is remarkable. Qualitatively, both for  $BRN = 15$  and for  $BRN = 2$ , the mean value of the total number of people that acquired the infection decreases if  $M_{50}$  decreases, but the variance of the distributions increases. For  $BRN = 2$ , at  $M_{50} = 10$  the distribution is characterised by a large initial peak followed by exponential-like decrease.

As for the extinction time (lower panels of 2), the simulated PDFs are also bimodal, corresponding to "initial" and "final" (i.e., after the outbreak occurred) extinction. Interestingly, for  $BRN = 15$  the mean values of the extinction time are increasing as  $M_{50}$  decreases (i.e., when outbreaks occur, the stronger the behavioral response the longer is the epidemic), and also the variances become larger. Moreover, at  $M_{50} = 10$  the PDF is more markedly bimodal. For  $BRN = 2$  and  $M_{50\%} = 10$  the PDF again shows an initial peak followed by exponential-like decrease.

Note that, for  $BRN = 15$  the location of PDFs of the extinction time for  $M_{50\%} = 50$  and especially for  $M_{50\%} = 10$  is so rightward that the chosen approximation of neglecting vital dynamics could be questioned. The same holds for  $BRN = 2$  and  $M_{50\%} = 50$ . This suggests that behavioral responses could significantly delay the epidemic extinction time and, not paradoxically, sustain the endemicity of the infection. Both these effects have been observed during the current COVID-19 pandemic.

Table 1: Statistics related to figure 2

Figure	Panel	$BRN$	$p$	$M_{50\%}$	mean	SD	median	min	max
1	A	15	1	(classic SIR)	93.27	25.05	100	0.01	100
1	A	15	1	200	88.97	23.65	95.23	0.01	96.72
1	A	15	1	50	86.56	23.42	92.87	0.01	95.50
1	A	15	1	10	81.21	22.97	87.70	0.01	92.75
1	B	2	1	(classic SIR)	39.80	39.83	0.27	0.01	83.07
1	B	2	1	200	24.12	24.35	0.16	0.01	58.56
1	B	2	1	50	16.61	17.68	0.08	0.01	49.58
1	B	2	1	10	1.33	2.35	0.04	0.01	92.75
1	C	15	1	(classic SIR)	69	20	71	<1	173
1	C	15	1	200	173	53	178	<1	371
1	C	15	1	50	434	131	454	<1	797
1	C	15	1	10	1260	381	1344	<1	1995
1	D	2	1	(classic SIR)	83	80	44	<1	297
1	D	2	1	200	289	297	31	<1	1212
1	D	2	1	50	456	496	20	<1	1966
1	D	2	1	10	116	192	13	<1	1705

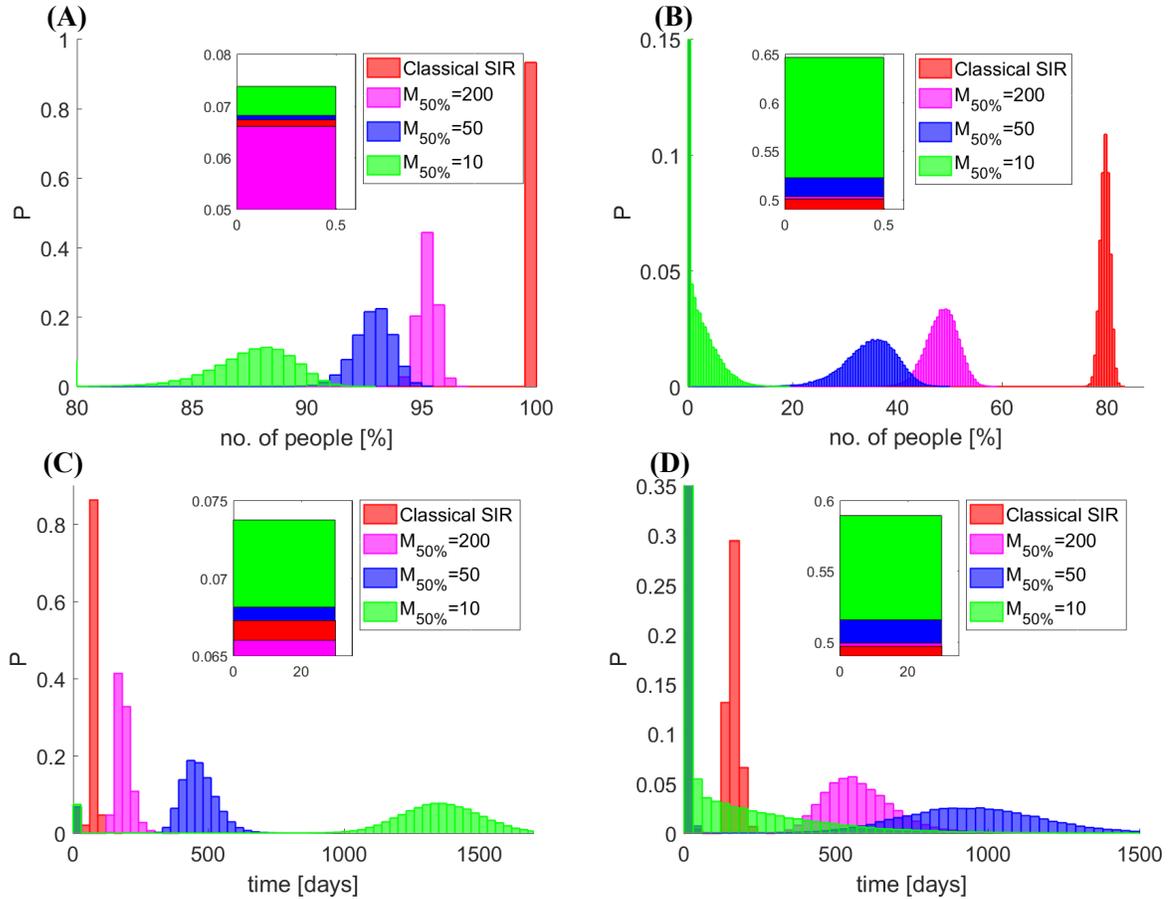


Figure 2: The memoryless case for  $p = 1$ : effects of  $M_{50\%}$  on the simulated distributions of the final attack rate and of the extinction time. Left Panels:  $BRN = 15$ ; Right Panels:  $BRN = 2$ . Higher Panels: PDFs of the final attack rate. Lower Panels: PDFs of the extinction time.

The PDFs of the maximum reached prevalence and of the time of occurrence of such maximum are reported in the upper and the lower panels of figure 3, respectively. For both  $BRN = 15$  and  $BRN = 2$ , the mean maximum epidemic size (upper panels of figure 3) and its variance are rapidly increasing as  $M_{50}$  increases. For  $BRN = 15$ , the mean and the variance of the time of occurrence of such maximum (lower panels of figure 3) are increasing as  $M_{50}$  decreases. Instead, for  $BRN = 2$ , the mean is non-monotone with  $M_{50\%}$ , whereas the variance increases. Summary statistics of the various PDFs of figures 2 and 3 are reported, respectively, in Tables 1 and 2.

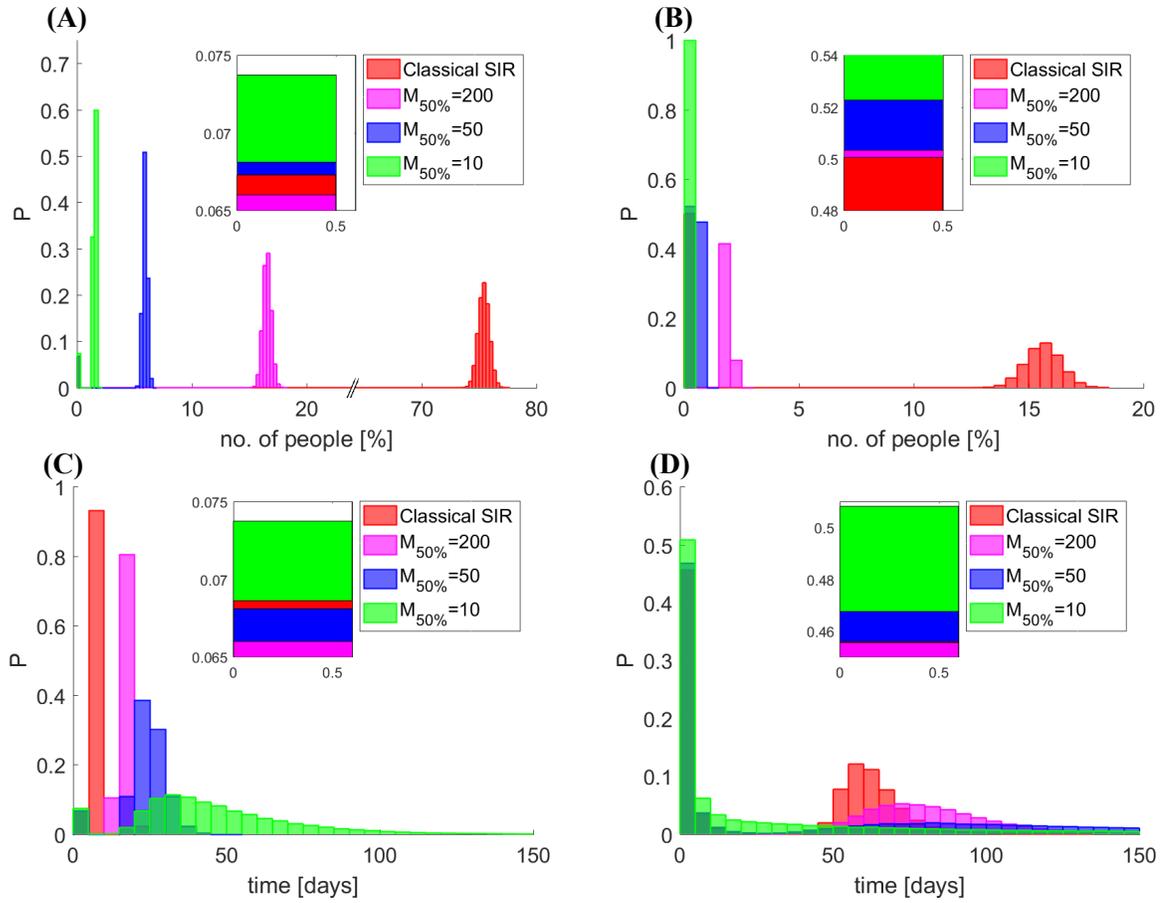


Figure 3: The memoryless case for  $p = 1$ : impact of  $M_{50\%}$ . Left Panels:  $BRN = 15$ ; Right Panels:  $BRN = 2$ . Higher Panels: effects of  $M_{50\%}$  on the PDF of the maximum prevalence. Lower Panels: effects of  $M_{50\%}$  on PDF of time at the maximum prevalence.

Table 2: Statistics related to figure 3

Figure	Panel	$BRN$	$p$	$M_{50\%}$	mean	SD	median	min	max
2	A	15	1	(classic SIR)	70.34	18.90	75.37	0.01	77.59
2	A	15	1	200	15.45	4.12	16.51	0.01	18.15
2	A	15	1	50	5.48	1.49	5.86	0.01	6.86
2	A	15	1	10	1.42	0.40	1.52	0.01	1.97
2	B	2	1	(classic SIR)	7.81	7.82	0.10	0.01	18.44
2	B	2	1	200	0.93	0.92	0.07	0.01	2.58
2	B	2	1	50	0.33	0.33	0.05	0.01	1.03
2	B	2	1	10	0.09	0.09	0.03	0.01	0.37
2	C	15	1	(classic SIR)	6	2	6	<1	11
2	C	15	1	200	16	4	17	<1	25
2	C	15	1	50	23	8	24	<1	50
2	C	15	1	10	45	25	41	<1	208
2	D	2	1	(classic SIR)	32	31	26	<1	135
2	D	2	1	200	41	43	17	<1	210
2	D	2	1	50	66	87	9	<1	846
2	D	2	1	10	51	99	5	<1	1197

#### 4.2.2 The effects of the shape of the response function

The effects of the shape of the behavioral response function  $\beta(M)$ , in absence of delays, is summarized by parameter  $p$ . Of course, even the threshold  $M_{50\%}$  plays a fundamental role. Figure 4 reports a sample of the simulated time-series of the infectious prevalence with  $M_{50\%} = 200$ , for different values of  $p$  (we keep  $BRN = 15$ ). This shows three main effects of increasing  $p$ . The first is related to the fact that, for  $p > 1$ , the function  $\beta(I; p)$  decreases slower with  $I$ , thereby increasing extinction time. This effects is better illustrated by the corresponding PDFs in figure 5. The second and the third are more interesting: i) the maximum epidemic size decreases with  $p$  and ii) the disease prevalence remains at large values for a larger time w.r.t. the case  $p = 1$ . Particularly interesting is the case  $p = 100$ , whose behavior can be better understood by resorting to the following approximation of function  $\beta(I)$

$$\beta(I) \approx \begin{cases} \beta(0) & \text{if } I < M_{50\%} \\ 0.5 & \text{if } I = M_{50\%} \\ 0 & \text{if } I \geq M_{50\%} \end{cases}$$

Initially, since for  $I < M_{50\%}$  it holds  $\beta(I) \approx \beta(0)$ , the epidemic grows free up to the level  $I = M_{50\%}$ . Sooner or later  $I$  switches to the value  $M_{50\%}$  where  $\beta = 0.5\beta(0)$ . However, sooner or later the next contagion event will occur, so that prevalence  $I$  reaches the value  $M_{50\%} + 1$  and the transmission rate abruptly falls to  $\beta = 0$ . This means that the contagion stops and prevalence will remain constant at the level  $I = M_{50\%} + 1$  until the next removal event occurs, so that the system now switches back and forth around  $M_{50\%}$ . In other words,  $I(t)$  will experience for a long time small stochastic oscillations around  $I = M_{50\%}$  until when, through a series of removal events disease extinction is reached.

By decreasing  $M_{50\%}$  the minimal extinction time remarkably increases. This is better shown in figure 5. As for the total number of infected people, the corresponding simulated PDF gradually switches rightwards as  $p$  increases (see supplementary figures). For the sake of completeness, Figures 8 and 9 report analogous results for the case  $BRN = 2$ . Note in particular the much larger stochasticity arising compared to  $BRN = 15$ , a well-known fact.

