Reduced modelling and optimal control of epidemiological individual-based models with contact heterogeneity

Clémentine Courtès, Emmanuel Franck, Killian Lutz, Laurent Navoret, Yannick Privat

Reduced modelling and optimal control of epidemiological individual-based models with contact heterogeneity

C. Courtès[*] E. Franck[†] K. Lutz[‡] L. Navoret[§] Y. Privat[¶ ‖]

December 22, 2022

Abstract

Modelling epidemics using classical population-based models suffers from shortcomings that so-called individual-based models are able to overcome, as they are able to take into account heterogeneity features, such as super-spreaders, and describe the dynamics involved in small clusters. In return, such models often involve large graphs which are expensive to simulate and difficult to optimize, both in theory and in practice.

By combining the reinforcement learning philosophy with reduced models, we propose a numerical approach to determine optimal health policies for a stochastic individual-based model taking into account heterogeneity in the population. More precisely, we introduce a deterministic reduced population-based model involving a neural network, designed to faithfully mimic the local dynamics of the more complex individual-based model. Then the optimal control is determined by sequentially training the network until an optimal strategy for the population-based model succeeds in also containing the epidemic when simulated on the individual-based model.

After describing the practical implementation of the method, several numerical tests are proposed to demonstrate its ability to determine controls for models with contact heterogeneity.

Contents

[*]IRMA, Université de Strasbourg, CNRS UMR 7501, Inria, 7 rue René Descartes, 67084 Strasbourg, France (clementine.courtes@unistra.fr).

[†]Inria, IRMA, Université de Strasbourg, CNRS UMR 7501, 7 rue René Descartes, 67084 Strasbourg, France (emmanuel.franck@unistra.fr).

[‡]Univ Lyon, Ecole Centrale de Lyon, CNRS UMR 5208, Institut Camille Jordan, F-69134 Ecully, France, (killian.lutz@ecl19.ec-lyon.fr).

[§]IRMA, Université de Strasbourg, CNRS UMR 7501, Inria, 7 rue René Descartes, 67084 Strasbourg, France (laurent.navoret@unistra.fr).

[¶]IRMA, Université de Strasbourg, CNRS UMR 7501, Inria, 7 rue René Descartes, 67084 Strasbourg, France (yannick.privat@unistra.fr).

[‖]Institut Universitaire de France (IUF).

Keywords: Individual-based models, Super-spreaders, Reduced models, Optimal control, Neural network

2020 AMS subject classifications: 92B05, 49M05, 68T07

1 Introduction

1.1 Population versus Individual-based models

Correclty modelling the spread of epidemics is of paramount importance in defining health policies to control their development. Models make it possible to estimate the so-called basic reproduction ratio $\mathcal{R}_0$, which indicates whether the epidemic is developing or not, and to propose optimal policies to limit the saturation of hospital services, for instance. In such a program, one difficulty is to take into account the presence of so-called *super-spreaders*. An individual is said to be a super-spreader if he/she is likely to infect many more people than a generic person would: indeed, the distribution of the number of contacts in the population is very heterogeneous. In the seminal work [28], the authors show that epidemics tend to be *rarer but more explosive* with super-spreaders. Indeed, super-spreaders tend to be infected in the early stages of an epidemic.

To tackle this problem, several levels of descriptions can be considered [23]. The two extreme types of models are population-based and individual-based models. *Population-based or mean-field models*, like the deterministic SIR model, are the simplest descriptions of epidemics: they describe the time evolution of the total number of susceptible (S), infected (I), retired (R) people or of other categories. At the opposite level of description, *individual-based models* are the most accurate. They describes the stochastic time evolution of the states (susceptible, infected, retired or other) of each

individual by taking into account the contact graph between individuals [23]. Similar deterministic individual-based versions could be used.

Individual-based models are much better suited to describe the effect of super-spreaders. Indeed, they rely on an precise description of the contact graph between individuals: each node corresponds to an individual and each edge corresponds to the contact between two individuals [37, 36, 11]. It is therefore easy to include contact heterogeneities by considering contact graphs with prescribed distributions of node degrees [40, 21]. The individual-based models are then constructed as a continuous-time Markov process, in which each individual can evolve between the different states (susceptible, infected, retired) at random times. Thus the infection of one susceptible individual depends on the number of connected infected neighbours as well as the individual transmission rate β_{ind}, while the transition to the retired state depends only on the recovery rate γ.

Individual-based models allow for more accurate modelling, but require more computing resources to simulate large or complex contact graphs [2]. It should be noted that models such as percolation processes have been proposed. The level of description is then intermediate and the state process is simplified. For instance, in [28, 14], the authors used this kind of models to describe the number of secondary infections. In some cases, population-based models can also be derived analytically from individual-based models, which is interesting for simulations [44, 45]. First, assuming some independence between the states of each individual, the individual-based stochastic models can be approximated by a deterministic model, where the state of each individual follows SIR differential systems coupled to those of other individuals [23]. Then, by assuming that the degree distribution has a small variance, this model can be further simplified and we recover the classical population based SIR models. Without this last assumption, SIR models structured by contact numbers can be derived but result in larger models.

Thus, to incorporate the effect of super-spreaders in population-based models, one possible approach is to expand the number of categories: for each health state, we can consider several compartments associated with sub-populations with more or less contacts [22]. Another strategy is to consider a single additional compartment with specific epidemiological properties at the population level (e.g. recovery rates, transmission rates, etc.) [33]. Other studies proposed to take into account super-spreaders by considering spatial inhomogeneities, which leads to an unequal distribution of the epidemic in the country [13]. The infection probability then depends on the distance between two individuals, and the difference between normal individuals and super-spreaders is taken into account via this dependency.

In another direction, recurrent neural networks have been proposed to improve population-based models [3]. Their long short-term memory is used to identify the transmission rate as a function of the observed mobility and social behaviour. Based on external data, the paper constructs a time dependent transmission rate modelling the effects of some social interactions which could include super-spreading events. Recently neural networks have been used to increase the accuracy of the compartmental models (e.g. additional compartment dynamics [47]) or to estimate the parameters of a SIR-like model [18, 29].

Despite the advantages of population-based models in terms of simplicity, only individual-based models are really capable of handling heterogeneous contact distributions and describing epidemics with stochastic effects and few individuals.

1.2 Optimal control issues

In this work, we propose a strategy to define an optimal control for an individual-based model. The optimal control problem we consider is to keep the number of infected individuals below a certain threshold by adjusting the average transmission rate and the parameters of the contact distribution

over time.

More specifically, we consider an individual-based model where contacts follow a negative binomial distribution. Such a distribution is used to model populations with super-spreaders [28] and is parametrized by its mean value $\alpha > 0$ and the so-called *dispersion coefficient* $\kappa > 0$. Misleadingly, since the variance in the contacts distribution is $\alpha + \alpha^2/\kappa$, a low value of the dispersion coefficient κ is associated with a high variance of the contact distribution and thus with the presence of super-spreaders. The dispersion control acts mainly on the super-spreaders while the control of the average transmission rate $\beta = \alpha\beta_{\text{ind}}$ is a uniform control on the population. We are therefore interested in defining an optimal pair $(\beta(\cdot), \kappa(\cdot)) = (b(\cdot)\beta_0, k(\cdot)\kappa_0)$, with the smallest deviation from the initial values (β_0, κ_0), in order to maintain the number of infected people below a certain threshold.

For small graphs, a possible approach is to use convex optimisation techniques such as geometric programming [35]. However this type of methods does not scale well to larger graphs. On the other hand, when the graph represents the links between certain cities, analytical approaches are also used to solve the control problem [38]. To define a control of the Markovian stochastic individual-based model, a classical approach consists in using dynamic programming algorithms. Indeed, the problem can be formulated as a Markov Decision Process (MDP) [42], with given probability transitions between the 3^N states of the SIR model, where N denotes the number of individuals. For large N, it is no longer possible to easily deal with such a model completely and a reinforcement learning approach should be used. In [6], a Deep-Q reinforcement learning algorithm is applied to large graphs using a global control on agents (in the case of partially observable MDP). For such global control problems, cooperative multi-agent approaches can also be used [27]. Another approach is to consider reinforcement learning based on a reduced model. As proposed in [2], it may be interesting to rely on simpler models, like population-based ones, for which the standard control theory framework applies. The optimal control strategy for the reduced model is then used as the starting point for designing controls for the individual-based model.

Regarding the control of population-based models, references are numerous: they are generally based on the use of optimality conditions such as the Pontryagin Maximum Principle (PMP). If few of them propose an analytical design of the controls (see for instance [4, 5]), many introduce adapted optimization algorithms based either on a discretization of the complete problem (discretize then optimize, see e.g. [31]) or an algorithm on the continuous problem applied on a discretized version of the model (optimize then discretize, see e.g. [4, 5]).

1.3 Initial objective, added value of the approach and organization of the article

The initial objective of this paper is to propose a local control approach based on a reduced model of the individual-based model. The focus is on the methodology and strategy of the approach.

This optimal control strategy on a reduced model could be developed and optimised thanks to the estimation of quantities of interest, such as the reproduction ratio $\mathcal{R}_0$, at first sight secondary, but in fact crucial to justify the choices made in the elaboration of the strategy. This is why we have clarified them in the appendix in order to motivate the choices made thereafter. Modelling (e.g. taking into account super-spreaders) and comparison with the reality of health policies (e.g. interpretation of the final controls obtained) are also secondary in our approach and, as such, will not be addressed in the numerical results.

In this work, the proposed strategy for designing a control of the individual-based model is decomposed into the three following steps:

(i) First, using neural networks, learn a reduced population-based SIR model via data coming from numerical simulations of the individual-based model.

(ii) Then, define a control of the parameters of the population-based models.

(iii) Finally, use a model-based reinforcement algorithm to improve the population-based model around the controlled solution and thus the control itself.

The data-driven population-based SIR model is intended to capture the effect of heterogeneity in the contact distribution and the stochastic effects due to relatively small size of the population. The model therefore depends on the dispersion parameter κ, the transmission rate β and the relative size of the population n, a coefficient depending on N describing the closeness to the large population regime.

The neural network is trained to compute the time variation of the number of susceptible people following an ordinary differential equation of the form $S' = -F_\theta(S, I; n, \beta, \kappa)$, where θ denotes a parameters vector. Note that throughout this study, the recovery rate γ is fixed, equal to $1/6$ day^{-1}. Then, the optimal control of the data-driven population-based SIR model is defined by means of an optimal control algorithm. This will define time-varying parameters $(\beta(t), \kappa(t))$. However, since the learned population-based model is not *a priori* trained with such time varying parameters, the latter control is not well adapted to the underlying individual-based model. This is why the reinforcement strategy is essential to obtain a meaningful control with respect to the individual-based model.

The outline of the article is as follows. In the second section, we discuss the principles underlying the construction of the individual-based SIR model as well as the methodology used to average time-series of this stochastic system. We then introduce the main problem we aim to solve in this paper: an optimal control problem of the individual-based model. In the third section, we present the proposed method to solve the optimal control problem. We first describe how the data-driven reduced SIR model is constructed from the data using a classical multi-perceptron neural network. A theoretical analysis is performed to show that the control of the data-driven SIR reduced model is well defined and then the reinforcement strategy is detailed. Finally, in the fourth section, several numerical tests are performed to assess the validity of the whole methodology. A few appendices conclude this paper by detailing technical points on the individual-based model (Appendix A), the reduced data-driven population-based model and its proper utility to evaluate epidemiological quantities, especially in parameter regimes not covered by the classical SIR model (Appendix B), some proofs of the control section (Appendix C) and numerical algorithms (Appendix D).

2 Optimal control problem of an individual-based SIR model

In this section, we first introduce the individual-based SIR model, hereafter denoted (IBM), which is able to take into account numerous aspects and most notably the epidemic dynamics involving contact heterogeneity among the population. We then explain the challenges and method used to approximate the averaged dynamics of the stochastic IBM. After that, we introduce an optimal control problem for the IBM in which we seek to curb the spread of the epidemic.

In Table 1, we gather all the parameters introduced in the models presented in this section, in order to facilitate its reading. The reader wishing to find the definition of the parameter as well as the place where it is introduced in the article can thus refer to the table below.

2.1 Individual based SIR model: graph and continuous-time Markov process

The individual-based SIR model consists of a graph with N vertices, together with the specification of the SIR dynamics. Each vertex represents one individual, whose epidemic state over time is denoted $X_j(t) \in \{s, i, r\}$ for $t \geqslant 0$, for susceptible (s), infected (i) and retired (r). The edges of the graph represent the contacts between individuals: the number of contacts of the j-th individual

N	total number of individuals	Section 1.2
n	population size ratio	Section 3.1
β_{ind}	individual transmission rate	Section 2.1
γ	recovery rate	Section 2.1
α	average number of contacts	Section 2.1
$\beta = \alpha\beta_{\text{ind}}$	mean transmission rate	Section 2.1
κ	dispersion coefficient	Section 2.1
β_0	initial transmission rate	Section 2.3
κ_0	initial dispersion coefficient	Section 2.3
$b(t) = \beta(t)/\beta_0$	control functions	Section 2.3
$k(t) = \kappa(t)/\kappa_0$		
$b_{\min} \in (0, 1]$	minimal pointwise value of the control variable b	Section 2.3
$k_{\max} > 1$	maximal pointwise value of the control variable k	Section 2.3
$I_{\text{hosp}} < 1$	threshold used in the cost function to penalize the number of infected individuals in hospitals	
$I_{\max} \in (I_{\text{hosp}}, 1]$	threshold used in the cost function to penalize the maximum number of infected individuals	

Table 1: Parameters of the model and the optimal control problem

is denoted $\nu_j \in \mathbb{N}$. Then the dynamics is described by a continuous-time Markov process, whose two main parameters (common to all nodes) are the individual transmission rate $\beta_{\text{ind}} > 0$ and the recovery rate $\gamma > 0$.

The state space of the Markov process consists of the 3^N possible configurations of the individuals of the graph and the continuous-time Markov process can be defined by the transition rate from one configuration to another as follows. If two configurations differ for the j-th individual only, which goes from susceptible to infected ($s \to i$), then the transition rate equals βd_j where d_j is the number of its infectious contacts and $\beta = \alpha\beta_{\text{ind}}$ is the mean transmission rate, with $\alpha > 0$ denoting the average number of contacts in the graph. If two configurations differ for the j-th individual only, which goes from infected to recovered ($i \to r$), then the transition rate equals γ. We refer to [23] for a detailed description of this model.

More precisely, the changes of state of the individuals in the graph occur one by one at random times $(T^m)_{m\in\mathbb{N}}$. Given the graph state at time T^m, the next time T^{m+1} and the associated transition are defined as follows. We assign random clocks C_j to any states and these clocks follow exponential distributions. If the j-th individual is infected then the exponential distribution has a rate γ. If the j-th individual is susceptible, then the exponential distribution has a rate βd_j. Then the next transition occurs for the j^*-th individual at time $T^{m+1} = T^m + C_{j^*}$, where j^* corresponds to the smallest clock time among all individuals: $C_{j^*} = \min_j C_j$. With such dynamics, the more contacts an individual has with infected neighbors, the more likely he is to be infected in turn.

To investigate the role of super-spreaders in the dynamics, we consider a heterogeneous distribution of contacts. Accordingly, the edges of the graph are distributed such that the number of contacts ν_j of the j-th individual follows a (generalized) negative binomial, also called Pólya, distribution[1]:

$$\nu_j \sim \mathcal{BN}\left(\kappa, \frac{\kappa}{\alpha + \kappa}\right),$$

[1]The probability distribution of the Pólya distribution writes: $P(\nu_j = k) = \frac{\Gamma(\alpha+k)}{k!\Gamma(\alpha)}(1-p)^k p^\alpha$ for all $k \in \mathbb{N}$, with $p = \frac{\alpha}{\alpha+\kappa}$.

where $\kappa > 0$ is the dispersion parameter. The mean of the distribution is α and the variance equals $\alpha+\alpha^2/\kappa$. Thus, a small dispersion coefficient κ corresponds to a large variance and so to the existence of super-spreaders (see Figure 1). For large κ, the Pólya distribution converges (in law) to the Poisson distribution used to model a homogeneous population (see Figure 1). With this distribution, the heterogeneity of contacts relatively to the average number of contacts is parametrized by α/κ. In practice, in order to fit this distribution, the edges are constructed using the Molloy-Reed algorithm [34].

The possible controls of this model can either act on the individual transmission rate β_{ind} (by imposing masks, for instance) or on the contact distribution parameters (α, κ) (one can think for example of the lockdown or the closing of the restaurants). However, the control of β_{ind} and α are quite similar as they both modify the mean transmission rate $\beta = \alpha\beta_{\text{ind}}$, which is the key parameter for epidemics developments. Thus the two main parameters that we are aiming to control are:

(i) the mean transmission rate $\beta = \alpha\beta_{\text{ind}}$,

(ii) the dispersion coefficient κ,

while the recovery rate γ and the population size N are two given quantities.

The model is one of the simplest model for the dynamics of epidemics on graphs. It is simulated by using a Gillepsie algorithm [15]. We refer to [23] for more details. We performed numerical simulations for this model using the Python packages EpidemicsOnNetworks [32] and NetworkX [19]. At time $t = 0$, a given proportion of states are randomly initialized as infected, the others being considered as susceptible.

2.2 Time shifted averaged dynamics

For modelling and control purposes, we are interested in deriving a population-based model which approximates the dynamics generated by the individual-based model.

We are therefore looking at the dynamics of the average number of susceptible, infected and recovered individuals in the graph:

$$S(t) = \frac{1}{N}\sum_{j=1}^{N} P(X_j(t) = s), \quad I(t) = \frac{1}{N}\sum_{j=1}^{N} P(X_j(t) = i), \quad R(t) = \frac{1}{N}\sum_{j=1}^{N} P(X_j(t) = r).$$

To this end, we calculate the average of several time series associated with the same set of parameters $(\beta, \gamma, \kappa, N)$. In cases where the population size is particularly small or where super-spreaders drive the epidemic, two main difficulties arise: (i) some simulations lead to immediate extinctions whereas others lead to outbreaks, (ii) the onset of the epidemic occurs at random times. Consequently, as illustrated in Figure 2, computing the average trajectory via a naive average frequently leads to severe underestimations of the total number of infected individuals $I(t)$, as already noted in [23, Appendix A.2]. A standard way to overcome this issue is to compute the standard point average only after time-translating the time series so that the outbreaks all occur at the same time. Details on the calculation of the average trajectories can be found in Appendix A. Although it seems that this method mostly solves the issues mentioned above and gives robust results, as a safeguard, individual stochastic trajectories will also be displayed when plotting the results.

Although we are interested in the averaged dynamics, there is no associated closed differential model for these quantities. Indeed, without the time shift, the averaged quantities solve the following

differential equations (see [23]):

$$S' = -\beta_{\text{ind}}[SI], \tag{1}$$
$$I' = \beta_{\text{ind}}[SI] - \gamma I, \tag{2}$$
$$R' = \gamma I, \tag{3}$$

where the quantity $[SI]$ denotes the average number of edges connecting an infected and a susceptible individual:

$$[SI](t) = \frac{1}{\#C} \sum_{(j,k)\in C} P(X_j(t) = s, X_k(t) = i).$$

where C denotes the set of contacts and $\#C$ its size. Obtaining a closed system requires a relation between $[SI]$ and the variables S, I. For a homogeneous contact graph, the relation $[SI] = \alpha SI/N$ is valid in the large population limit. However, since we consider contact heterogeneity and time-shift, this relation is no longer valid.

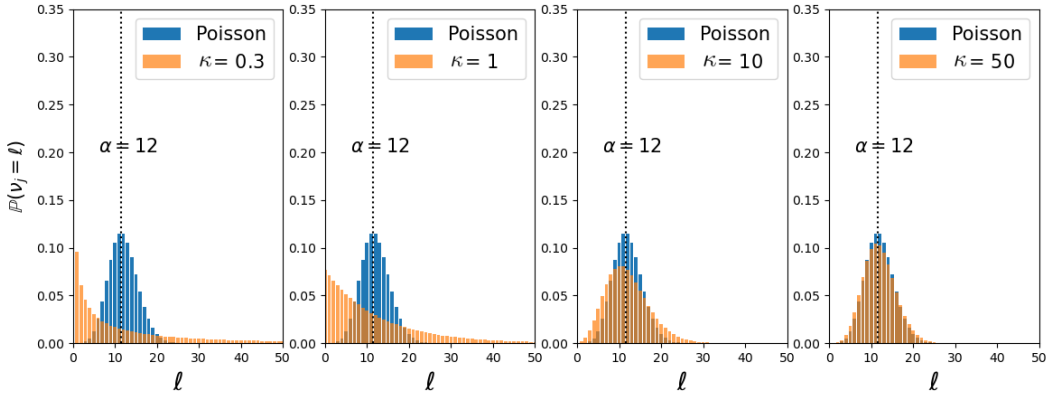


Figure 1: Pólya distribution (orange) of the number of contact ℓ for increasingly large dispersion coefficient κ and fixed mean α. Comparison with the Poisson distribution of mean α (blue) to illustrate convergence of the Pólya distribution to the Poisson one as $\kappa \to \infty$.