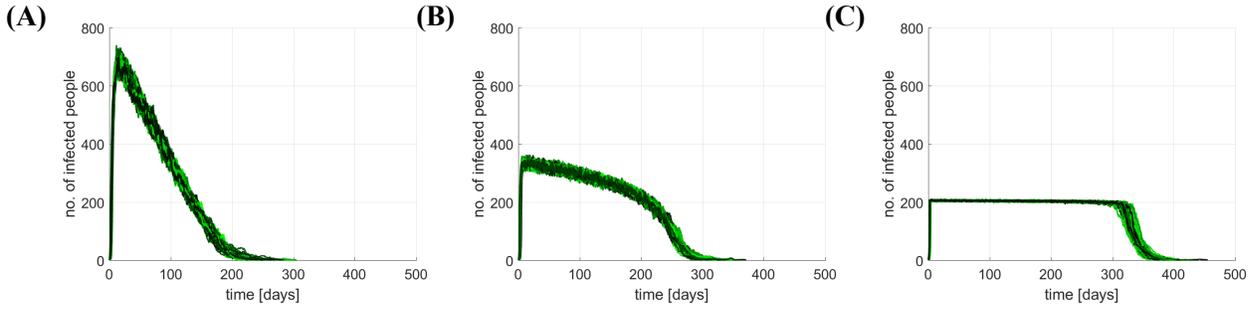


Figure 4: The stochastic model in the memoryless case. Effects of parameter  $p$  on the time-course of infectious prevalence  $I(t)$  (a sample of 100 realizations) for  $BRN = 15$  and  $M_{50\%} = 200$ . (A)  $p = 2$  (B)  $p = 5$  (C)  $p = 100$

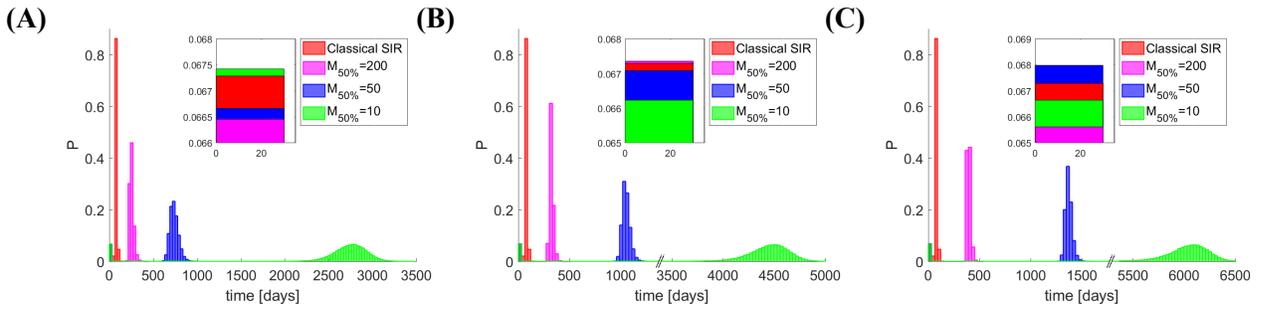


Figure 5: The stochastic model in the memoryless case. Effects of parameter  $p$  on the simulated PDF of the infection extinction time for  $BRN = 15$  and different values of  $M_{50\%}$ . A)  $p = 2$  (B)  $p = 5$  (C)  $p = 100$

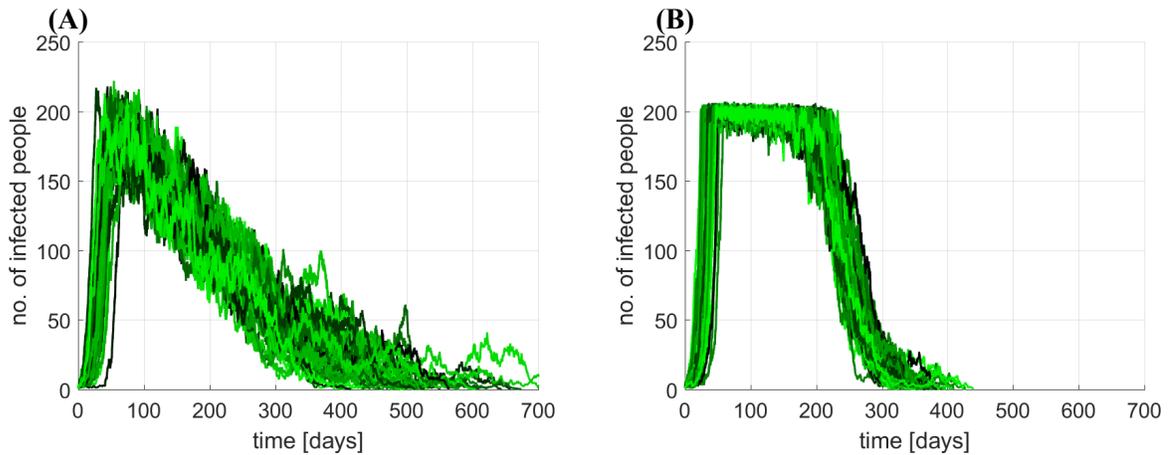


Figure 6: The stochastic model in the memoryless case. Effects of parameter  $p$  on the time-course of infectious prevalence  $I(t)$  (a sample of 100 realizations) for  $BRN = 2$  and  $M_{50\%} = 200$  (A)  $p = 2$  (B)  $p = 100$

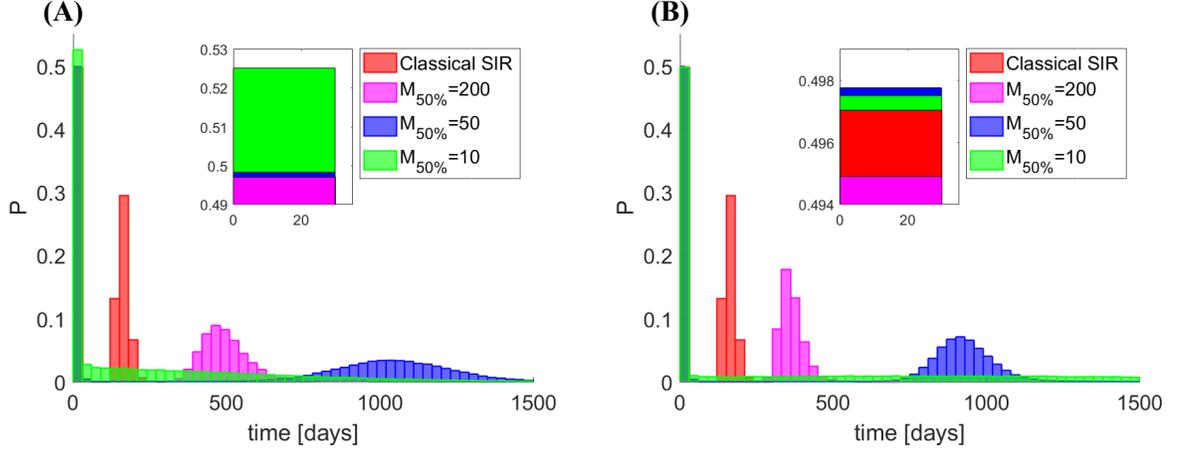


Figure 7: PDF of infection extinction time for  $BRN = 2$ : impact of  $p$  and  $M_{50\%}$ . (A)  $p = 2$  (B)  $p = 100$ . Other values of  $p$  produces PDFs similar to that of panel (B).

## 5 Results. Exponentially fading kernels and multiple stochastic epidemic waves

In this section we investigate the effects of delayed social distancing response function  $\beta(M)$  on the epidemic course. In particular, we will study the EFK case by considering two possible values for the parameter  $a$  tuning the delayed response:  $a = 0.1/day$  and  $a = 0.05/day$ , corresponding to an average response delay of 10 and 20 days, respectively.

We assessed the effects of a delayed behavioral response in both the stochastic and in the deterministic case (see appendix), for  $BRN = 15$  (figure 8) and  $BRN = 2$  (figure 9) and for distinct values of  $p$  ( $p = 1, 2, 10$ ). **Both** deterministic and stochastic simulations show that multiple epidemic waves can actually occur, as a consequence of the behavioral response of individuals, when lagged information is used in the response process. The different cases reported in the figures clarify mechanisms and determinants of these waves. Notably, these waves cannot occur in the absence of the delay. In particular, the oscillatory pattern will be more marked the larger the amplitude of the time delay (i.e., the lower  $a$ ), the sudden the behavioral response (i.e., the larger  $p$ ), and the larger the response threshold  $M_{50\%}$ .

### 5.1 Statistical features of simulated epidemics

In what follows, we reports the main statistics of our simulative result where multiple waves occur. To do so, in the computation of the simulated PDFs we did not include those realizations where no second peak was clearly detectable. In the present experiments we keep  $M_{50\%} = 50$ , whereas  $a \in \{0.05, 0.1\}$ ,  $p = 1, 2, 10$ ,  $BRN \in \{2, 15\}$ . Figure (10) shows the PDF of infectious prevalence at the first epidemic peak. We note that of course in all cases there is a peak at low prevalence levels, corresponding to a rapid extinction of the epidemic. A second peak occurs for larger values of prevalence. We note that for  $BRN = 2$  (bottom part of the Figure) the PDFs are located at low values of prevalence, they are overlapping and the average values are increasing with  $p$ . On the contrary, for  $BRN = 15$  (top part of the Figure), the PDFs are located at large values of the prevalence, they are not overlapping and the average values are decreasing with  $p$ . Obviously, the larger the delay in the behavioral response, the larger the expected magnitude at the first peak. Figure (11) reports the PDF of the infectious prevalence at the *second* epidemic peak. As detailed before, this is done only for the case of a large value of the  $BRN$  ( $BRN = 15$ ), in order to magnify occurrence and detectability of the second wave. For  $a = 0.1$  the prevalence at the second time-series peak decreases with  $p$ , whereas for  $BRN = 15$  and  $a = 0.05$  the histograms overlap but their location

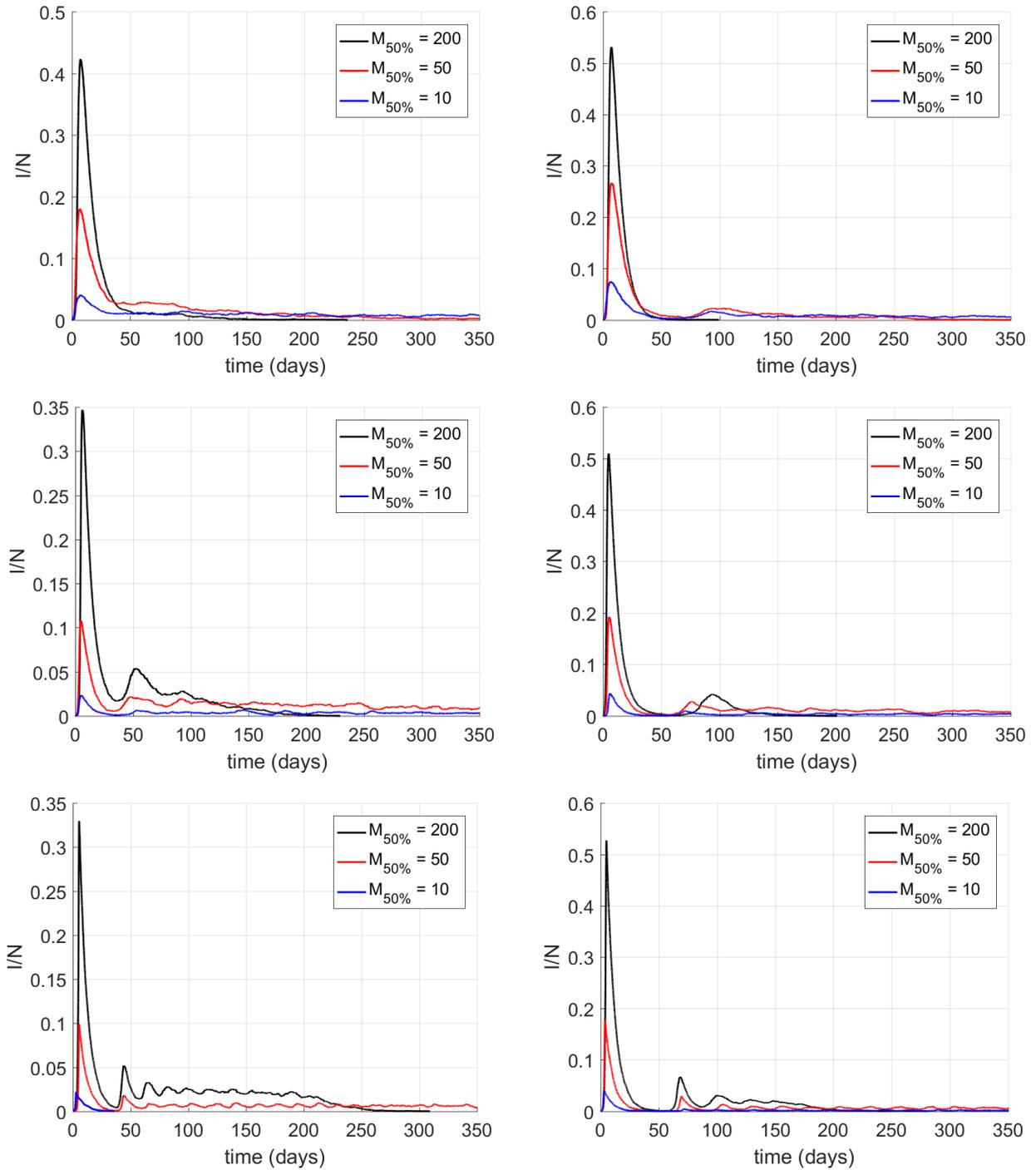


Figure 8: Stochastic multiple epidemic waves of infectious prevalence for an exponentially fading behavioral response, under  $BRN = 15$ , for different values of the memory rate  $a$ , and of parameters  $p$  and  $M_{50\%}$  tuning the behavioral response. Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

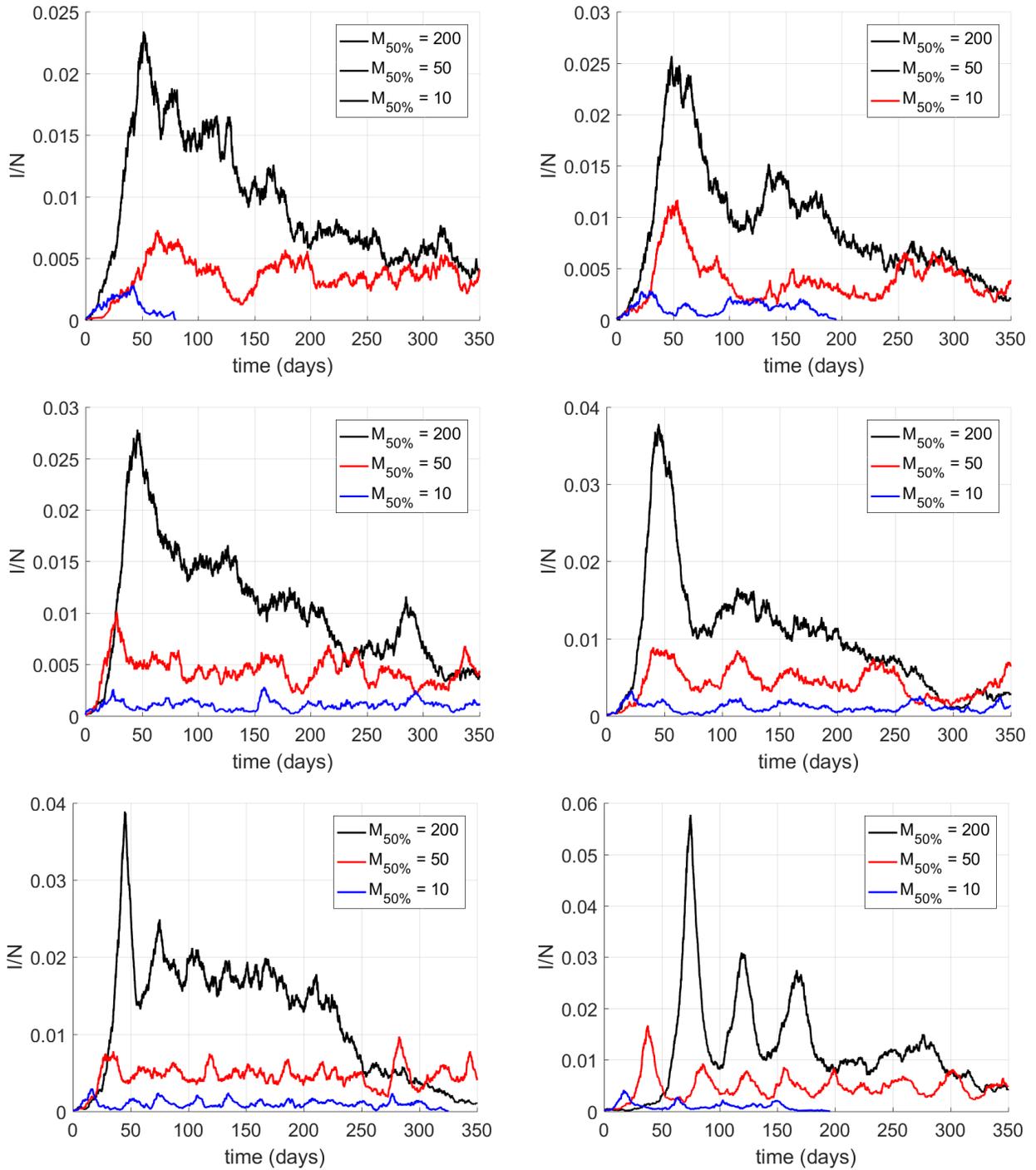


Figure 9: Stochastic multiple epidemic waves of infectious prevalence for an exponentially fading behavioral response, under  $BRN = 2$ , for different values of the memory rate  $a$ , and of parameters  $p$  and  $M_{50\%}$  tuning the behavioral response. Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