2.3 Optimal control problem of the IBM

In this section, we define a control problem for the individual-based model with heterogeneous contacts. The aim is to minimize the maximum number of infected individuals by including an important constraint reflecting the limited capacity of hospitals. In what follows, we will use the coefficients β and κ as optimization variables to contain the epidemic. Action on β can be seen as mandatory measures affecting the whole population (e.g. lockdowns or wearing masks indoor) contrary to action on κ which focuses on super-spreaders (e.g. cancelling of large events or imposing one to hold a valid COVID-19 certificate).

Suppose that at the initial time, only a small fraction of a population of size $1 = S(t) + I(t) + R(t)$ has contracted a disease whose transmission rate is estimated to be β_0 and that the coefficient of dispersion is approximately known to be κ_0. Furthermore, let $T > 0$ be the time horizon up to which we wish to study the effect of given health policies and $T_c < T$ be the non-negative time required for a sufficient number of secondary infections to occur and the health authorities to intervene. At this

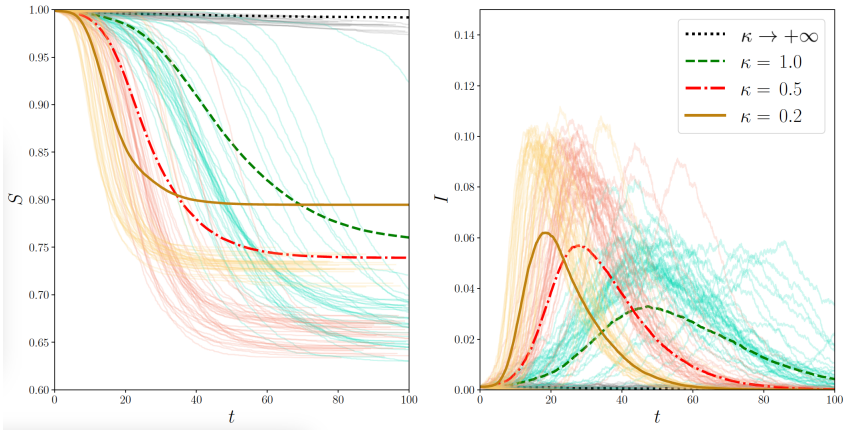


Figure 2: Simulations of $S(t)$ (left) and $I(t)$ (right) illustrating the insufficiency of a naive average: the delays in the start of the epidemic due to stochasticity lead to an underestimation of the instantaneous number of infected people. It also illustrates the importance of modelling the dispersion since here, in the homogeneous case ($\kappa = +\infty$, corresponding to a Poisson distribution), there is immediate extinction. $(n, \beta, \gamma, i_0) = (0.25, 0.15, 0.15, 0.001)$.

time, the state of the population is denoted by the non-negative numbers S_c and I_c. We consider health policies $b(\cdot) := \beta(\cdot)/\beta_0$ and $k(\cdot) := \kappa(\cdot)/\kappa_0$ with bounded values over the time interval $[T_c, T]$. More precisely, the set of admissible control is

$$\mathcal{U} = \left\{(b,k) \in L^\infty(T_c, T; \mathbb{R}^2) : b_{\min} \leq b(t) \leq 1,\ 1 \leq k(t) \leq k_{\max} \text{ a.e.}\right\}, \tag{4}$$

where $b_{\min} \in (0,1]$ and $k_{\max} > 1$ are given and reflect the fact that a perfect application of sanitary measures is unrealistic.

Towards an optimal control problem for the IBM. There are different aspects that we want to include in the definition of the optimal control problem:

- on the one hand, the application of sanitary measures can be detrimental in the long run, both to the mental health of the citizens but also to the economy. We therefore choose to integrate weights in the definition of the criterion, which represent the trade-offs a political decision-maker has to consider. This makes it possible to penalize or not certain types of measures (confinement or closure of certain public places) in the cost of control.

- on the other hand, we wish that the epidemic dies out as soon as possible without putting too much pressure on the health infrastructures (hospitals, intensive care units). This stress translates mathematically to proportions of infected individuals above a given threshold I_{hosp}.

These considerations lead us to balance the costs using a convex combination of terms involving three non-negative weights ω_β, ω_κ and ω_{hosp}. The last term of the cost function aims at penalizing strongly, say by $1/\varepsilon$ for a small positive ε, any control leading to proportions of infected individuals above a certain threshold $I_{\max} \in (I_{\text{hosp}}, 1]$. This constraint can be understood as a strong constraint such as the one on intensive care beds for example.

Thus, denoting by I_{IBM} the second component of the (S, I) solution to the IBM associated with a (b, k) health policy, the previous elements are taken into account in the fixed-time optimal control problem

$$\boxed{\inf_{(b,k)\in\mathcal{U}} J_{\text{IBM}}[b,k]} \tag{5}$$

relying on the cost functional J_{IBM} defined on the set of admissible controls $\mathcal{U}$ by

$$J_{\mathrm{IBM}}[b,k] = \frac{1}{2}\int_{T_c}^{T} \omega_\beta \left(1-b(t)\right)^2 + \omega_\kappa \left(k(t)-1\right)^2 + \omega_{\mathrm{hosp}} \left(\frac{I_{\mathrm{IBM}}(t)}{I_{\mathrm{hosp}}} - 1\right)_+^2 + \frac{1}{\varepsilon}\left(\frac{I_{\mathrm{IBM}}(t)}{I_{\max}} - 1\right)_+^2 \mathrm{d}t,$$

Notice that, in the above definition, the purpose of the positive part function is to avoid penalizing efficient health policies that limit the quantity of sick individuals to proportions lower than I_{hosp}.

3 Optimal control method

Determining solutions to the optimal control problem (5) for the IBM is a complex and computationally expensive task. The proposed strategy is to solve the optimal control problem for an associated population-based reduced model which faithfully reproduces the time-shifted averaged dynamics of the IBM. This population-based reduced model is constructed using a neural network and iteratively improved in order to best fit the the solution around the optimal control trajectory. The overall methodology is summarized in Figure 3.

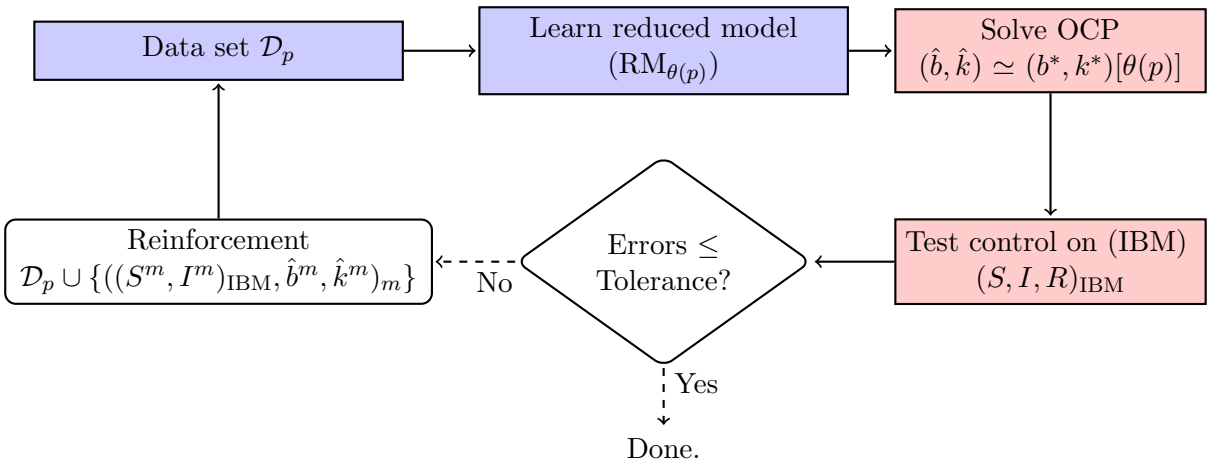


Figure 3: Blue blocks: learning of a reduced model ($RM_{\theta(p)}$) described in Section 3.1. Red blocks: resolution of the optimal control problem on the reduced model described in Section 3.2. White blocks: reinforcement step described in Section 3.3. Index p refers to the iteration of the reinforcement algorithm consisting in sequentially updating the data set and the reduced model.

3.1 Construction of a data-driven population-based reduced model

This subsection deals with the construction of a deterministic closure for the population-based model (1)-(2)-(3) which faithfully approximates the dynamics of the stochastic individual-based model. Using supervised machine learning techniques, the main goal is to capture, through an SIR-like system of autonomous differential equations, the effects of contact heterogeneity and population size on the dynamics of an epidemic. More precisely, we assume the following relationships hold between the state variables S, I, R and their time derivatives:

$$\begin{array}{rcl} S' &=& -F_\theta(S,I;n,\beta,\kappa), \\ I' &=& F_\theta(S,I;n,\beta,\kappa) - \gamma I, \\ R' &=& \gamma I, \end{array} \qquad (\mathrm{RM}_\theta)$$

where γ is still the individual recovery rate and the function $F_\theta : \mathbb{R}^5 \to \mathbb{R}$ is a parametrized incidence function [30] whose purpose is to approximate the term $\beta_{\text{ind}}[SI]$.

Its inputs are assumed to be both the instantaneous proportion of susceptible S and infected I individuals, as well as three positive parameters: the population size ratio $n = \min\{1, N/N_{\max}\}$, the mean transmission rate β and the dispersion coefficient κ. The introduction of the parameter $N_{\max}$, taken equal to $20,000$, allows us to extend our approach to large population sizes. From a practical point of view, we choose $N_{\max}$ empirically in such a way that the population dynamics did not vary anymore when increasing $N_{\max}$, but the approach stays valid for a larger value of $N_{\max}$. In other words, n is a ratio describing the closeness to the large population regime.

In order to ensure that the state variables remain in the interval $[0, 1]$, it is further assumed that the incidence function writes:

$$F_\theta(S, I; n, \beta, \kappa) = f_\theta(S, I; n, \beta, \kappa)\, SI, \tag{6}$$

where $f_\theta : \mathbb{R}^5 \to \mathbb{R}$ is another function, with the same inputs, called the *transmission rate function*. For the sake of clarity, we postpone to Appendix C details on regularity assumptions on both functions f_θ and F_θ, as well as well-posedness issues (existence and uniqueness of an absolutely continuous global solution with non-negative components) of System (RM_θ). From now on, we will assume that such regularity properties on F_θ are satisfied so that System (RM_θ) has a unique solution which is moreover Lipschitz, with non-negative components. Note that because $S+I+R = 1$ (remember that these quantities are proportions), all state variables of (RM_θ) are bounded from above by 1 and one of the three equations in (RM_θ) is redundant. Hence, the equation corresponding to recovered individuals is hereafter omitted.

Neural network structure. The function f_θ is built by means of a fully-connected neural network (or multilayer perceptron), which shows a great ability to learn non-linear functions [46, 17]. The function is then defined by composition of layers: each layer performs an affine transformation on its inputs and then applies a non-linear function (a so-called activation function) which is determined in advance. All the coefficients involved in the affine transformations are thus parameters of the function f_θ and correspond to the vector-valued parameter θ. The neural network structure is then determined by so-called hyperparameters, for instance, the number of layers, the input and output sizes of each layer and the activation functions used. The hyperparameters that we have selected are specified in Table 2 (left column).