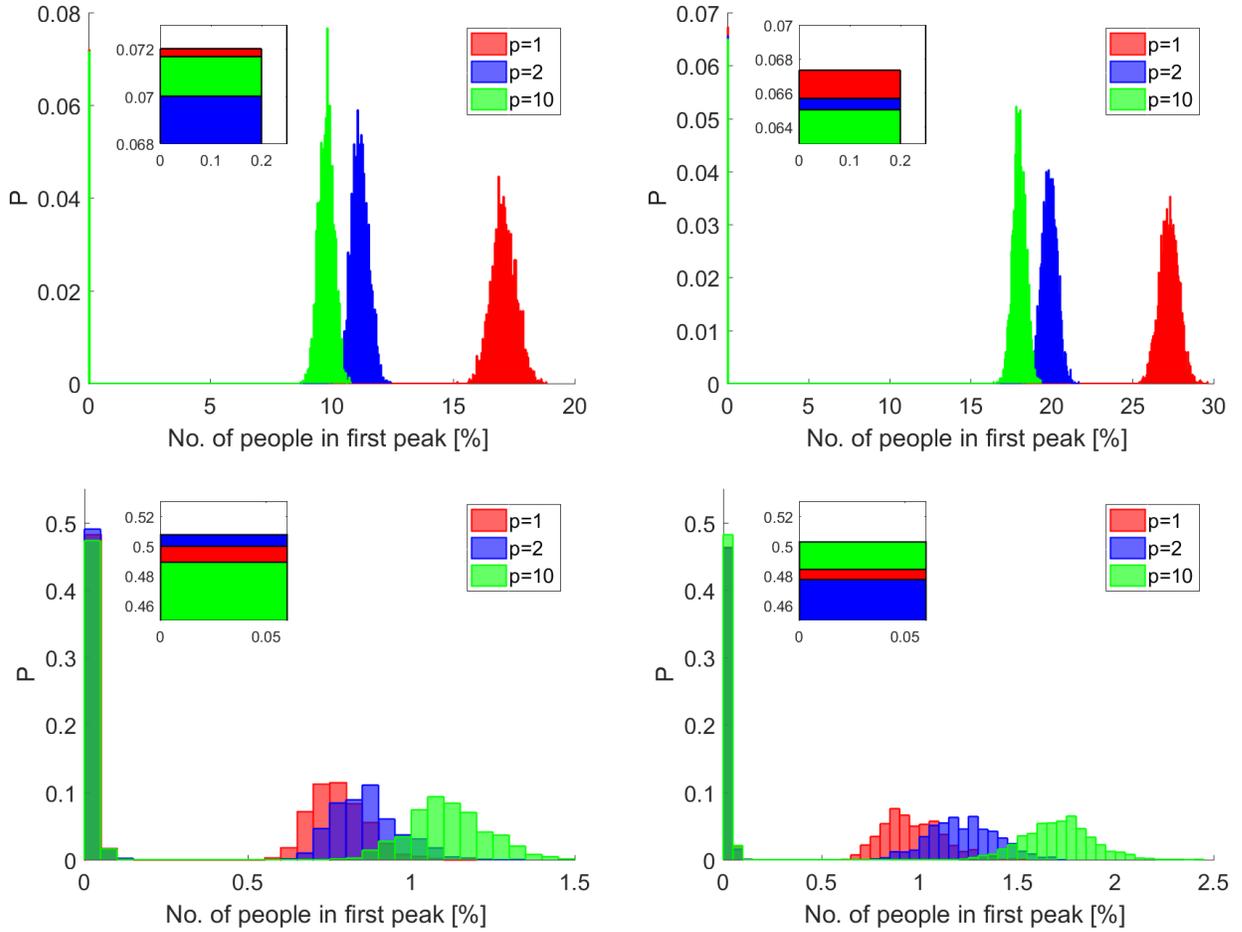


Figure 10: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the PDF of the infectious prevalence at the first epidemic peak for  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $BRN = 15$ ; lower panels:  $BRN = 2$ .

appears somewhat non monotone with  $p$ . As for the simulated PDF of the time of occurrence of the first epidemic

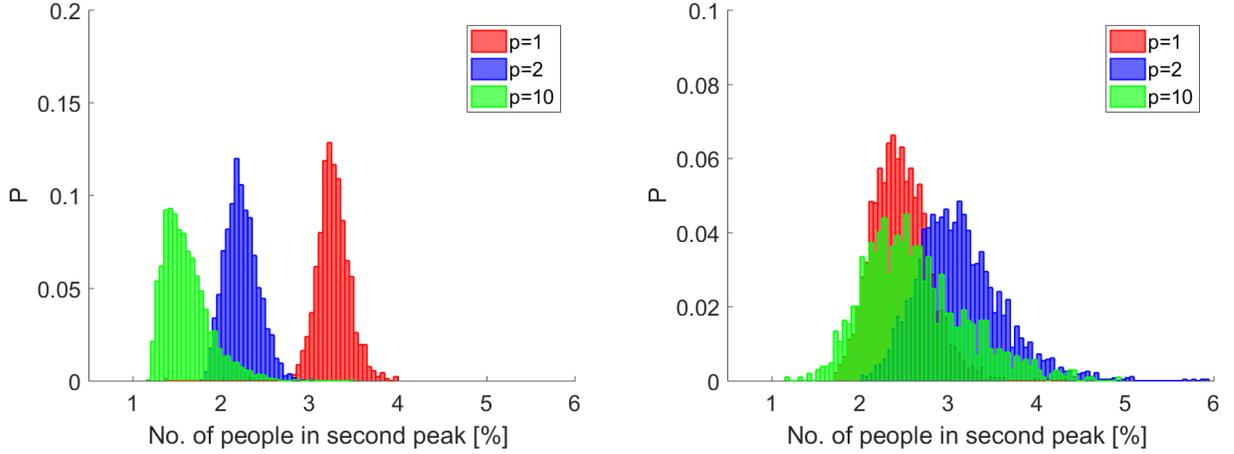


Figure 11: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the PDF of the infectious prevalence at the second epidemic peak for  $BRN = 15$ ,  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panel:  $a = 0.1$ ; right panel:  $a = 0.05$ .

peak (Figure 12), we note that (disregarding the initial peak due to stochastic extinction) for  $BRN = 15$ , the PDFs are well-separated with an average peak-time decreasing in  $p$  (from 4 to 8 days), while for  $BRN = 2$ , the PDFs largely overlap especially for longer behavioral response delays. In particular, the effects of the response delay are essentially negligible on the timing of the first peak. Instead, the corresponding PDFs of the time at the *second* epidemic peak (still drawn for  $BRN = 15$  only) reveal (Figure 13) a marked effect of the time-delay in the behavioral response.

We report in the Table (3) the main summary statistics of the PDFs shown in the previous figures.

Table 3: Statistics related to figure 13 - times of second peak

Figure	Panel	$BRN$	$a$	$p$	mean	SD	median	min	max
13	A	15	0.1	1	70	7	69	53	195
13	A	15	0.1	2	55	8	53	46	170
13	A	15	0.1	10	44	1	44	41	63
13	B	15	0.05	1	104	5	104	86	132
13	B	15	0.05	2	79	2	79	71	96
13	B	15	0.05	10	71	1	71	67	76

For the sake of completeness, we report also the PDFs of the time to epidemic extinction (Figure 14). Still disregarding initial stochastic extinction, it can be noted that no relevant role was played by the delay in the behavioral response. However, the dependency on  $p$  and the  $BRN$  remains more complicated. As for the epidemic *final attack rate*, the corresponding simulated PDFs (Figure 15), suggest - as usual - a major role of the shape of the behavioral response (tuned by  $p$ ) for  $BRN = 2$ . For  $BRN = 15$  a counter-intuitive role of the delay emerges at high values of  $p$ : for  $p = 10$  and  $a = 0.05$ , the PDF of the final attack rate has two nontrivial modes. One of these is the traditional one at high levels of the FAR. The second one occurs for levels of the FAR around value 20%. This second mode reflects the interesting fact by which a too delayed behavioral response cannot prevent the occurrence of the outbreak but, if the response is sufficiently intense, it can halt it and bringing it to extinction.

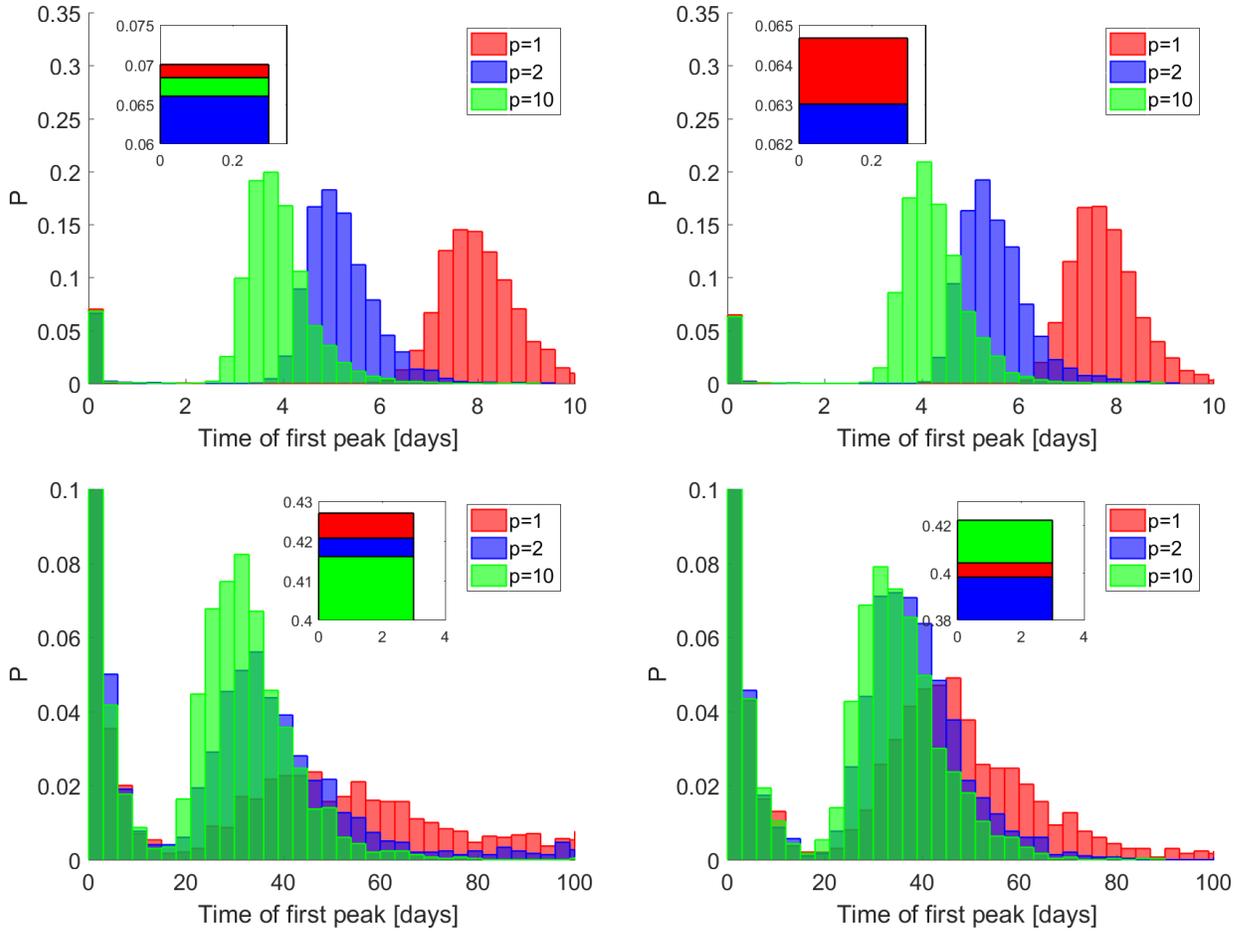


Figure 12: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the simulated PDF of the time at the first epidemic peak for  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $BRN = 15$ ; lower panels:  $BRN = 2$ .

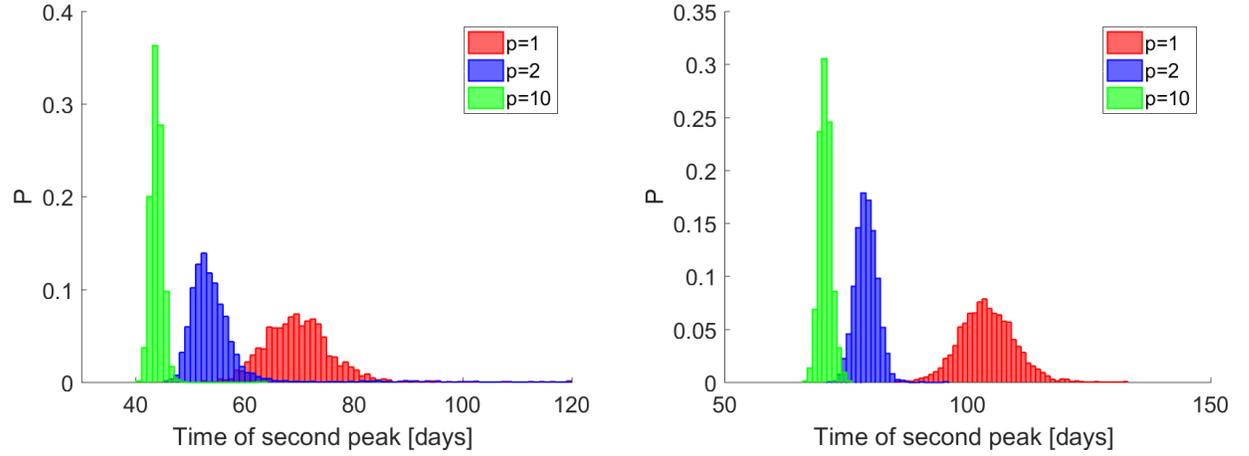


Figure 13: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the PDF of the time at the second epidemic peak for  $BRN = 15$ ,  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panel:  $a = 0.1$ ; right panel:  $a = 0.05$ .

Finally, the PDF of the number of nontrivial peaks (Figure 16) in the more interesting case ( $BRN = 15$ ), reveals the number of (detectable) peaks increases with the delay of the behavioral response (that is, for  $a = 0.05$  compared to  $a = 0.1$ ).

## 5.2 The filtering effect

For  $BRN = 2$ , figure 17 compares, for a single realization of the stochastic process, the simulated time series of the information index  $M(t)$  and that of the infectious prevalence  $I(t)$ . **The figure well shows** the analogy between the concept of an exponentially fading memory and the concept of 'low-pass filter' used in system control theory. Calling these quantities as the (original) 'signal' ( $I(t)$ ) and the 'filtered signal' ( $M(t)$ ), it happens that the signal is noisy and the filtered signal is *smooth*, for  $a = 0.01$  and for  $(a, p) = (0.1, 1)$ , whereas in all the other cases also the filtered signal is noisy, especially for  $a = 0.05$ . This is a natural consequence of the fact that the 'low-pass frequency'  $a$  is too close to the typical frequency of the stochastic events of  $I(t)$ . **Case  $BRN = 15$  is shown in the Appendix.**

## 6 Concluding Remarks

This article investigated the effects of agents' behavioral responses on the dynamics of stochastic SIR models for epidemic outbreaks. The agents' behavioral response was represented by means of an information-dependent transmission rate specified as a function of an appropriate *information index*, as first proposed in [15, 17]. In particular, agents were assumed to respond either to the current or past prevalence of infection, where the latter was specified according either to an exponentially fading memory or to an acquisition-fading memory. **This resulted in a family of hybrid stochastic models. In order to numerically simulate specific models belonging to the family, we needed a suitable extension of classical Gillespie algorithm. In particular, we developed analytical formulas valid for two classes of models.**

Our analysis provides a thorough classification of the possible outcomes of behaviorally-modulated stochastic SIR epidemics depending, besides basic reproduction numbers, on: (i) the form and strength of the behavioral response, (ii) the information time-lag with which this behavioral response is enacted by individuals.