Hyperparameters	Values	Learning parameters	Values
Neurons per layer	64 / 128 / 64 / 16	Initial learning rate	10^{-3}
No. inputs / outputs	5 / 1	Validation split	15%
Inputs normalization	Centered and reduced	Cost function	Mean squared error
Initialization	Orthogonal	Optimizer	Adam
Activations [Last]	ReLu [Linear]	Batch size	512
Learning rate schedule	Exponential	Epoch	15

Table 2: Practical details regarding the learning of the transmission rate function f_θ using the open-source Python library Keras [12].

Learning the transmission rate function from data. According to the authors in [1], there are two approaches for data-driven closures: model regression and trajectory regression. The model

regression approach consists in solving a regression problem on the quantity we aim to approximate

$$\sum_{([SI],S,I;n,\beta,\kappa)\in\mathcal{D}} |\beta_{\text{ind}}[SI] - F_\theta(S,I;n,\beta,\kappa)|^2$$

This approach seems costly since it requires computing the probability $[SI]$ using the IBM. In the second approach, the regression problem involves the left term $\mathrm{d}S/\mathrm{d}t$ instead of the term $\beta_{\text{ind}}[SI]$ which only requires simulating the IBM and approximating the time-derivative of S. In the remaining, we will use the latter approach.

We now discuss how the closure is constructed. The parameter θ is set in such a way that the transmission rate function f_θ best approximates the rate observed on the individual-based model simulations. More precisely, the parameter θ is found by regression, i.e. by minimizing the mean squared error:

$$L(\theta) = \sum_{(\tilde{S},S,I;n,\beta,\kappa)\in\mathcal{D}} \left\| f_\theta(S,I;n,\beta,\kappa) - \frac{\tilde{S}-S}{\Delta t\, SI} \right\|^2,$$

where $\mathcal{D}$ denotes the data set composed of samples $(\tilde{S},S,I;n,\beta,\kappa)$, where S,I are the average number of the susceptible and infected populations at a given time t, $\tilde{S}$ the value at time $t+\Delta t$, obtained after averaging individual-based simulations with parameters (n,β,κ). The hope is that, by considering a diverse enough data set, the function corresponding to an optimal parameter will manage to capture the underlying trend which relates the inputs to the output, especially for input values lacking in the data set. Details about the practical implementation of the learning algorithm are displayed in Table 2 (right column).

To generate the data set, parameters $(n,\beta,\kappa,I(0))$ are chosen randomly according to the distributions given in Table 3. The susceptible state at $t=0$ is then $S(0)=1-I(0)$. Next, we run the individual-based model over a given time interval. We then average the time-series of the corresponding susceptible and infected populations over 50 simulations at discrete times $t^m = m\Delta t$. For the simulations, we choose $\Delta t \simeq 0.28$. Repeating this process a significant number of times and storing the results makes up the training data set $\mathcal{D}$ which, in our case, contains about 7.4 millions of samples.

Parameters	Lower bound	Upper bound	Units	Interpretation
n	0.1	1	-	Population size ratio
β	0.075	0.9	days^{-1}	Transmission rate
κ	0.1	10	-	Dispersion coefficient
$I(0)$	10^{-4}	10^{-3}	-	Initial proportion of infected people

Table 3: Samples used to learn the function f_θ have their input parameters randomly drawn (uniformly for n,β and log-uniformly for κ and $I(0)$) in a subset of their possible values.

Global model validation Once the parameters defining the transmission rate function f_θ have been determined, the validation step is carried out to prevent over-fitting and evaluate the model accuracy. To do so, we select unseen values of (n,β,κ) in the ranges of interest (see Table 3) for which the population-based model (RM_θ) is numerically solved over a given time horizon. Then, the corresponding trajectory of (S,I) is compared to the one simulated via the (IBM) initialized with the same parameters. In each case, the time-series of the number of susceptible and infected individuals are compared based on their qualitative behaviour as well as quantitative error criteria. To be more

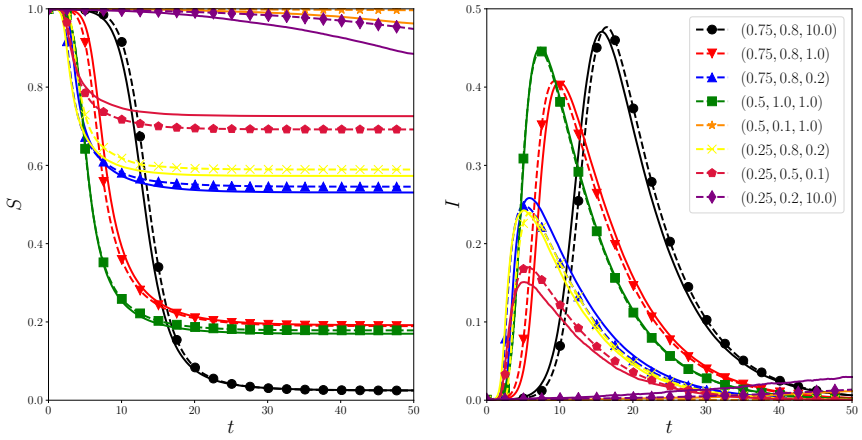


Figure 4: Response (susceptible evolution in the left, infected evolution in the right) of the learned model (RM_θ) to parameters that are constant over time. In the legend, these are specified in the format (n, β, κ). Learning is based on the data set $\mathcal{D}$. In dotted lines with markers: predictions of the (RM_θ) learned model. In continuous lines: average of the IBM trajectories.

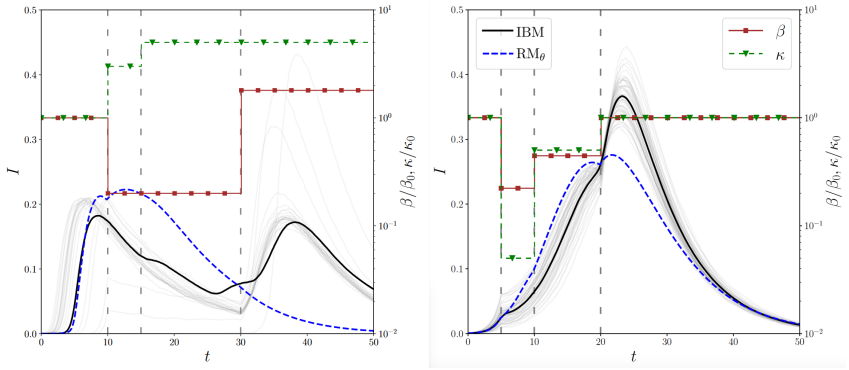
specific, we were mostly interested in the ability of the learned model (RM_θ) to correctly predict immediate disease extinctions, even in parameter regimes in which stochasticity plays a significant role (low dispersion coefficient κ or small population size ratio n).

On Figure 4, we illustrate our approach on several examples by comparing, for different parameters (n, β and κ), the average of the IBM and our data-driven population-based model. The outcome is pretty convincing and the reduced model shows decent accuracy for a wide range of parameter values (low population size ratio and dispersion coefficient, etc.). These results suggest that the model captures well the dynamics involved in a heterogeneous epidemic and might be used to analyze those. Indeed, the reduced model is overall accurate and capable of predicting immediate extinctions (see the purple and orange trajectories). In Appendix B, we present some other features of the reduced model and in particular its ability to predict key epidemiological quantities such as the threshold number $\mathcal{R}_0$.

Limitations of a global reduced model for time-dependent parameters The results of the previous paragraph show that the data-driven population-based model (RM_θ) is able to capture complex dynamics with super-spreading in a wide range of population size ratios. Moreover, the neural network structure of the incidence function makes numerical computations easier and the resulting function F_θ enjoys nice properties (at least local Lipschitz continuity) helping with the theoretical analysis of the ODE model.

However, all previous simulations were concerned with parameters β and κ that were constant over time. When considering time varying parameters and seeking to build a robust reduced model able to handle numerous values of n, β and κ (cf. Table 3), which is what would be required to control its dynamics, the model no longer works well as observed on Figure 5. Indeed, in the latter, we simulate two epidemics with time piece-wise constant β and κ parameters and the main observation is that the reduced model approximates well the average of the IBM until the first time at which the parameter-values change. Then, the accuracy worsens more and more as the number of changes increases. For instance, the reduced model fails to capture the last epidemic rebound in Figure 5 (a).

The incidence function is a function mapping a subset of $\mathbb{R}^5$ into $\mathbb{R}$. When training the reduced



(a) $n = 0.5$, $\beta = (0.5, 0.1, 0.1, 0.9)$, $\kappa = (0.2, 0.6, 1, 1)$

(b) $n = 0.75$, $\beta = (0.9, 0.2, 0.4, 0.9)$, $\kappa = (10, 0.5, 5, 10)$

Figure 5: Comparison of the trajectory I between the (RM$_\theta$) model (dashed line blue) and IBM (in black on the figure) in two cases where the parameters β (red squares markers) and κ (green triangle markers) vary over time.

model using constant coefficients the samples generated by the simulations only cover a smaller part of this subset compared to the varying coefficient case. For example we observe cases where S is very close to zero and I is reaching a peak. A possible explanation of the difficulty to approximate the target function when dealing with time-varying parameters could be the conjunction of a (globally) badly-behaved function (difficult to learn) and a subset of data larger than in the constant case.

Provided that the above explanation is correct, there is hope that learning the reduced model based on a training data set involving fewer combinations of the parameters will lead to better results. This issue together with finding the right subset of parameters to include in the data set is dealt with in Section 3.3. For now, we will attempt to find an optimal control of the reduced model based on the optimal control problem (5) defined for the IBM.

3.2 Optimal control algorithm of the reduced model

Recall that we choose to use a reduced model to determine an optimal policy for the IBM. However, as mentioned in the previous paragraph, the reduced model is not accurate anymore whenever β and κ vary over time. This is problematic. Indeed, this lack of accuracy means that an optimal control for the reduced model will likely fail to contain the epidemic on the IBM, since a candidate control is likely to involve time-variations of $\beta(t)$ and $\kappa(t)$. To avoid this, we will use the principle of *model-based reinforcement learning* [26]. That is, we will learn a local model around a trajectory, compute the associated control and then learn the controlled trajectory. By repeatedly applying these steps, we expect to obtain a control that is relevant to the original individual-based model.