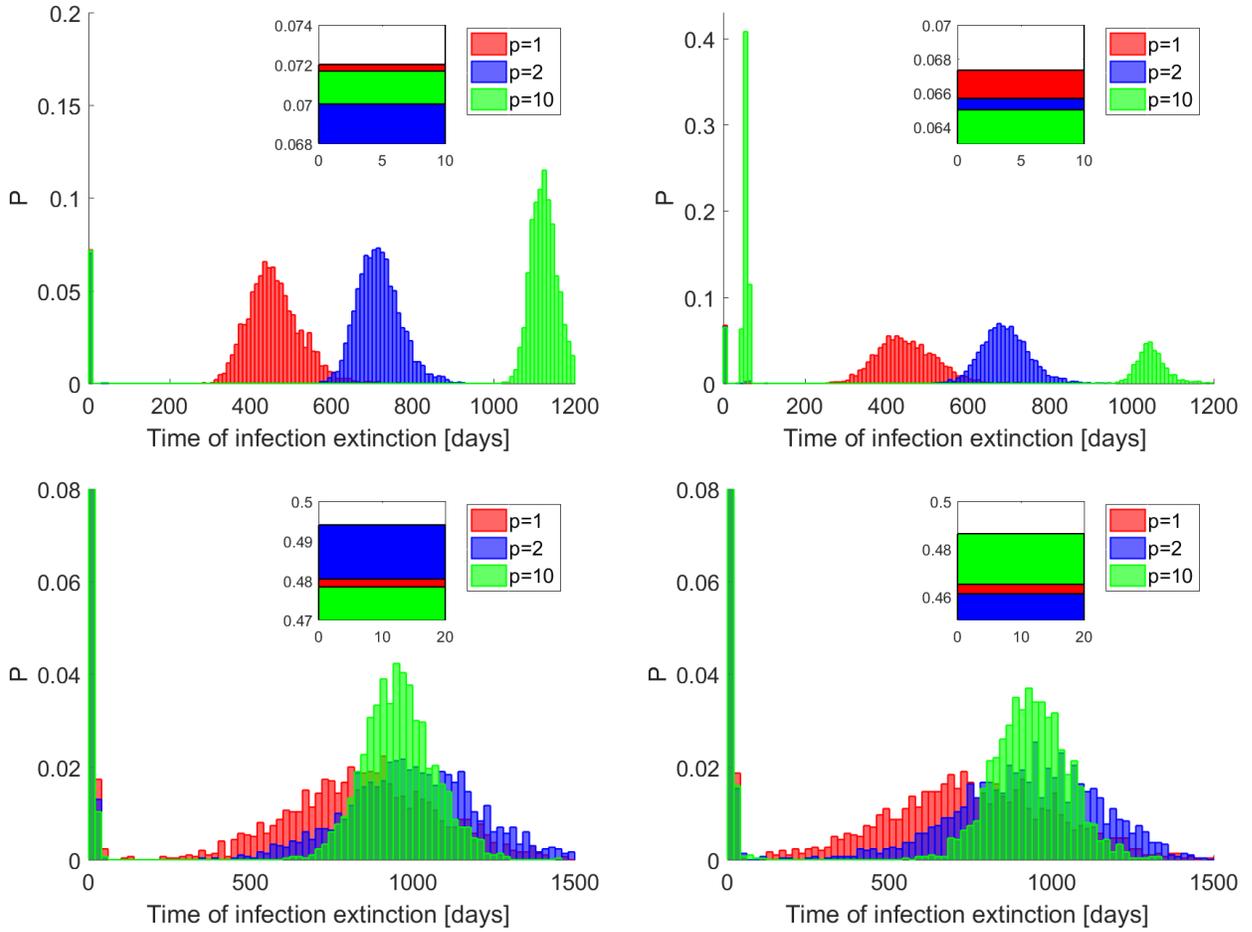


Figure 14: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the simulated PDF of epidemic extinction time for  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $BRN = 15$ ; lower panels:  $BRN = 2$ .

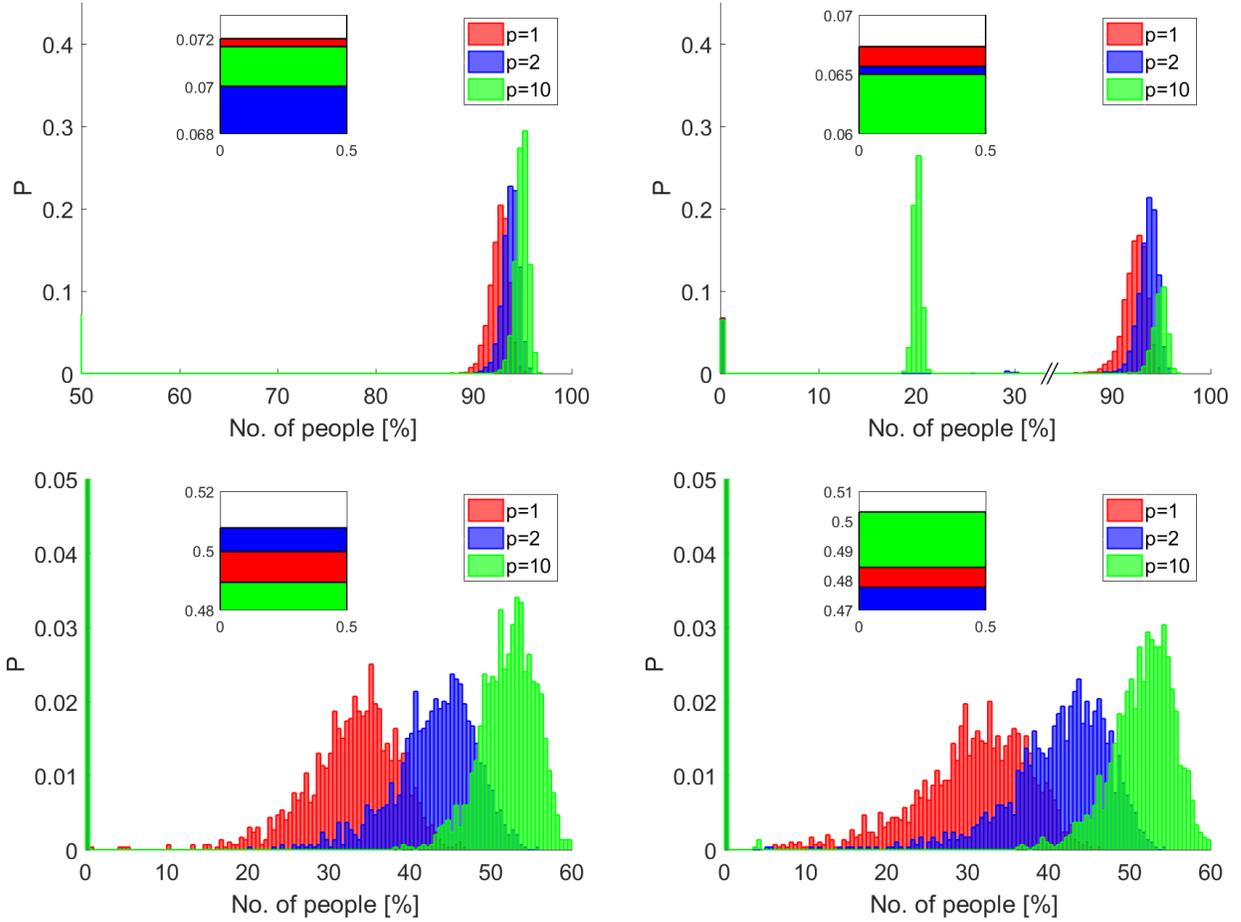


Figure 15: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the simulated PDF of the final attack rate for  $p = 1, 2, 10$  and  $M_{50\%} = 50$ . Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $BRN = 15$ ; lower panels:  $BRN = 2$ .

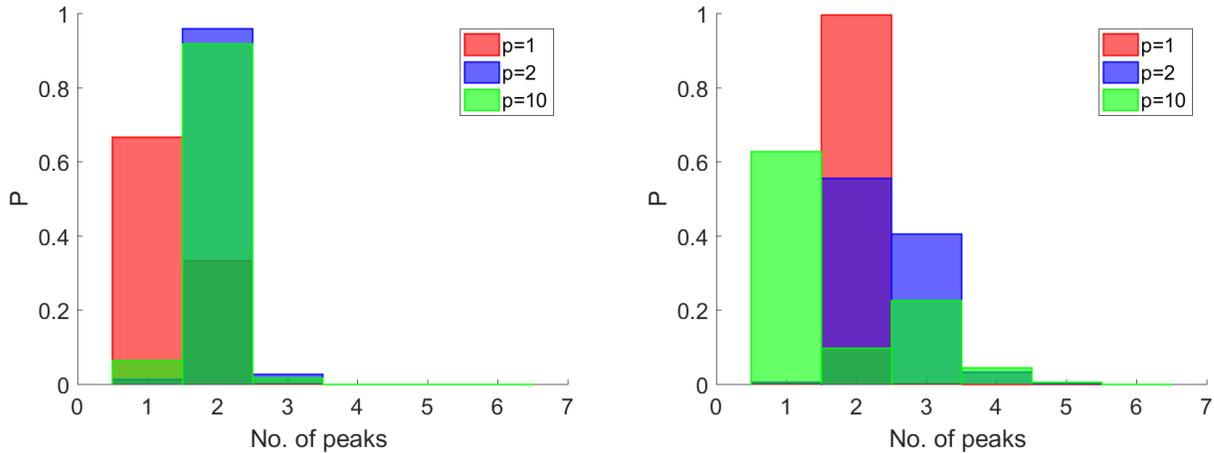


Figure 16: The stochastic model in the case of an exponentially fading memory kernel. Histograms of the simulated PDF of the number of epidemic peaks for  $BRN = 15$ ,  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF. Left panel  $a = 0.1$ , Right panel  $a = 0.05$ .

This offers a range of theoretical results. For example, even in the absence of delays in the behavioral response (i.e., when the agents' response is only based on current prevalence), the behavioral response may substantially increase the variance of the extinction time, thereby favoring infection persistence (especially for highly transmissible infections). This behavior-triggered persistence might link with other factors, such as demographics or the appearance of new variants of the pathogen, favoring endemicity, as might sadly be the case with COVID-19. These phenomena becomes richer when the behavioral response is also modulated by past information. In this case, we showed that multiple epidemic waves can result from delayed agents' responses already in the underlying deterministic models, and remain also in the stochastic formulation.

In general the topic of recurrent waves, especially in relation to endemic infectious diseases (e.g. for measles) has been a major topic of mathematical epidemiology over decades [6]. Instead, the issue of multiple waves during an epidemic outbreak has become popular during the pandemic preparedness in the early 2000s. This started from retrospective studies on the 1918–1919 Spanish flu pandemic and developed after the H1N1 2009 pandemic. A variety of explanations were proposed, ranging from exogenous changes in transmissibility [42], to policy intervention [43], to the presence of multiple strains [44], up to behavioral changes [45]. **Recently, elegant analytical work has stressed the role of inter-regional commuting in inducing multiple (behavior-unrelated) epidemic waves** [46]. The multiple waves observed during the COVID-19 pandemic in industrialized countries were primarily due to the activation and relaxation of social distancing measures during the first pandemic year and to the onset of new variants of concern during 2021 [47], but also behavioral explanations were proposed [48].

**In relation to the previous debate, the main innovation of this work lies in the fact that it represents the first general theoretical assessment of the role of delayed behavioral responses in promoting multiple waves in the most basic stochastic setting. This in turn required to consider an innovative formal setting, namely that of Stochastic Hybrid Automata, whose continuous component mirrors (in our case) a distributed delay/memory.**

Of course, the proposed model is minimalistic: many refinements can be considered. **First, the model represents spontaneous individual behaviors in the simplest manner i.e., implicitly by the information index approach** [15, 17]. As such, the model might also represent a situation where a government uses the same information index to implement a policy of forced behavior change in individuals as e.g., a lockdown. **Modeling behavior in an explicit manner** [7, 15, 16] can refine such scenarios.

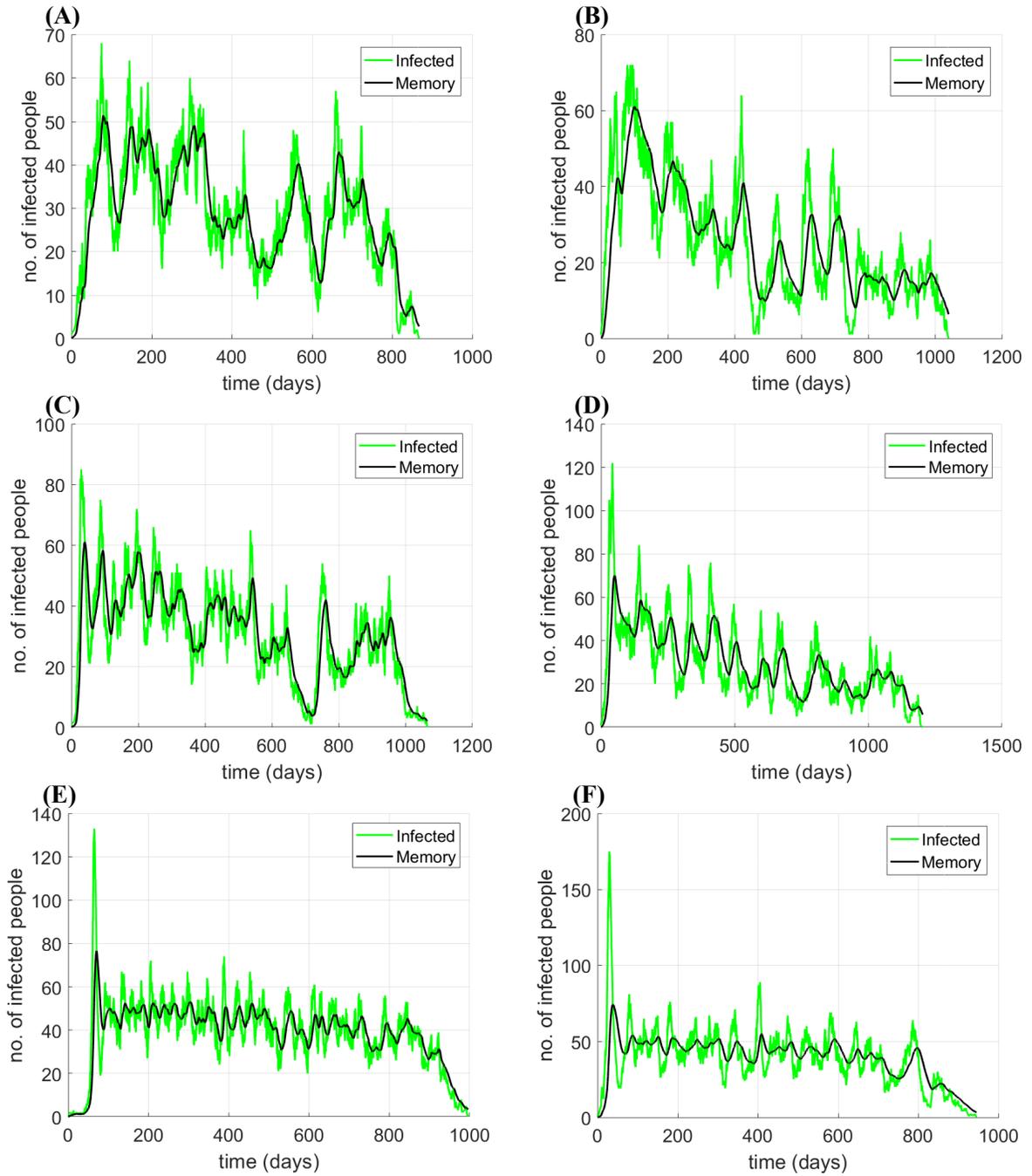


Figure 17: Exponentially Fading Kernel: effect of 'filtering' with  $BRN = 2$ . Plot of  $I(t)$  (green lines) and  $M(t)$  (black lines) for a single instance of the stochastic process. Left panels:  $\alpha = 0.1$ , Right panels:  $\alpha = 0.05$ . (A) and (B):  $p = 1$ ; (C) and (D):  $p = 2$ ; (E) and (F):  $p = 10$ .

Further improvements **might** include: (i) further epidemiological detail, e.g. the presence of a latency period; ii) the effects of exogenous seasonal fluctuations in transmission; (iii) the combination of individuals' behavior with governmental responses; (iv) the spatial dimension **modeled as continuous or by a discrete metapopulation approach** ([46]): **it is well-known that the metapopulation network topology can relevantly impact on disease spread** [7, 49] ; (v) the presence of multiple pathogen strains; (vi) **the possibility that, besides the transmission rate, the behavioral response affect also the recovery rate. For COVID-19 this surely occurred e.g., in the propensity to testing and self-isolating by pauci-symptomatic individuals** ; (vii) **the modeling of the propagation of information through its articulated interpersonal and virtual networks** [49–51].

## Declarations

**Acknowledgements** The authors warmly thank three anonymous referees of this Journal whose comments allowed to greatly improve the exposition of the manuscript. Usual disclaimers apply.

**Funding** : A d'O and PM did not receive funding for this work; Dr Ochab research stage at IPRI was supported by: i) a EU 'Erasmus+' grant; ii) following French Work Law on Research Stages, Dr Ochab received monthly a 'remuneration de stage' (research internship payment) by IPRI; KP's project was partially supported by Statutory Research of the Department of Systems Biology and Engineering, Silesian University of Technology, Poland. Grant No. 02/040/BK21/1010.

**Conflicts of interest or Competing interests** The authors declare that they have no conflicts of interest.

**Availability of data and material** Since part of this work has been done during the stay of Dr Ochab at the international Prevention Research Institute (IPRI), a private firm, the authorization to get the simulated time series obtained during her stay at IPRI must be asked to IPRI.

**Code availability** Since part of this work has been done during the stay of Dr Ochab at IPRI, a private firm, the authorization to get the codes prepared at IPRI must be asked to IPRI. However, this work contains a detailed description of the algorithms allowing all readers to replicate their implementation.

**Authors' contributions** Research Design: Ad'O; Model Design: Ad'O; Algorithm Design: Ad'O; Algorithm implementation: MO; Numerical Experiments design : Ad'O PM, KP; Numerical Experiments: MO; Results Analysis: Ad'O, PM, KP; Discussion: Ad'O, PM, MO, KP; Writing of the manuscript: Ad'O, PM, KP, MO.

## APPENDIX

### A Global Stability of the Endemic Equilibrium in the no-memory case

In this appendix we show the Global Stability of the Endemic Equilibrium in the no-memory case  $K(x) = \delta(x)$  without imposing technical conditions. In paper [36] , it was shown that the endemic equilibrium is globally stable provided that  $F(I) = I\beta(I)$  is either increasing or having at most an inflexion point. Here we remove this requirement.

Since model (1)–(2) has a stable manifold the set  $I = 0$  it is convenient to investigate (in our case in the set

$$\Gamma = \{(S, I) | S \geq 0, I > 0, S + I \leq 1\}$$

) an appropriately chosen topologically equivalent dynamical system (TEDS) obtained by multiplying the velocity field by a Dulac multiplier [52]. In our case, we employed the following multiplier:  $1/(I\beta(I))$ . The resulting TEDS reads as follows

$$\frac{d}{d\tau}S = \frac{A(I)}{I}\mu(1-S) - S \quad (36)$$

$$\frac{d}{d\tau}I = S - cA(I) \quad (37)$$

where we set for the sake of notation simplicity:  $N = 1$ ,  $A(I) = 1/\beta(I)$ , and  $c = \gamma + \mu$ . The rescaled time  $\tau$  is linked to the original time  $t$  by

$$\tau = \int_0^\tau I(t)\beta(I(t))dt.$$

Note that at the endemic equilibrium it is

$$\mu(1 - S_e) = cI_e,$$

which implies that at equilibrium the Jacobi matrix  $J$  is such that

$$J_{12} = \partial_I \left( \frac{d}{d\tau}S \right) = \mu(1 - S_e) \left( \frac{A'(I_e)}{I_e} - \frac{A(I_e)}{I_e^2} \right) = c \left( A'(I_e) - \frac{A(I_e)}{I_e} \right).$$

Thus the characteristic equation reads as follows

$$\lambda^2 + \left( cA'(I_e) + 1 + \mu \frac{A(I_e)}{I_e} \right) \lambda + c \frac{A(I_e)}{I_e} (1 + \mu A'(I_e)) = 0$$

which cannot have positive real part roots or zero roots. Thus, the endemic equilibrium is LAS. Moreover, applying the Dulac-Bendixon theorem yields

$$\text{div} \left( \frac{d}{d\tau}S, \frac{d}{d\tau}I \right) = -1 - \frac{A(I)}{I} - cA'(I) < 0$$

It easily follows that the Endemic Equilibrium of system of (36)–(37) is GAS in the set  $\Gamma$ . Due to the topological equivalence, even the endemic equilibrium of model (1)–(2) is GAS in  $\Gamma$ .

## B The deterministic Memoryless case

### B.1 Effects of $M_{50\%}$ in the deterministic case with $p = 1$

The output of the deterministic model for  $p = 1$  is reported in figure 18, where the effects of three different values of parameter  $M_{50\%}$  are shown.

### B.2 The effects of the shape of the response function in the deterministic case

For sake of comparison with our stochastic simulation, in figure 19 the behavior of the deterministic model in the memoryless case is reported for two values of  $p$ .

## C Exponentially fading kernels and multiple deterministic epidemic waves

Here we illustrate the effects of a delayed behavioral response in the deterministic case, for  $BRN = 15$  (figure 20) and  $BRN = 2$  (figure 21) and for distinct values of  $p$  ( $p = 1, 2, 10$ ).

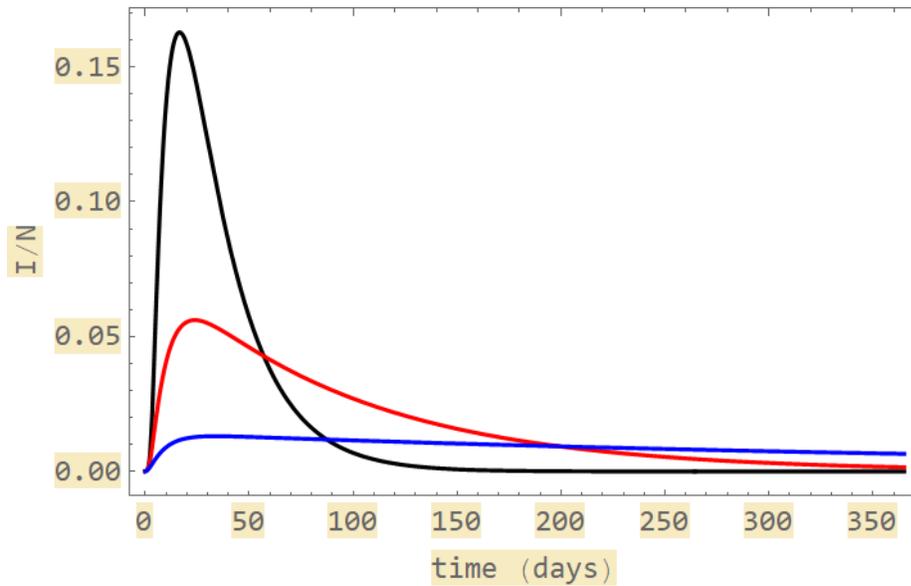


Figure 18: Pattern of infectious prevalence in the deterministic model for  $BRN = 15$ , and a memoryless behavioral response with  $p = 1$ . Black curve:  $M_{50\%} = 200$ , red curve:  $M_{50\%} = 50$ , blue curve  $M_{50\%} = 10$ .

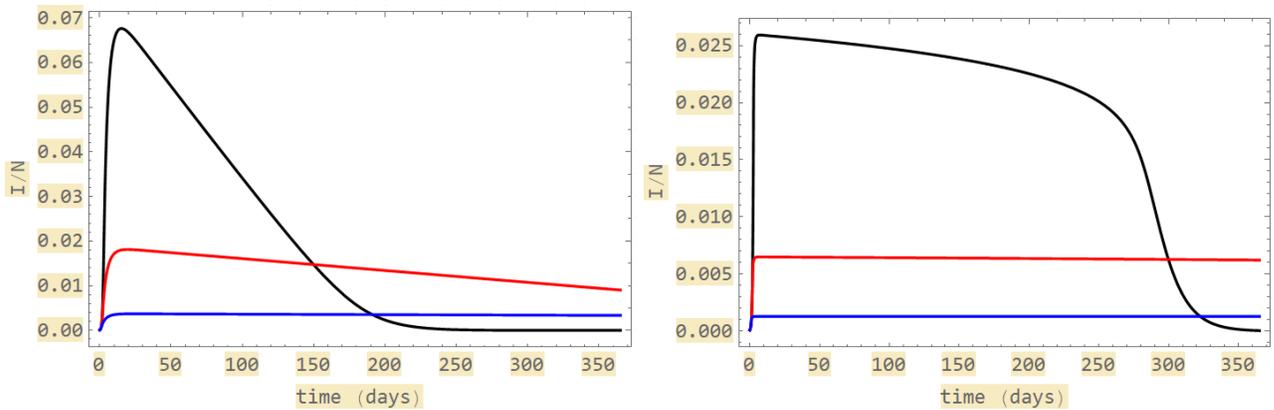


Figure 19: The deterministic model in the memoryless case. Effects of parameter  $p$  on the time-course of infectious prevalence for  $BRN = 15$ . Left panel:  $p = 2$ , right panel:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , red curves:  $M_{50\%} = 50$ , blue curves  $M_{50\%} = 10$ .

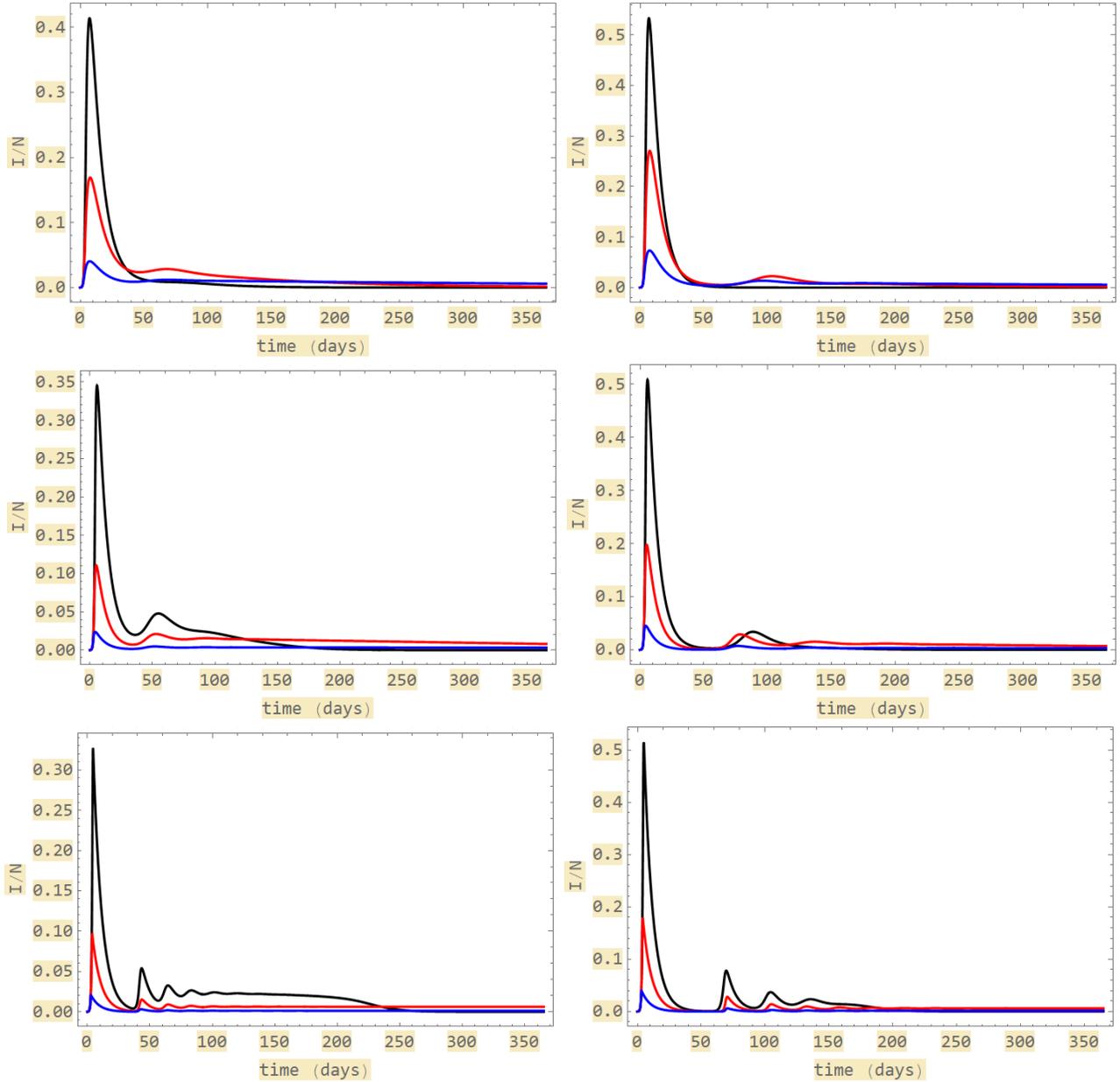


Figure 20: Deterministic multiple epidemic waves of infectious prevalence for an exponentially fading behavioral response, under  $BRN = 15$ , for different values of the memory rate  $a$ , and of parameters  $p$  and  $M_{50\%}$  tuning the behavioral response. Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

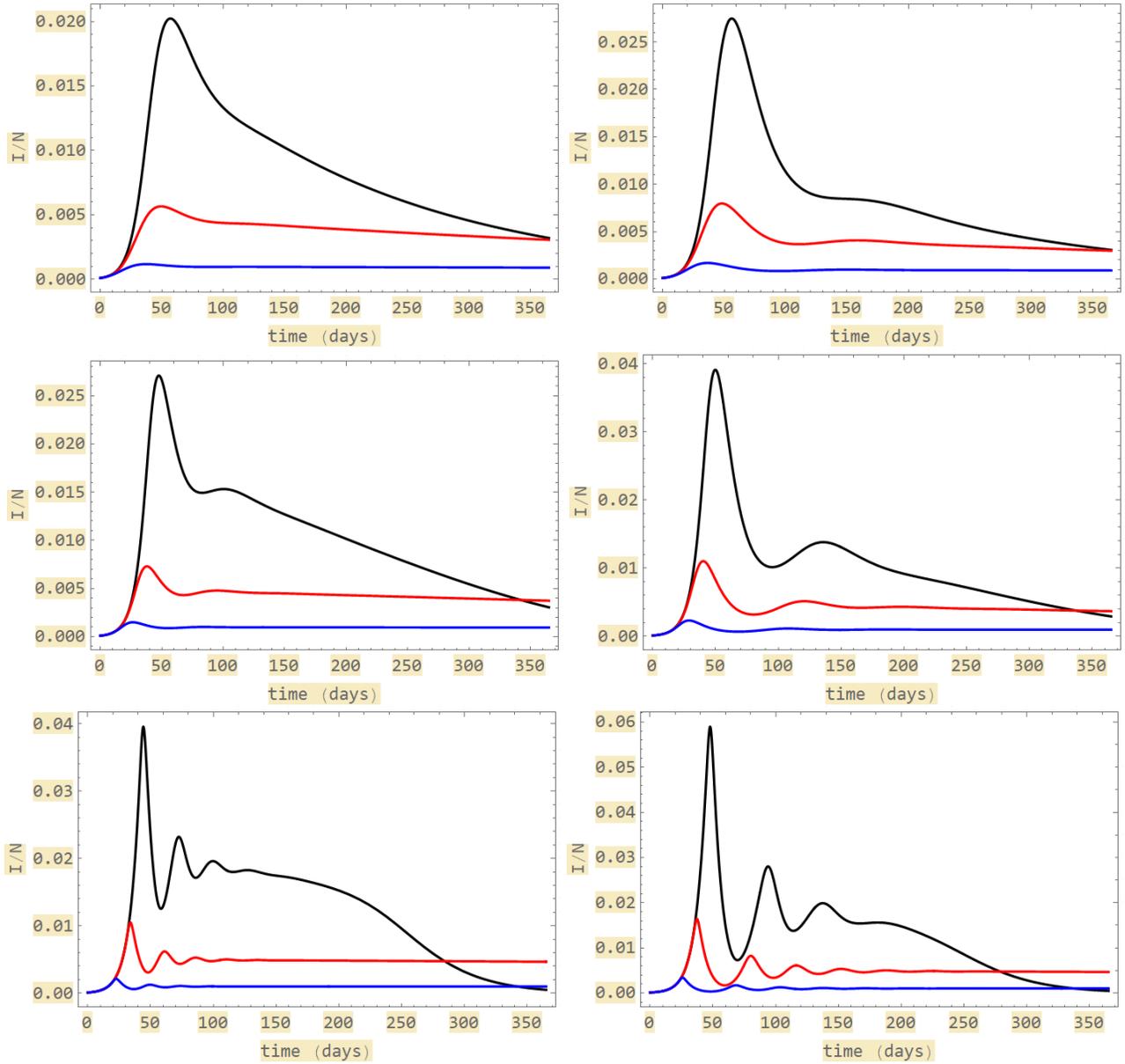


Figure 21: Deterministic multiple epidemic waves of infectious prevalence for an exponentially fading behavioral response, under  $BRN = 2$ , for different values of the memory rate  $a$ , and of parameters  $p$  and  $M_{50\%}$  tuning the behavioral response. Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

## D Exponentially fading Memory: the filtering effect for $BRN = 15$

Figure 22 illustrates the effect of 'filtering' with  $BRN = 15$  induced by an Exponentially Fading Kernel.