To implement the above strategy, we first need to define and be able to solve an optimal control for the data-driven population-based SIR model constructed in Section 3.1. This section is thus dedicated to applying the standard theory of optimal control (OC) to the learned model (RM$_\theta$) involving the incidence function F_θ whose expression has been obtained thanks to a neural network. Recall that the weights of the latter neural network have been optimized so that the output of F_θ accurately estimates the rate of change of the proportion of susceptible individuals (see Section 3.1). However, it is not clear that the partial derivatives of $F_\theta = F_\theta(S, I; n, \beta, \kappa)$ are also good approximations of the corresponding quantities. Thus, when solving the equations numerically, difficulties may arise due to the fact that the tools of OC theory rely heavily on the differentiation of the system dynamics with respect to the state variables and control.

Reduced system under control. The equations describing the dynamics of the reduced system under the admissible health policies $(b,k) \in \mathcal{U}$ are given over the time interval $[T_c, T]$ as

$$\begin{pmatrix} S' \\ I' \end{pmatrix} = g\left(S, I, bv(k), k\right), \qquad \begin{pmatrix} S \\ I \end{pmatrix}(T_c) = \begin{pmatrix} S_c \\ I_c \end{pmatrix}, \tag{7}$$

where we have defined the right-hand side by

$$g : \mathbb{R}^4 \ni (a,b,c,d) \mapsto (-F_\theta(a,b;n,c\beta_0,d\kappa_0), F_\theta(a,b;n,c\beta_0,d\kappa_0) - \gamma b) \in \mathbb{R}^2.$$

Let us also comment on the choice of transmission rate $bv(k)$ appearing in model (7). The mapping v is defined as a non-increasing function of k by

$$v : [1, k_{\max}] \ni k \mapsto 1/\left(1 + \log_{10}(k)\right) \in \mathbb{R}.$$

It is a relatively simple way to account for the fact that controlling super-spreaders (e.g. closing down certain public places or introducing mandatory COVID-19 certificates) invariably has an influence on the whole population. Mathematically, we take this into account by expressing that the action on κ (the action on the super-spreaders) also has a small influence on β (on the whole population). Introducing an effective rate $bv(k)$ enables us to couple both the controls. On top of that, we will from now on make the following regularity assumptions on the function F_θ:

$$F_\theta : \Omega \to \mathbb{R} \text{ belongs to } W^{1,\infty} \text{ and is of the form given in Equation (6)}, \tag{$\mathcal{H}_{F_\theta}$}$$

where $\Omega = [0,1]^3 \times [\beta_0 b_{\min}, \beta_0] \times [\kappa_0, \kappa_0 k_{\max}]$. Under this assumption, System (7) has a unique global solution that belongs to $W^{1,\infty}(T_c, T; \mathbb{R}^3)$, according to Appendix C. Moreover, since the total population size is constant in time, it is enough to consider only the (S,I) equation in the control problem.

Cost function definition. The optimal control problem for the reduced model mimics the one we defined for the IBM (see Equation (5)), except that the cost functional J also penalizes the cost of the control. More precisely, we consider the optimal control problem

$$\boxed{\inf_{(b,k)\in\mathcal{U}} J[b,k]} \tag{OCP}$$

where, denoting by $I_{b,k}$ the second component of the (S,I) solution to (7) associated with a (b,k) health policy,

$$J[b,k] = \frac{1}{2}\int_{T_c}^{T} \omega_\beta \left(1 - b(t)\right)^2 + \omega_\kappa \left(k(t) - 1\right)^2 + \omega_{\text{hosp}} \left(\frac{I_{b,k}(t)}{I_{\text{hosp}}} - 1\right)_+^2 + \frac{1}{\varepsilon}\left(\frac{I_{b,k}(t)}{I_{\max}} - 1\right)_+^2 \mathrm{d}t.$$

On the existence of an optimal control. In the problem (OCP), the control intervenes in the dynamics in a strongly nonlinear way. It is known that, for such control problems, it is not guaranteed that a solution exists and phenomena such as relaxation or homogenization of the minimizing sequences, leading to numerical pathologies, may occur. For this reason, we will in fact slightly modify the previous optimal control problem by adding a regularization term to the cost function J. We have decided to consider here a BV regularization[2] of (OCP), by introducing the following

[2]Recall that if Ω denotes an open set of $\mathbb{R}^n$, f belongs to $\mathrm{BV}(\Omega)$ whenever f belongs to $L^1(\Omega)$ and

$$\mathrm{TV}(f) < +\infty \quad \text{where} \quad \mathrm{TV}(f) = \sup_{\substack{\psi \in C_c^1(\Omega;\mathbb{R}^n) \\ \|\psi\|_{L^\infty(\Omega)} = 1}} \int_\Omega f \,\mathrm{div}\,\psi.$$

The Banach space BV is endowed with the norm $\|\cdot\|_{\mathrm{BV}(\Omega)}$ defined by $\|f\|_{\mathrm{BV}(\Omega)} := \|f\|_{L^1(\Omega)} + \mathrm{TV}(f)$.

problem:

$$\boxed{\inf_{(b,k)\in\mathcal{U}} J_\delta[b,k]}, \tag{OCP$_\delta$}$$

where $\delta > 0$ is a parameter standing for the strength of the regularization and

$$J_\delta[b,k] = J[b,k] + \delta(\mathrm{TV}[b] + \mathrm{TV}[k]).$$

Such a regularization is interesting from several points of view. For example, if the control is of the bang-bang type[3], the BV regularization imposes a maximum number of switches, which may reflect an economic cost. On the other hand, this term imposes that the control belongs to the BV Banach space, which leaves the freedom to choose the control functions among a large variety of functions, not necessarily continuous.

We claim that Problem (OCP$_\delta$) has a solution (b,k). Since the arguments are rather standard, we refer to Appendix C for additional explanations. Let us now introduce the first-order optimality conditions for this problem, which are at the heart of the numerical solution algorithm that we then implement. The optimality conditions for this problem involve the notion of subdifferential $\partial\,\mathrm{TV}$ of the total variation operator. For the sake of readability, the proof of the following result is postponed to Appendix C.

Theorem 3.1. *Let $M_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k)$ denote the matrix*

$$M_\theta = \begin{pmatrix} -\partial_1 F_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k) & -\partial_2 F_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k) \\ \partial_1 F_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k) & \partial_2 F_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k) - \gamma \end{pmatrix} \tag{8}$$

and let $[p_1,q_1,p_2,q_2]$ denote the solution of the (linear) adjoint system

$$-\frac{d}{dt}\begin{pmatrix} p_1 \\ q_1 \\ p_2 \\ q_2 \end{pmatrix} = \begin{pmatrix} M_\theta^\top & 0_{\mathcal{M}_2(\mathbb{R})} \\ 0_{\mathcal{M}_2(\mathbb{R})} & M_\theta^\top \end{pmatrix}\begin{pmatrix} p_1 \\ q_1 \\ p_2 \\ q_2 \end{pmatrix} + \left(\frac{\omega_{\mathrm{hosp}}}{I_{\mathrm{hosp}}}\left(\frac{I}{I_{\mathrm{hosp}}}-1\right)_+ + \frac{1}{I_{\max}\varepsilon}\left(\frac{I}{I_{\max}}-1\right)_+\right)\begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \tag{9}$$

completed with the terminal conditions

$$p_1(T) = q_1(T) = p_2(T) = q_2(T) = 0.$$

The functionals $\mathcal{U} \ni [b,k] \mapsto (S,I) \in [W^{1,\infty}(T_c,T)]^2$ and J are differentiable, where the pair (S,I) denotes the solution of (7) associated with the control choice (b,k). Furthermore, the differential of J is given by

$$\langle dJ[b,k],[h_1,h_2]\rangle \;=\; \int_{T_c}^{T} \left(h_1\partial_b J(b,k) + h_2\partial_k J(b,k)\right)\,\mathrm{d}t \tag{10}$$

for every $[b,k]\in\mathcal{U}$ and every admissible perturbation[4] $[h_1,h_2]$, where

$$\begin{aligned} \partial_b J(b,k) &= \omega_\beta\,(b-1) + \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}\cdot\begin{pmatrix} -\beta_0 v(k)\partial_4 F_\theta \\ \beta_0 v(k)\partial_4 F_\theta \end{pmatrix} \\ \partial_k J(b,k) &= \omega_\kappa\,(k-1) + \begin{pmatrix} p_2 \\ q_2 \end{pmatrix}\cdot\begin{pmatrix} -\beta_0 bv'(k)\partial_4 F_\theta - \kappa_0\partial_5 F_\theta \\ \beta_0 bv'(k)\partial_4 F_\theta + \kappa_0\partial_5 F_\theta \end{pmatrix}. \end{aligned}$$

[3] In other words, if the control takes only two distinct values.

[4] More precisely, we call "admissible perturbation" any element of the tangent cone $\mathcal{T}_{[b,k],\mathcal{U}}$ at $[b,k]$ to the set $\mathcal{U}$. The cone $\mathcal{T}_{[b,k],\mathcal{U}}$ is the set of functions $[h_1,h_2]\in L^\infty(T_c,T;\mathbb{R}^2)$ such that, for any sequence of positive real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$ decreasing to 0, there exists two sequences of functions $h_{i,n}\in L^\infty(T_c,T)$ converging to h_i, $i=1,2$, as $n\to+\infty$, and $[b,k]+\varepsilon_n[h_{1,n},h_{2,n}]\in\mathcal{U}$ for every $n\in\mathbb{N}$.

Now, let us assume that $(\mathcal{H}_{F_\theta})$ *is true and let* (b,k) *denote a solution to Problem* (OCP_δ)*. There exist* $T_b \in \partial\,\mathrm{TV}(b)$ *and* $T_k \in \partial\,\mathrm{TV}(k)$ *such that*

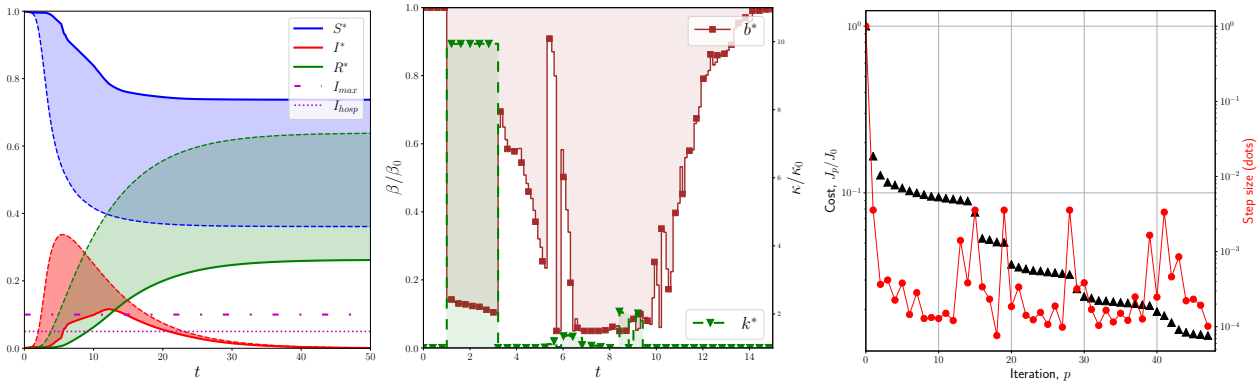
$$\forall (B,K) \in L^\infty(T_c,T;[b_{\min},1]) \times L^\infty(T_c,T;[1,k_{\max}]), \quad \begin{cases} \langle \partial_b J - T_b, B-b\rangle_{L^2(T_c,T)} \geq 0 \\ \langle \partial_k J - T_k, K-k\rangle_{L^2(T_c,T)} \geq 0. \end{cases}$$

Remark 3.2 (subdifferential of the total variation)**.** *Let us recall that, according to [10, Proposition I.5.1], the subdifferential of the total variation is given by*

$$\partial\,\mathrm{TV}(b) = \{\eta \in C^0([T_c,T]) \mid \|\eta\|_\infty \leq 1 \text{ and } \int \eta \mathrm{d}b = \mathrm{TV}(b)\}.$$

From a practical point of view, we will not directly use these optimality conditions which remain rather abstract written as they are. Instead, we will regularize the TV term and introduce a descent method using the differential calculation established in Theorem 3.1, using the adjoint state (p_1,q_1,p_2,q_2). The implemented algorithm is introduced in Appendix D.