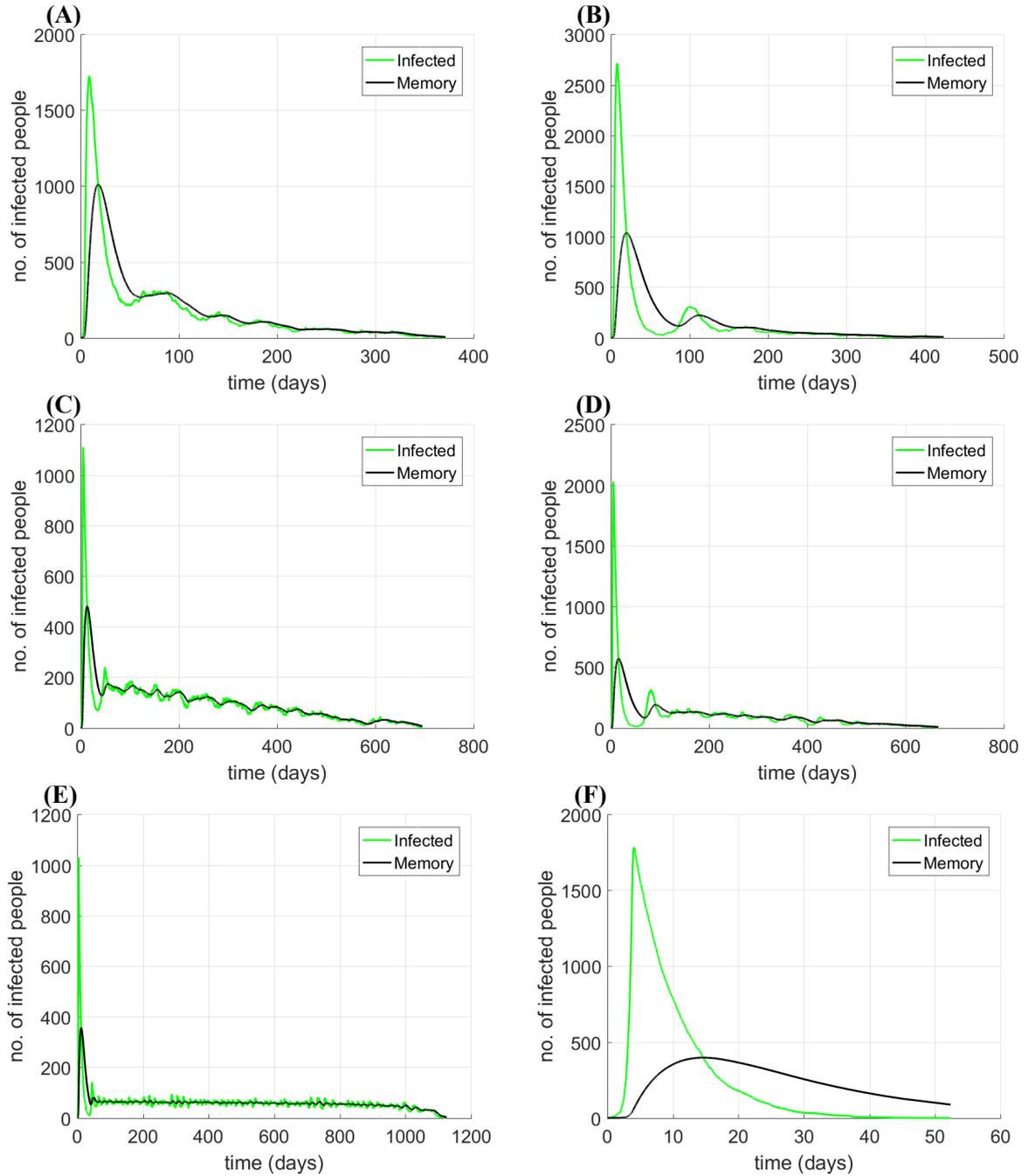


Figure 22: Exponentially Fading Kernel: effect of 'filtering' with  $BRN = 15$ . Plot of  $I(t)$  (green lines) and  $M(t)$  (black lines) for a single instance of the stochastic process. Left panels:  $a = 0.1$ , Right panels:  $a = 0.05$ . (A) and (B):  $p = 1$ ; (C) and (D):  $p = 2$ ; (E) and (F):  $p = 10$ .

## E Multiple Stochastic Epidemic Waves: the case of the Acquisition Fading Kernel

In this section, we will illustrate the results of the simulations in case of Acquisition Fading Kernel. Since the acquisition phase is short, we set  $a_1 = 1$ , whose characteristic time is  $T_1 = 1$ . On the contrary, the fading phase is much slower so we considered two values:  $a_2 = 0.1$ , i.e.  $T_2 = 10 \gg T_1$ , and  $a_2 = 0.05$ , i.e.  $T_2 = 20 \gg T_1$ .

### E.1 Deterministic and Stochastic Multiple Waves

Deterministic simulations for  $(a_1, a_2) = (1, 0.1)$  and  $(a_1, a_2) = (1, 0.05)$  are shown, respectively, in left and right panels of figure 23, where  $BRN = 15$ . The deterministic model suggests that multiple epidemic waves can occur for  $BRN = 15$  for large populations. Similarly, multiple waves are also observed for  $BRN = 2$  (not shown).

Patterns of multiple waves are also observed in the stochastic model, for both  $BRN = 15$  (figure 24) and  $BRN = 2$  (figure 25).

### E.2 Statistical Assessment

Figure (26) shows that the prevalence at the first peak. For  $BRN = 15$  it is interesting to note as the location of the PDFs depends non-monotonically on  $p$ . As far as the prevalence at the second peak, Figure (27) shows that for  $a_2 = 0.1$  the PDF for  $p = 1$  overlaps with the PDF for  $p = 2$ , which for  $a_2 = 0.05$  moves rightwards and overlaps partially with the PDF for  $p = 10$ . As far as the time of the first peak, Figure (28) shows that the PDFs seem mixtures between an unimodal PDF and an exponential PDF (faster decaying for  $BRN = 15$ , slower for  $BRN = 2$ ). As far as the time of at the second peak, Figure (29) shows that for  $BRN = 15$  the average decreases with  $p$  and also the variance. The distribution of the time at epidemic extinction time is of interest (see figure 30). Indeed for  $(BRN, a_2) = (15, 0.1)$  the location of the PDFs is increasing with  $p$  but the PDF for  $p = 10$  has also an appreciable earlier mode. On the contrary; for  $(BRN, a_2) = (15, 0.05)$  all the three PDF have an earlier peak; which in the case  $p = 10$  is the largest one. For  $BRN = 2$  the PDFs overlaps are mixtures between unimodal and fastly decaying exponentials. Finally, even the distribution of the final attack rates are of some interest (figure 31). Indeed, for  $(BRN, a_2) = (15, 0.1)$  the PDF are increasing with  $p$  but for  $p = 10$  the PDF has a smaller but appreciable earlier mode. For  $(BRN, a_2) = (15, 0.05)$ , all the three PDFs have an intermediate mode, and for  $p = 10$  this mode is the largest one. Finally, figure 32 shows as  $p$  is a key factor influencing the average and shape of the histogram of main peaks.

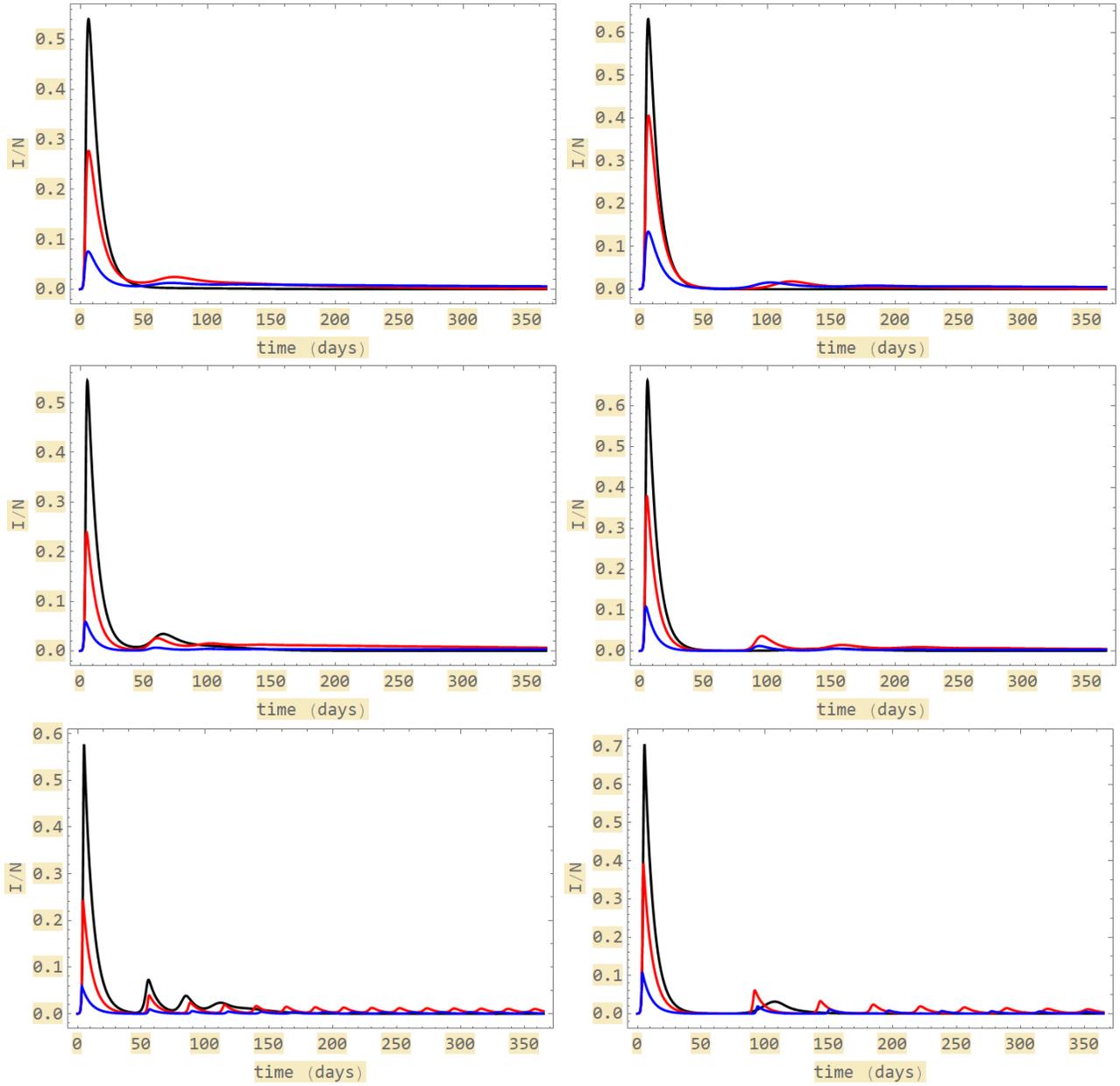


Figure 23: Acquisition Fading Kernel: Multiple Epidemic Waves predicted by the deterministic model. Left panels:  $a = 0.1$ ; right panels:  $a = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

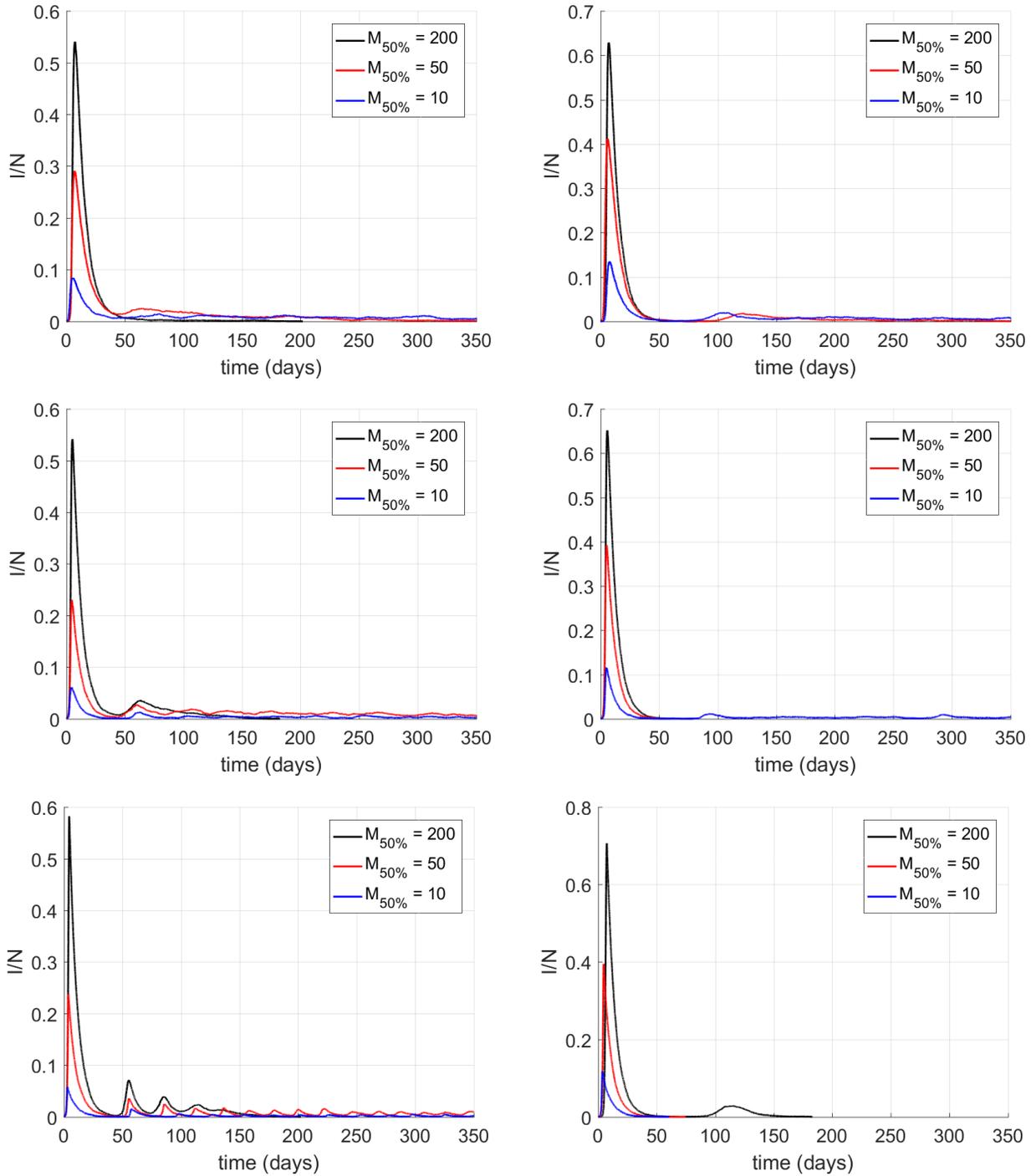


Figure 24: Acquisition Fading Kernel: Multiple Epidemic Waves predicted by the stochastic model for  $BRN = 15$ . Left panels:  $a_2 = 0.1$ ; right panels:  $a_2 = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

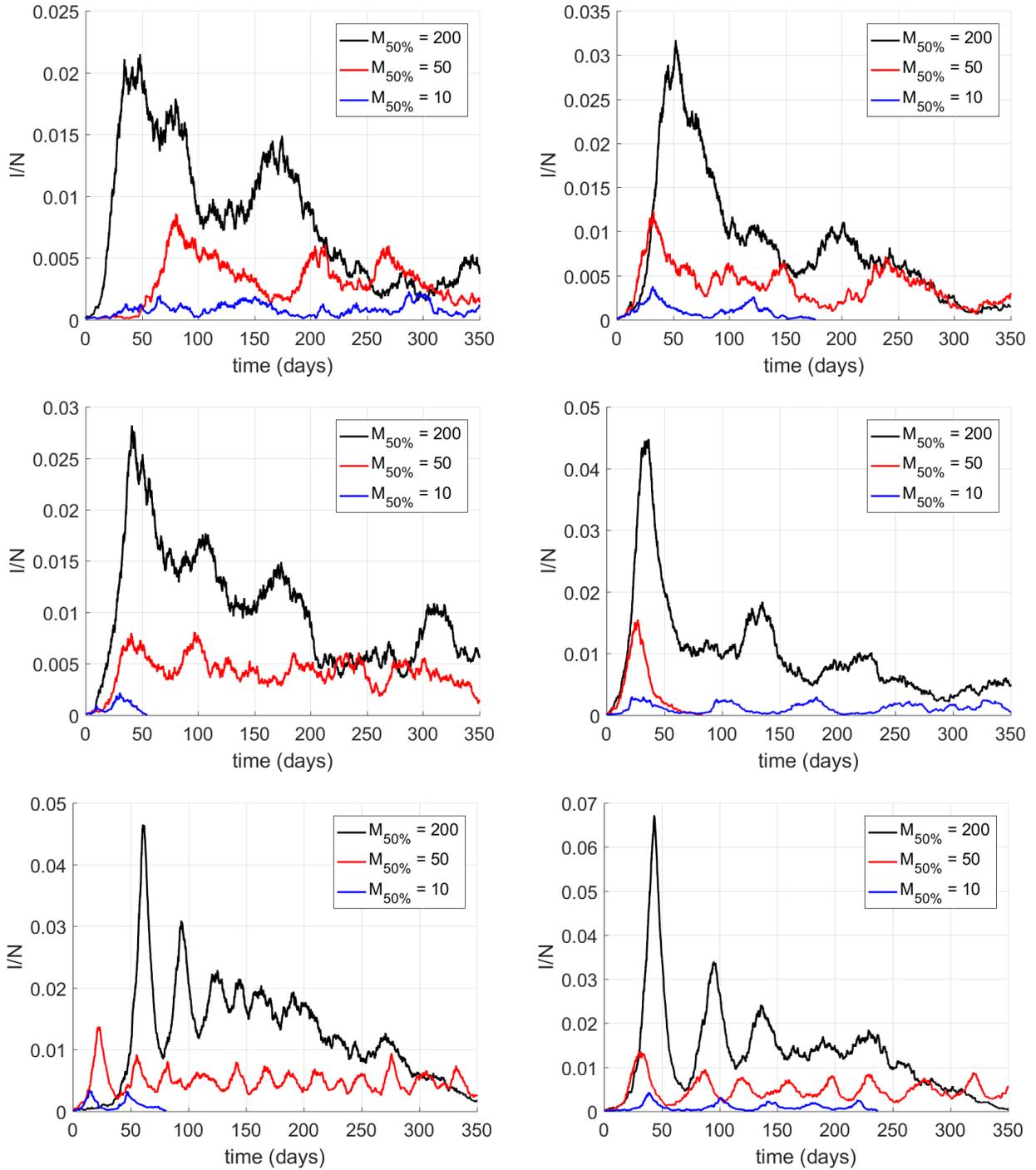


Figure 25: Acquisition Fading Kernel: Multiple Epidemic Waves predicted by the stochastic model for  $BRN = 2$ . Left panels:  $a_2 = 0.1$ ; right panels:  $a_2 = 0.05$ . Upper panels:  $p = 1$ ; central panels:  $p = 2$ ; lower panels:  $p = 10$ . Black curves:  $M_{50\%} = 200$ , Red curves:  $M_{50\%} = 50$ , Blue Curves:  $M_{50\%} = 10$ .

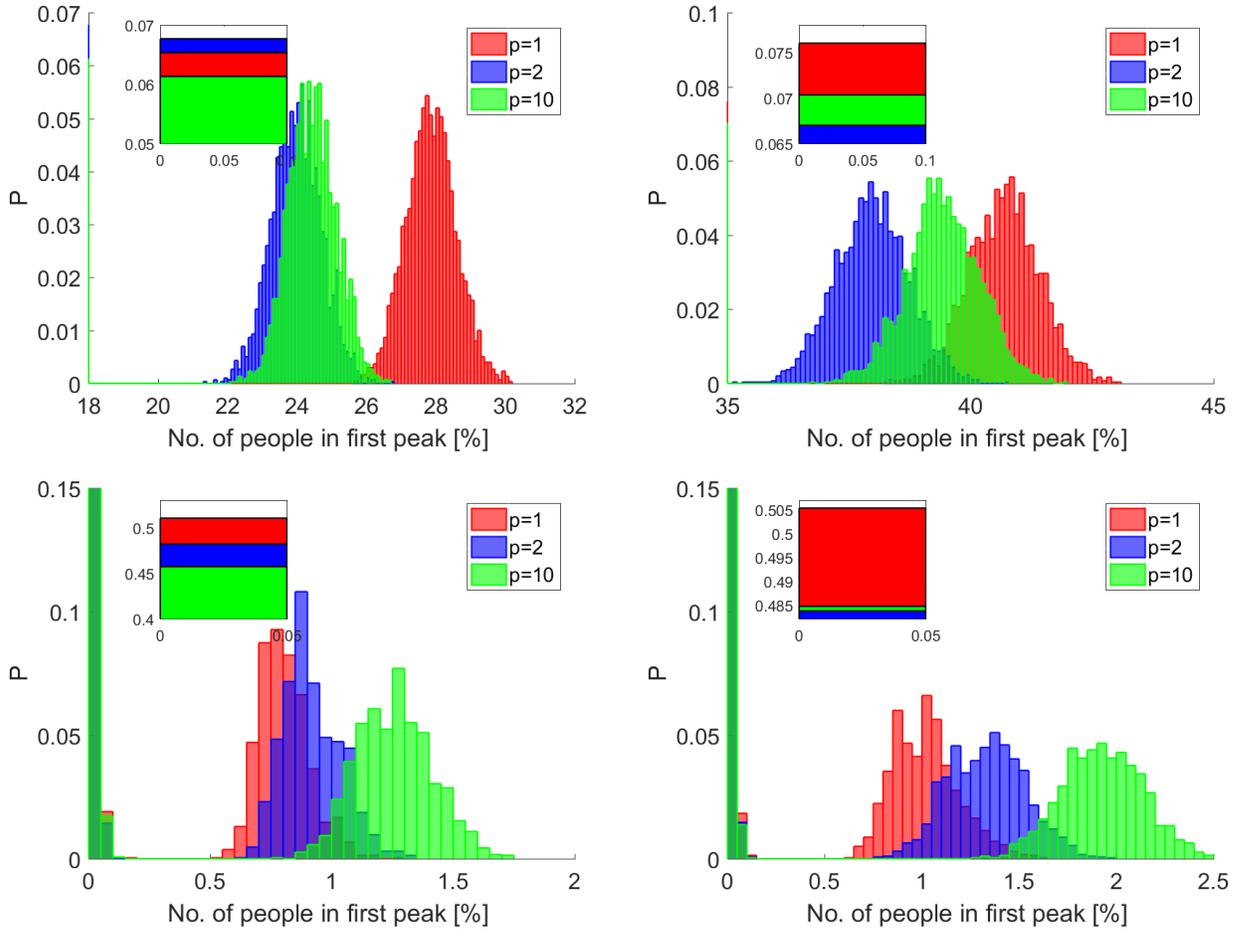


Figure 26: Acquisition Fading Kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the prevalence at the first epidemic peak. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$ ; (C)  $(BRN, a_2) = (2, 0.1)$ . (D)  $(BRN, a_2) = (2, 0.05)$

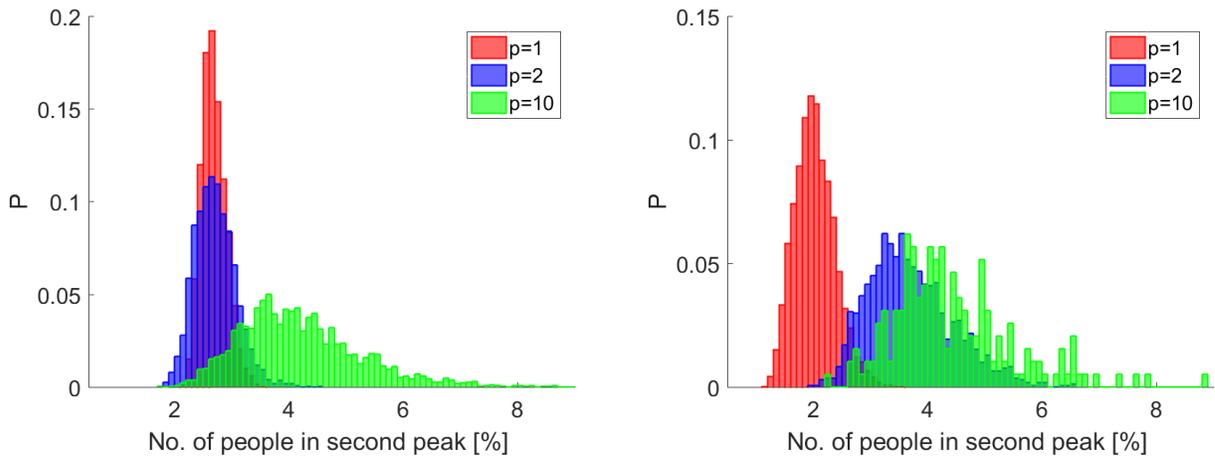


Figure 27: Acquisition Fading memory kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the prevalence at the **second** epidemic peak. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$

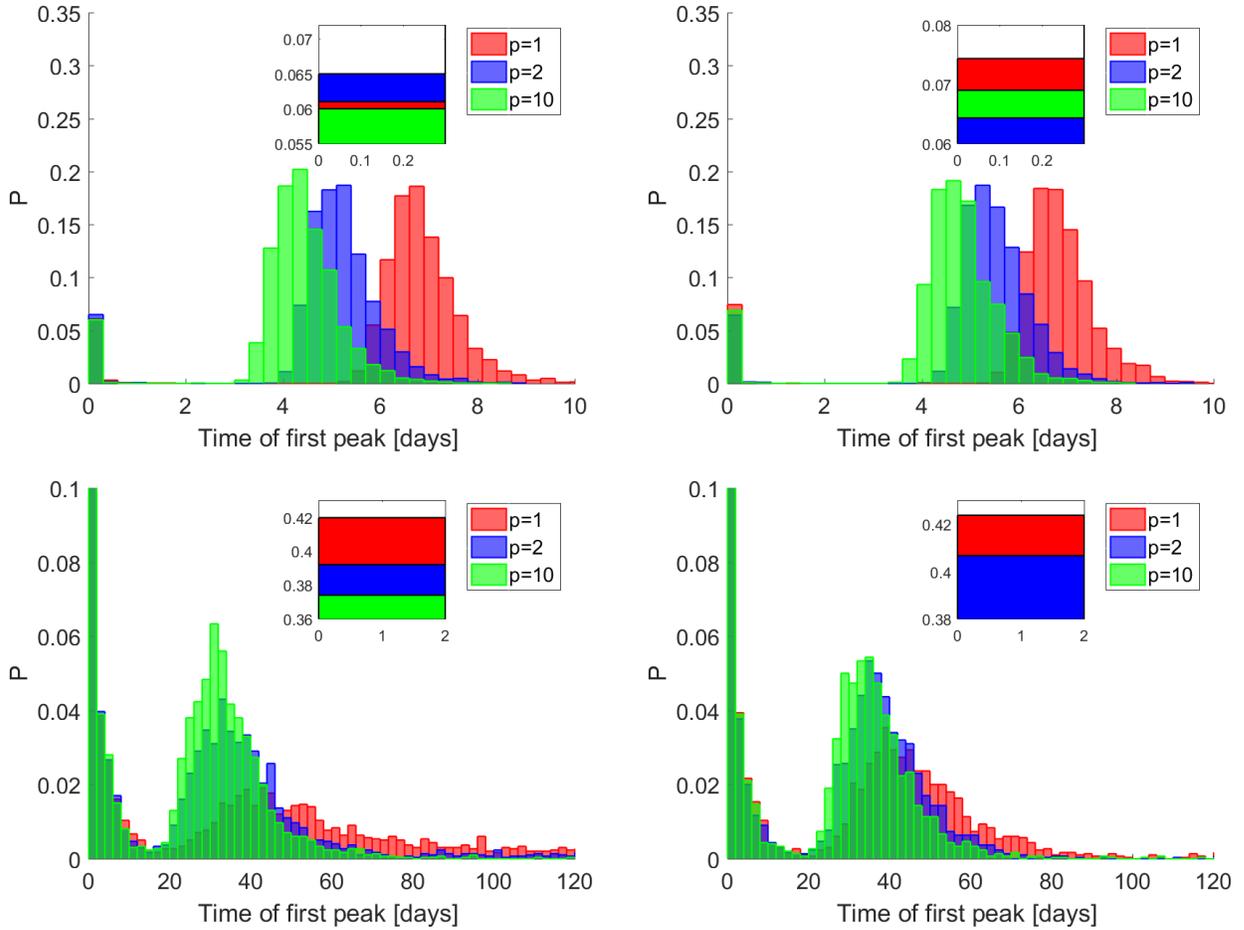


Figure 28: Acquisition Fading memory kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the time of the first epidemic peak. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$ ; (C)  $(BRN, a_2) = (2, 0.1)$ . (D)  $(BRN, a_2) = (2, 0.05)$

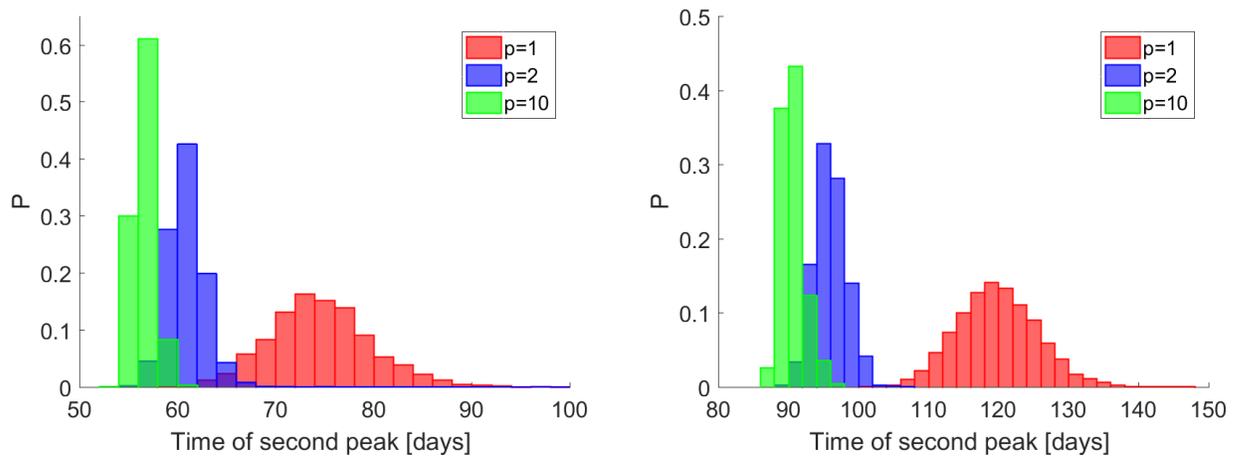


Figure 29: Acquisition Fading memory kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the time of the **second** epidemic peak. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$