As explained earlier, we will consider an optimal control computed from the reduced nonlinear model that we will apply to the individual-based model. To numerically compute an estimate of the control solving the (OCP_δ) problem, we use a direct approach consisting in discretizing the differential systems involved via a regular $\mathcal{S}$ subdivision of the $[T_c,T]$ interval with step-size Δt. This also allows us to transform the optimal control problem into a nonlinear program whose decision variables are the control values (b,k) evaluated at each point of $\mathcal{S}$. The optimization of the latter values is performed using a relatively simple adaptive step projected gradient algorithm, using a linear search of the step size taken in the direction of greatest descent[5]. In order to limit the computational cost, the latter online search is performed using a gradient-free method called *golden-section search* [39]. Details are provided in Appendix D.



(a) Trajectories with / without (dashed line) control.

(b) Evolution of the controls over the first 15 days. Afterwards, the controls remain constant.

(c) Cost (black triangle markers) and descent step evolution.

Figure 6: Example of controlled trajectories. Parameters: $(n,\beta_0,\kappa_0) = (0.6,0.8,0.4)$, $T_c = 1$, $T = 50$, $\Delta t = 0.1$, $\delta = 10^{-7}$, $\varepsilon = 10^{-2}$, $(I_{\mathrm{hosp}},I_{\max}) = (5\%,10\%)$, $(\omega_\beta,\omega_\kappa,\omega_{\mathrm{hosp}}) = (0.2,0.2,0.6)$. Here 50 iterations are required for the gradient descent to converge. On the right figure, J_p denotes the value of J_δ at the p-th iteration.

In Figure 6, we give an example of control computations on the reduced model independently of the individual-based model. On the left, we see an example of uncontrolled (dashed lines) and

[5]In other words, such that the next control leads, after projection onto the set of constraints, to the greatest possible decrease in the value of the cost functional

controlled (solid lines) trajectories. We obtain that the maximum number of infected is exceeded during a very short time compared to the uncontrolled trajectories. Since the beginning of the control is delayed (by T_c) and since it is not realistic to set β (resp. κ) too low (resp. too high), it is sometimes not possible to avoid exceeding $I_{\max}$. Let us also note that the implemented gradient algorithm allows *a priori* to determine only local minima: there is no guarantee that a global minimizer has been obtained. In the middle, we plot the associated controls of the coefficients β and κ. Finally, we show on the right the evolution of the cost function and size of the descent step. As expected, we obtain a decreasing cost function. This example illustrates the accuracy of the algorithm used to find optimal controls of the reduced system. Based on the latter, we will now present the main algorithm of this paper which gathers all together the tools we developed so far to construct approximate solutions to the control problem (5) for the IBM.

3.3 Reinforcement learning strategy to improve controls

In the model-based reinforcement learning problems, there are two families of approaches: the global model-based methods and the local model-based methods (see e.g. [24, 8]). At the end of Section 3.1, we have seen that it is difficult to build a versatile (global) model capable of handling time-varying parameters (β, κ). Therefore, we propose to use a more local approach.

Recall that in Section 3.1, the complexity of the individual-based model was simplified to obtain a reduced model (RM_θ) consisting of only two deterministic ODEs, at the cost of richness and accuracy. The reduced model approaches the dynamics of the individual-based model over a wide range of constant parameters (i.e. large ranges of values of n, β, and κ). In Section 3.2, we looked for an optimal health policy specifically for the reduced model. A question then naturally arises: to what extent can a control minimizing the cost function of the optimal control problem for the reduced system be used to obtain a "good" control for the original individual-based model? As explained in Section 3.1, this naive approach is lacking because a versatile reduced model can fail to be accurate when β and κ vary over time (which is the case when applying a control policy). We now seek to overcome this issue by constructing a reduced model specialized around a given set of parameters (n, β_0, κ_0) and whose sole purpose is to approximate only locally, but very accurately, the dynamics of the IBM. This is done by restricting the number of parameter combinations in the training data set and raises the following question: how does one select the relevant subset of parameters $\mathscr{S}$ to construct a local reduced model?

The subset $\mathscr{S}$ depends on the choice of (n, β_0, κ_0) which makes it difficult to determine before starting the learning process. Thus, we propose to follow an approach inspired by the theory of *model-based reinforcement learning* in which the subset $\mathscr{S}$ is built "on the fly" and determined by "trial and error". More precisely, the approach consists in alternating between a learning step on the reduced model and an optimal control step. At each iteration, the control trajectory is recomputed based on the IBM and added to the data set used to train the reduced model.

It is common practice to use a valid linear model just around a state $(S(t^m), I(t^m))$. However, since we are able to efficiently control a non-linear system, we propose to compute a valid non-linear system around a complete trajectory. This choice appears to be a compromise between a local and a global model. We will now describe the algorithm used to control the IBM.

Local model-based reinforcement approach. The idea is to sequentially training the network until an optimal control strategy for the corresponding reduced model manages to equally well contain the epidemic when simulated on the IBM. Suppose that, at the beginning of an epidemic, the authorities have recorded a percentage of infected people I_0 with an estimated coefficient of dispersion κ_0. Moreover, we assume that the transmission rate β_0 of the disease is known, and that

an optimal control problem was defined for the local model (7) we aim to construct.

To begin with, the function $F_{\theta(0)}$ is trained on $\mathcal{D}_0$, a very small fraction (e.g. 5-10%) of the shuffled training dataset $\mathcal{D}$, introduced in Section 3.1. In other words, the neural network defining F_θ receives information about the IBM dynamics corresponding to a reduced but representative region of the parameter space (n, β, κ). The main objective of this exploration concerns stability: the reduced model becomes accurate enough to take into account the dominant behaviour of the IBM, which ensures that the solution to the associated ODE does not blow-up under control. Recall that the definition of the population size ratio n has been introduced and commented in Section 3.1.

Let us now describe a current iteration of the control algorithm using the MPC approach. Assume that the p-th iteration of the algorithm begins and that the neural network has a weight configuration $\theta(p)$. Figure 3 already showed a flowchart of the proposed approach based on the MPC method. Following the steps described in Section 3.2, we can estimate a control (b^*, k^*) optimally driving the reduced model (7) based on the knowledge induced by the weight configuration $\theta(p)$. Depending on the fineness of the partition of the time interval of interest, this optimal health policy may correspond to measures evolving freely on an unrealistic time scale (a few days or a few hours). For this reason, the control (b^*, k^*) is approximated via a regression tree (computed using the SK-learn library) by two piece-wise constant functions, denoted $(\hat{b}, \hat{k})$, taking at most 8 different values over a time horizon of 200 days. Starting from the initial configuration corresponding to the operating point $(S_0, I_0, n, \beta_0, \kappa_0)$, the IBM is then simulated under this last policy and the corresponding scenario denoted by $(S, I, R)_{\text{IBM}}$.

We decide to stop the algorithm when the obtained health policy is sufficiently efficient. The stopping criterion requires a sufficient decrease in the value that the cost function J_{IBM} evaluated at the current iteration control takes. More precisely, we say that the current control $(\hat{b}, \hat{k})$ is *acceptable* with tolerance $\tau_{\text{RL}} > 0$ if

$$J_{\text{IBM}}[\hat{b}, \hat{k}] \leq \tau_{\text{RL}} J^0_{\text{IBM}},$$

where J^0_{IBM} is the cost associated with the control-free IBM solution. Recall that T_c is the time at which the intervention of the health authorities begins (detailed in Section 2.3). In addition to requiring that the control is acceptable, the p-th scenario must, under the same $(\hat{b}, \hat{k})$ control, be associated with a lower cost than the cost of the reduced model under the $(\hat{b}, \hat{k})$ control. In other words, the algorithm should not stop unless the inequality $J_{\text{IBM}}[\hat{b}, \hat{k}] \leq J[\hat{b}, \hat{k}]$ holds. This stopping criterion allows us to relate the performance of the control on the reduced model and on the IBM.

Since the success of the algorithm depends upon the ability of the reduced model to accurately predict the output of the IBM, the stopping criterion also involves the following three error metrics: we retrieve global information by computing the discrete L^2-norm of the difference between the reduced model and the IBM for the state variables S and I, and by estimating the mismatch between the final proportion R_∞ of removed people, defined by (13). Accuracy is also assessed by measuring the delay between the time at which the infection peak (IP) occurs, respectively for the IBM and reduced model. The numerical values of the associated tolerances $\tau_{L^2}, \tau_{R_\infty}$ and τ_{IP} are shown in Table 4.

If the stopping criterion is not satisfied, then the reduced model is strengthened by training the weights $\theta(p+1)$ on a larger training set $\mathcal{D}_{p+1}$ containing not only $\mathcal{D}_p$, but also the local information corresponding to the p-th scenario $(S, I, R)_{\text{IBM}}$ as well as the parameters defining the candidate control $(\hat{b}, \hat{k})$. The above sequence of steps repeats until these criteria are satisfied, at which point the output of this algorithm is the $(\hat{b}, \hat{k})$ health policy corresponding to the last iteration.

Note that, in an attempt to reduce the computational cost and to escape as much as possible from local minima wells, at each p reinforcement step, the optimal control algorithm is initialized with the control obtained at the end of the previous reinforcement step. Moreover, the more we advance in the reinforcement algorithm, the more precise the optimal control algorithm must be

(more iterations, smaller step sizes). The reasoning behind this last point is that in the first few iterations of reinforcement, high accuracy is not so important because the behavior of the reduced model under control is likely to be an unfaithful approximation of that of the IBM.

4 Numerical results

In this section, we present the behaviour of our optimal control method (Section 3.3) for different regime of parameters. In each case, we provide the quantities n, β_0, κ_0 and the number of iterations of the reinforcement learning algorithm. We plot on each of them the trajectories relative to susceptible individuals on the left, the trajectories for the infected in the middle, and the control for the IBM model (red for the β control, green for the κ control) on the right. Scales for β/β_0 and κ/κ_0 are shown respectively on the left and right sides of the right figure. In Table 4, we specify the parameters common to all the test cases. If one of these parameters were to change, it would be indicated in the legend of the figure.

Parameters	Values	Param.	Val.	Param.	Val.
S_0	99.95%	$(I_{\text{hosp}}, I_{\max})$	$(0.025, 0.1)$	T_c	1
I_0	0.05%	$(\omega_\beta, \omega_\kappa, \omega_{\text{hosp}})$	$(0.2, 0.2, 0.6)$	T	200
γ	1/6	$(b_{\min}, k_{\max})$	(0.1,10)	Δt	2/7
τ_{RL}	10^{-3}	ε	10^{-2}	τ_{R_∞}	10^{-3}
τ_{L^2}	1	δ	10^{-7}	τ_{IP}	6

Table 4: Numerical values of the main parameters involved in the reinforcement algorithm (Section 3.2) and (OCP_δ). These are common to all results shown in Section 4.

Since the legend of the figures is the same for all the tests, let us explain the notations we use:

- IBM denotes an average trajectory (based on 50 simulations) for the IBM without control,
- IBM^C denotes an average trajectory (based on 50 simulations) for the IBM with the final control obtained by the reinforcement learning algorithm (individual trajectories are in grey),
- RM^C denotes a trajectory produced by the reduced model trained only on the initial data set $\mathcal{D}_0$ with the final control of the algorithm,
- RM^{RLC} denotes a trajectory produced by the reduced model after model-based reinforcement with the final control of the algorithm.
- $\hat{b}$ and $\hat{k}$ denote the final controls, piece-wise constant, provided by the algorithm. The vertical dashed segments indicate times at which changes in control values occur.

The presentation of the results is divided into three parts. In the first one, we illustrate the efficiency and flexibility of the proposed algorithm via simulations representative of the different regimes observed with the IBM (super-spreaders or homogeneous contact distribution, little or large population sizes, etc.).Then, we focus on how the number of iterations of reinforcement affects the reduced model accuracy and the effectiveness of the associated control. Lastly, we turn our attention to the limitations and drawbacks of the proposed reinforcement learning approach.

4.1 Algorithm versatility in the different parameter regimes

We now present results corresponding to different parameter configurations. To begin with, we consider two cases where the population size ratio and κ are large (large population and homogeneous contact regime), meaning that they are in the validity regime of the classical SIR dynamics. We observe the results on this type of configuration in Figures 7-8. In the first one, for the black curve, we observe that the strong constraint on $I_{\max}$ is preserved and the number of infected stays close to I_{hosp}. In this case, the reduced model learned only with $\mathcal{D}_0$ (blue curve) is arguably as accurate as the reinforced one. In the second case (Figure 8), similar results are obtained, but we observe that the reinforcement step allows to increase the accuracy of the reduced model (red versus blue curves) which subsequently improves the control efficiency.

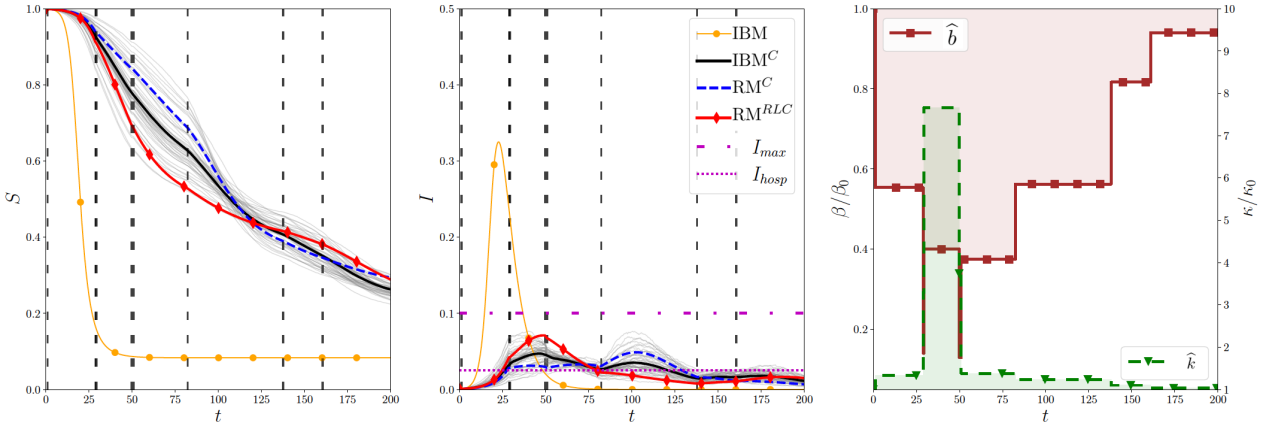


Figure 7: (Large population size and large dispersion) $n = 0.95$, $\beta_0 = 0.5$, $\kappa_0 = 9$, 10 iterations. In the classical SIR regime, very few iterations are needed to generate an effective control.

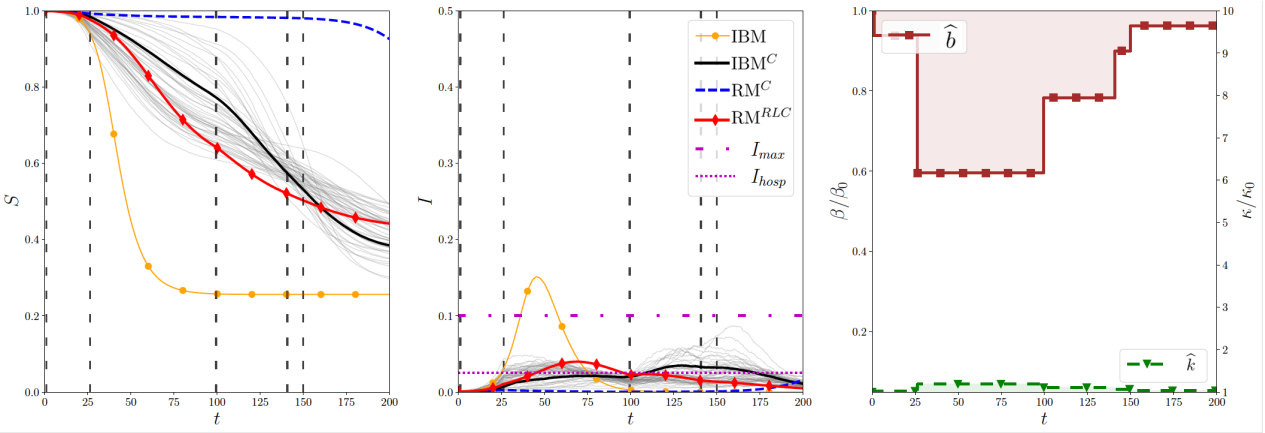


Figure 8: (Large population size and large dispersion) $n = 0.85$, $\beta_0 = 0.8$, $\kappa_0 = 10$, 6 iterations. In the classical SIR regime, very few iterations are needed to generate an effective control.

In Figures 9-10, we stray away from classical population-level regimes by considering intermediate population sizes and dispersion coefficients. In this slightly more complicated regime, stochastic behaviours are commonly observed in the IBM simulations. Nevertheless, in the first case (Figure 9), the reinforced reduced model (red curve) faithfully approximates the IBM average trajectory (black curve) and the control is effective enough to ensure that I does not exceed the threshold $I_{\max}$.

Similarly the control policy remains satisfying in the second case (Figure 10), although it is more difficult for the reduced model to capture the averaged random behaviour of the IBM due to the very low value of the parameter n.

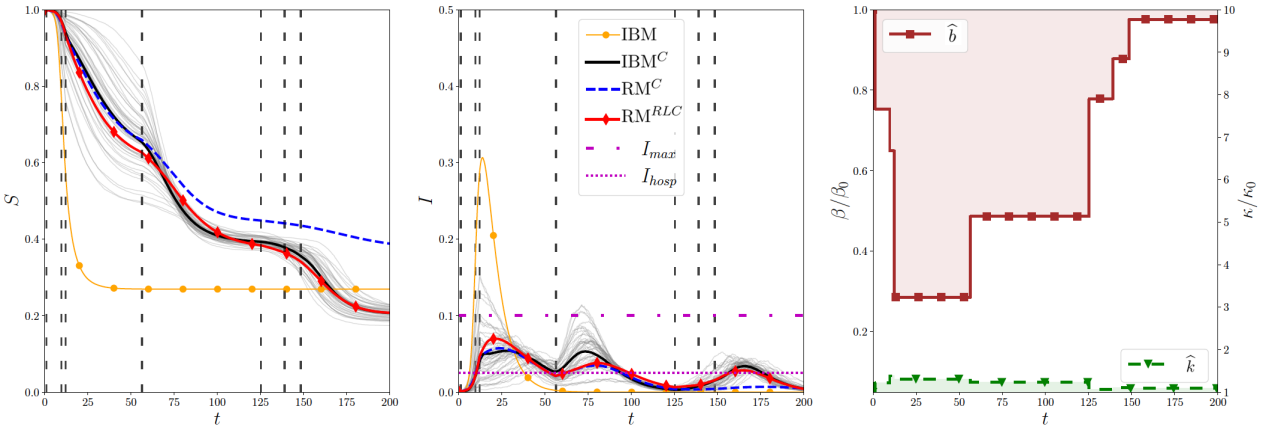


Figure 9: (Intermediate population size and dispersion) $n = 0.6$, $\beta_0 = 0.5$, $\kappa_0 = 1$, 15 iterations. The proposed algorithm seems effective in spite of the particularly strong stochastic behaviour of the IBM.