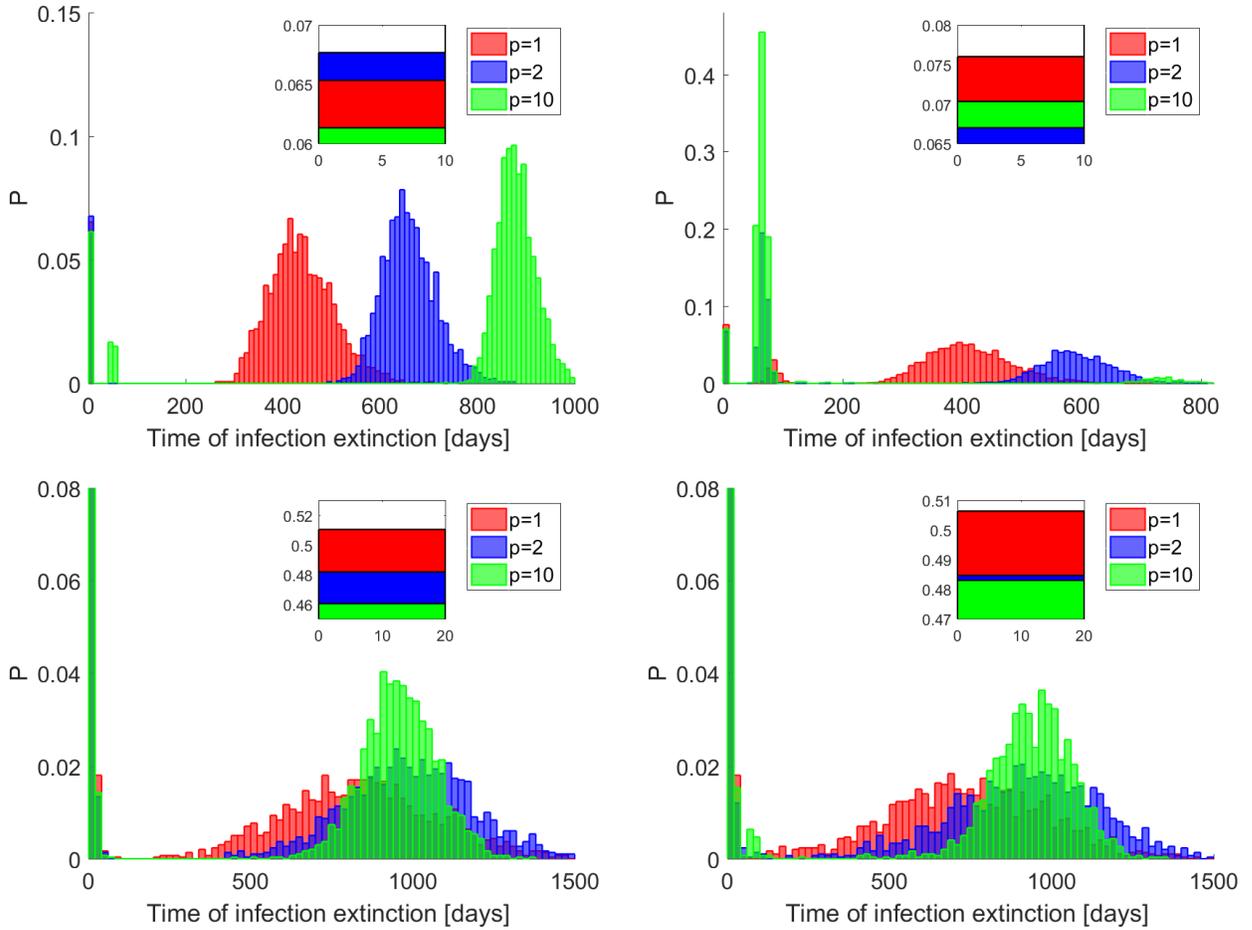


Figure 30: Acquisition Fading Kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of infection extinction times. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$ ; (C)  $(BRN, a_2) = (2, 0.1)$ . (D)  $(BRN, a_2) = (2, 0.05)$

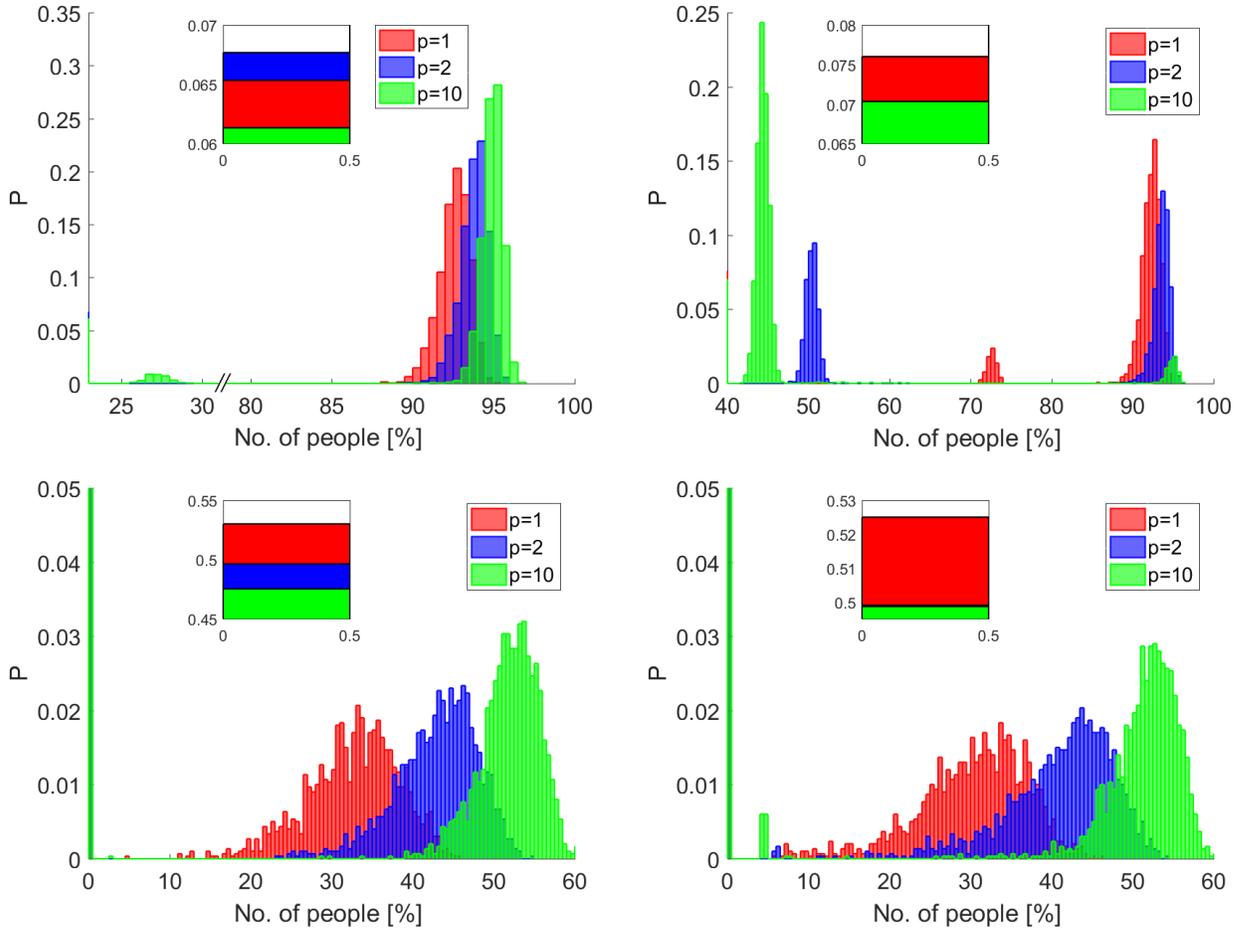


Figure 31: Acquisition Fading memory kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the final attack rate. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$ ; (C)  $(BRN, a_2) = (2, 0.1)$ . (D)  $(BRN, a_2) = (2, 0.05)$

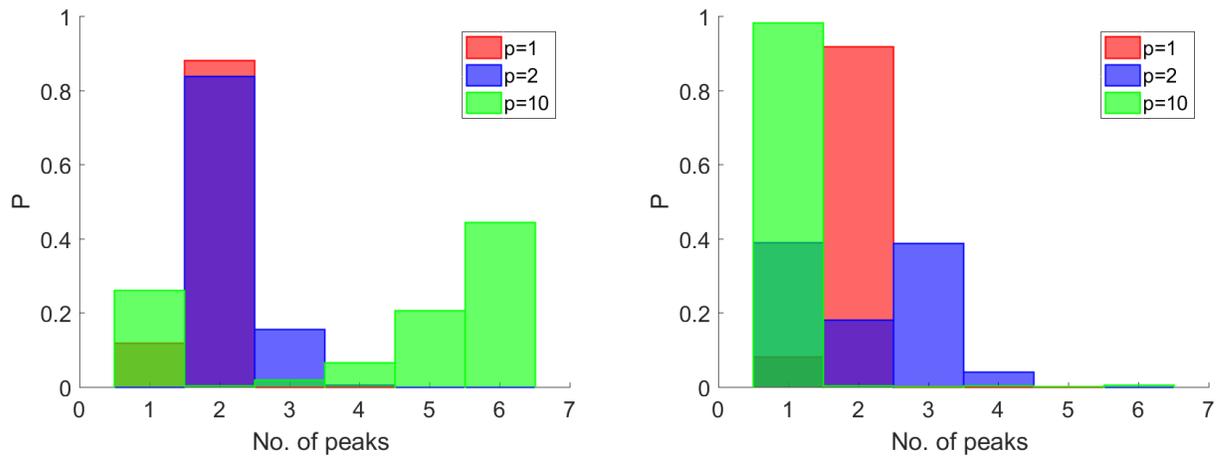


Figure 32: Acquisition Fading memory kernel. Histograms for  $p = 1, 2, 10$  and  $M_{50\%} = 50$  of the PDF of the number of peaks. (A)  $(BRN, a_2) = (15, 0.1)$ , (B)  $(BRN, a_2) = (15, 0.05)$ ;

## References

- [1] Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638):204–230, 1916.
- [2] William O 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.
- [3] Vincenzo Capasso. *Mathematical structures of epidemic systems*. Springer, 1993.
- [4] B Buonomo, N Chitnis, and A d’Onofrio. Seasonality in epidemic models: a literature review. *Ricerche Di Matematica*, 67(1):7–25, 2018.
- [5] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.
- [6] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.
- [7] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d’Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. *Physics Reports*, 664:1–113, 2016.
- [8] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, 10(1):190, 2010.
- [9] Nicola Perra. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports*, 2021.
- [10] Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.
- [11] Chris Bauch, Alberto d’Onofrio, and Piero Manfredi. Behavioral epidemiology of infectious diseases: an overview. *Modeling the interplay between human behavior and the spread of infectious diseases*, pages 1–19, 2013.
- [12] Piero Manfredi and Alberto D’Onofrio. *Modeling the interplay between human behavior and the spread of infectious diseases*. Springer, 2013.
- [13] Vincenzo Capasso and Gabriella Serio. A generalization of the kermack-mckendrick deterministic epidemic model. *Mathematical Biosciences*, 42(1-2):43–61, 1978.
- [14] Vincenzo Capasso. Global solution for a diffusive nonlinear deterministic epidemic model. *SIAM Journal on Applied Mathematics*, 35(2):274–284, 1978.
- [15] Alberto d’Onofrio, Piero Manfredi, and Ernesto Salinelli. Vaccinating behaviour, information, and the dynamics of sir vaccine preventable diseases. *Theoretical population biology*, 71(3):301–317, 2007.
- [16] Piero Manfredi and Alberto d’Onofrio. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. Springer, New York, 2013.

- [17] Alberto d’Onofrio and Piero Manfredi. Information-related changes in contact patterns may trigger oscillations in the endemic prevalence of infectious diseases. *Journal of Theoretical Biology*, 256(3):473–478, 2009.
- [18] Horacio J Marquez. *Nonlinear control systems: analysis and design*. Wiley-Interscience Hoboken, 2003.
- [19] John Bechhoefer. Feedback for physicists: A tutorial essay on control. *Reviews of Modern Physics*, 77(3):783, 2005.
- [20] Mathukumalli Vidyasagar. *Nonlinear systems analysis*, volume 42. Siam, 2002.
- [21] Daizhan Cheng, Xiaoming Hu, and Tielong Shen. *Analysis and design of nonlinear control systems*. Springer, 2010.
- [22] Carlo Cosentino and Declan Bates. *Feedback control in systems biology*. Crc Press, 2011.
- [23] James P Keener and James Sneyd. *Mathematical physiology*. Springer, 1998.
- [24] Norman TJ Bailey. *The elements of stochastic processes with applications to the natural sciences*. John Wiley & Sons, 1990.
- [25] Crispin Gardiner. *Stochastic Methods*. Springer, 2009.
- [26] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403 – 434, 1976.
- [27] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [28] Aurélien Alfonsi, Eric Cances, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In *ESAIM: proceedings*, volume 14, pages 1–13. EDP Sciences, 2005.
- [29] Giulio Caravagna, Alberto d’Onofrio, Paolo Milazzo, and Roberto Barbuti. Tumour suppression by immune system through stochastic oscillations. *Journal of theoretical biology*, 265(3):336–345, 2010.
- [30] Giulio Caravagna, Giancarlo Mauri, and Alberto d’Onofrio. The interplay of intrinsic and extrinsic bounded noises in biomolecular networks. *PLoS One*, 8(2):e51174, 2013.
- [31] Giulio Caravagna, Alex Graudenzi, and Alberto d’Onofrio. Distributed delays in a hybrid model of tumor-immune system interplay. *Mathematical Biosciences & Engineering*, 10(1):37, 2013.
- [32] Marino Gatto, Enrico Bertuzzo, Lorenzo Mari, Stefano Miccoli, Luca Carraro, Renato Casagrandi, and Andrea Rinaldo. Spread and dynamics of the covid-19 epidemic in italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117(19):10484–10491, 2020.
- [33] Bruno Buonomo and Rossella Della Marca. Effects of information–induced behavioural changes during the COVID–19 lockdowns: the case of Italy. *Royal Society Open Science*, 7(10):201635, 2020.
- [34] Enrico Bertuzzo, Lorenzo Mari, Damiano Pasetto, Stefano Miccoli, Renato Casagrandi, Marino Gatto, and Andrea Rinaldo. The geography of covid-19 spread in italy and implications for the relaxation of confinement measures. *Nature communications*, 11(1):1–11, 2020.

- [35] Piero Poletti, Marco Ajelli, and Stefano Merler. The effect of risk perception on the 2009 h1n1 pandemic influenza dynamics. *PloS one*, 6(2):e16460, 2011.
- [36] Wendi Wang. Epidemic models with nonlinear infection forces. *Mathematical Biosciences & Engineering*, 3(1):267, 2006.
- [37] Norman MacDonald. *Time lags in biological models*. Springer, 1978.
- [38] Cruz Vargas-De-León and Alberto d’Onofrio. Global stability of infectious disease models with contact rate as a function of prevalence index. *Mathematical biosciences and engineering*, 14(4):1019–1033, 2017.
- [39] Denis V Efimov and Alexander L Fradkov. Oscillatory of nonlinear systems with static feedback. *SIAM Journal on Control and Optimization*, 48(2):618–640, 2009.
- [40] Denis Efimov, Wilfrid Perruquetti, and Anton Shiriaev. Conditions of existence of oscillations for hybrid systems. *IFAC Proceedings Volumes*, 46(23):223 – 228, 2013.
- [41] Daniel T Gillespie. *Markov processes: an introduction for physical scientists*. Elsevier, 1991.
- [42] Stefano Merler, Piero Poletti, Marco Ajelli, Bruno Caprile, and Piero Manfredi. Coinfection can trigger multiple pandemic waves. *Journal of theoretical biology*, 254(2):499–507, 2008.
- [43] D Rios-Doria and G Chowell. Qualitative analysis of the level of cross-protection between epidemic waves of the 1918–1919 influenza pandemic. *Journal of Theoretical Biology*, 261(4):584–592, 2009.
- [44] Ilaria Dorigatti, Simon Cauchemez, and Neil M Ferguson. Increased transmissibility explains the third wave of infection by the 2009 h1n1 pandemic virus in england. *Proceedings of the National Academy of Sciences*, 110(33):13422–13427, 2013.
- [45] Martin CJ Bootsma and Neil M Ferguson. The effect of public health measures on the 1918 influenza pandemic in us cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593, 2007.
- [46] E Sadurní and G Luna-Acosta. Exactly solvable sir models, their extensions and their application to sensitive pandemic forecasting. *Nonlinear dynamics*, 103(3):2955–2971, 2021.
- [47] WHO official page on coronavirus disease (covid-19) pandemic. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed: 2022–FEB–05.
- [48] Joshua S Weitz, Sang Woo Park, Ceyhun Eksin, and Jonathan Dushoff. Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proceedings of the National Academy of Sciences*, 117(51):32764–32771, 2020.
- [49] Chengyi Xia, Li Wang, Shiwen Sun, and Juan Wang. An sir model with infection delay and propagation vector in complex networks. *Nonlinear Dynamics*, 69(3):927–934, 2012.
- [50] Qian Shao, Chengyi Xia, Lin Wang, and Huijia Li. A new propagation model coupling the offline and online social networks. *Nonlinear Dynamics*, 98(3):2171–2183, 2019.
- [51] Zhishuang Wang and Chengyi Xia. Co-evolution spreading of multiple information and epidemics on two-layered networks under the influence of mass media. *Nonlinear Dynamics*, 102(4):3039–3052, 2020.
- [52] Stephen Wiggins, Stephen Wiggins, and Martin Golubitsky. *Introduction to applied nonlinear dynamical systems and chaos*. Springer, 1990.