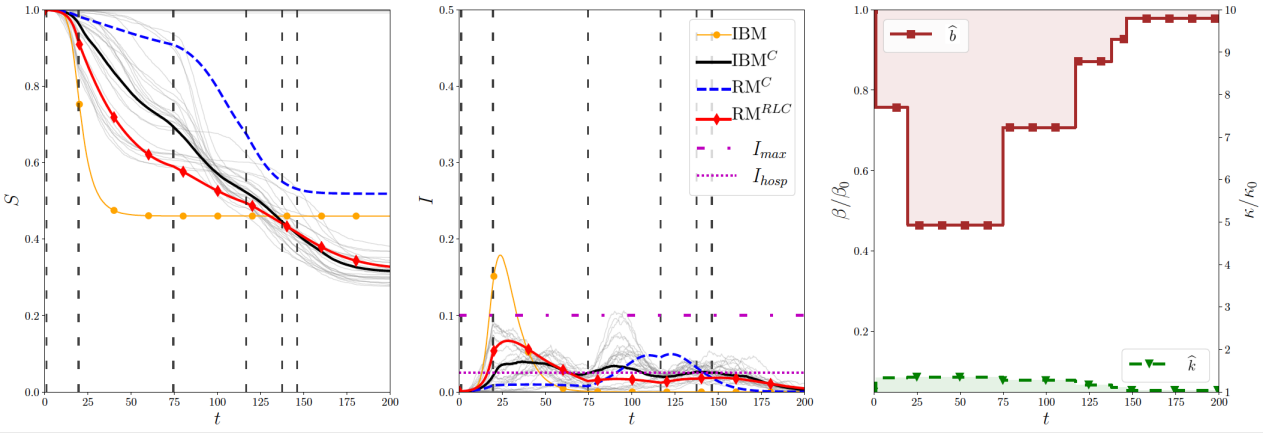


Figure 10: (Intermediate population size and dispersion) $n = 0.2$, $\beta_0 = 0.3$, $\kappa_0 = 0.8$, 7 iterations.

In the following results, we increase one step further the difficulty by considering even lower population size ratios and dispersion coefficients. Figure 11 deals with a very large heterogeneity in the population (low κ). The resulting control is accurate and the black curve remains far away from the strong constraint. By comparing the red curve to the blue one, we also observe that the reinforcement learning allows to improve a lot the reduced model which seems to be more and more faithful to the IBM trajectory, even if we observe a discrepancy between the final values of S_∞. Indeed, since $S_\infty = 1 - \int_0^\infty \gamma I$, the accumulation of non-compensating errors on I seems to lead to a poor estimate of S_∞. Note that, in the optimal control problem, the cost functional does not involve S_∞. In the end, this is not an issue because the main objective of the method is to compute an optimal control for the IBM which reduces the infection peak. Accurately predicting the behaviour of $S(t)$ is not necessary to achieve this goal.

Figure 12 highlights results in a very low population size regime where the graph and stochastic

effects are important, yet the results are also convincing. It is worth noting that the reinforcement learning procedure (red curve) drives to an improvement of the reduced model (blue curve) which in the end captures correctly the infection peaks.

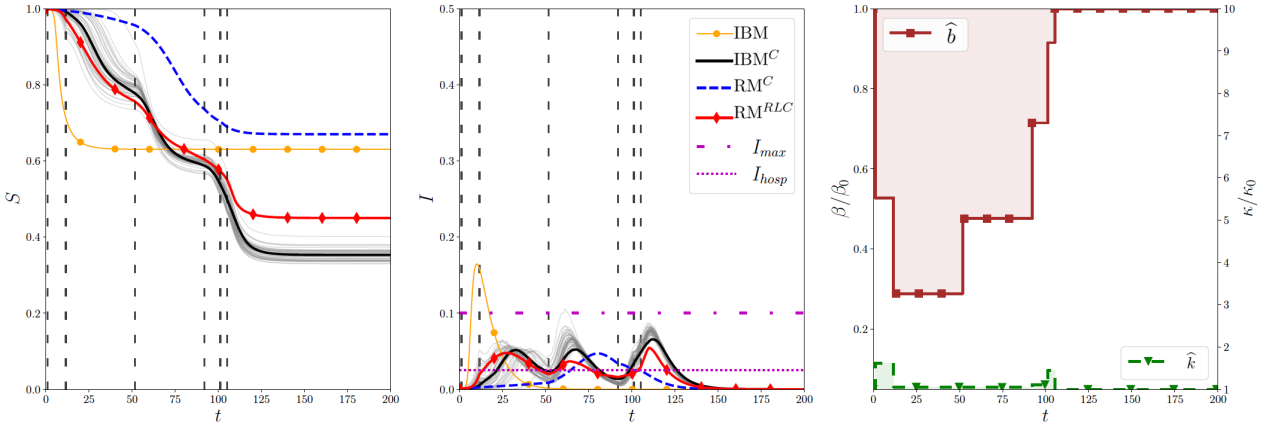


Figure 11: (Large population size and small dispersion) $n = 0.8$, $\beta_0 = 0.3$, $\kappa_0 = 0.2$, 44 iterations with $(\omega_\beta, \omega_\kappa, \omega_{\text{hosp}}) = (0.5, 0.1, 0.4)$. Populations with large contact heterogeneity are prone to epidemic rebounds, requiring more iterations of the proposed algorithm to meet the stopping criteria.

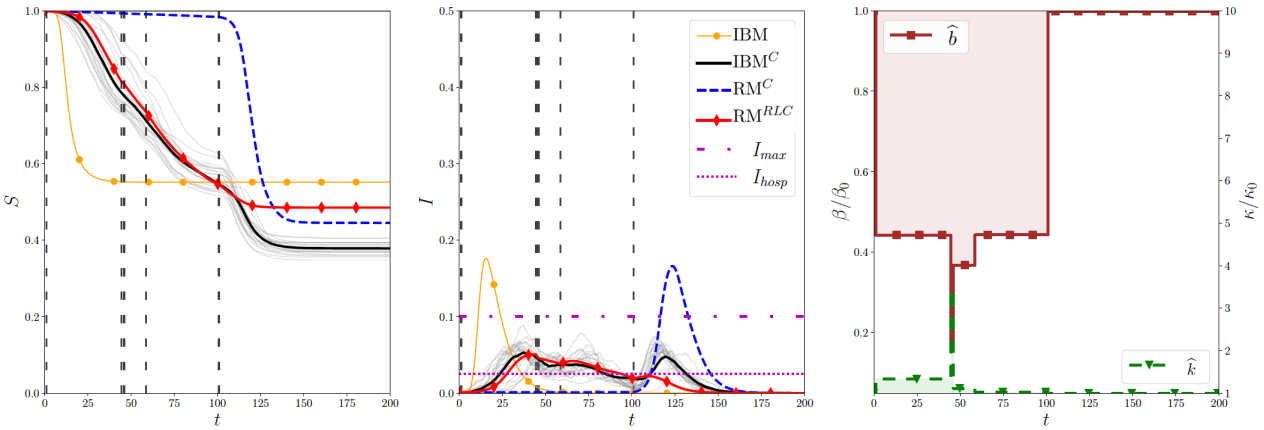


Figure 12: (Small population size and small dispersion) $n = 0.2$, $\beta_0 = 0.3$, $\kappa_0 = 0.4$, 23 iterations. Significant accuracy improvement (red) of the reduced model compared to the naive approach (blue).

4.2 Overall improvement of the control strategy with the number of iterations

In this second subsection, we investigate how the number of reinforcement iterations affects the results of the algorithm. First, we consider in Figure 13 a case with large dispersion. We observe that making some additional iterations increases a little bit the accuracy of the reinforcement learning reduced model and allows to compute a better control, since the amplitude of the infection peak remains less close to the constraint $I_{\max}$.

We now consider a low dispersion regime (with super-spreaders) in Figure 14. Since the epidemic is small, capturing the threshold is generally more complicated for the reduced model. Here, we compare one training after 18 and 34 iterations respectively, as well as a new training with 17 iterations and a smaller initial data set. As before, we observe that increasing the number of

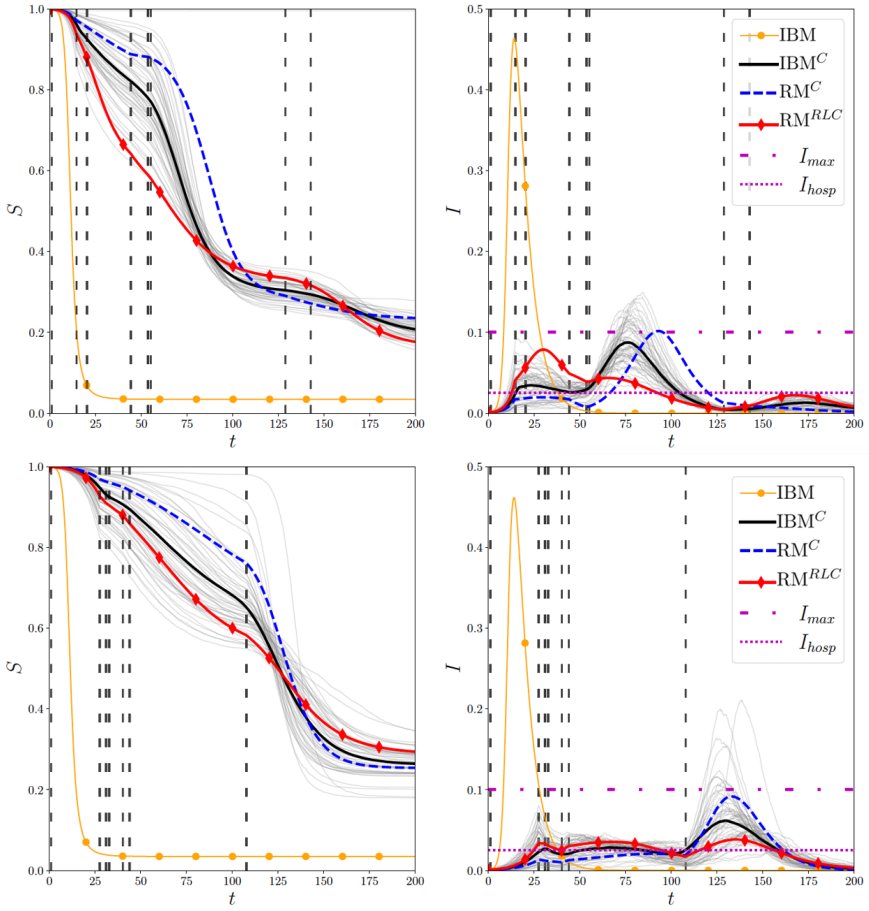


Figure 13: (Variation of the number of iterations of the control algorithm with large dispersion) $n = 0.5$, $\beta_0 = 0.8$, $\kappa_0 = 9$. 14 iterations for the top versus 30 for the bottom, leading to an improvement of the control acceptability (lower infection peak).

iterations improves the accuracy of the reduced model and control. Indeed at the top of Figure 14, the run corresponding to 18 iterations does not preserve the strong constraints, contrary to the second one in the middle of Figure 14 which generates a trajectory satisfying the constraints. However, the reduced model is not flawless and previous tests show that this impacts the accuracy of the control.

To improve control accuracy, we propose in this case to reduce the size of the initial data set $\mathcal{D}_0$. This modification allows us to obtain a better reduced model and comparable control. A possible explanation is that the initial training may lead the neural network to learning a trajectory that deviates too far from the test case, making it difficult to explore the space of admissible trajectories. In other words, its ability to adapt to new samples may be impaired. This shows that the size of the data set $\mathcal{D}_0$ and its diversity may impact the efficiency of the algorithm.

Figure 15 deals with a test case involving moderate dispersion and population size ratio. This example illustrates that usually, during the algorithm, the control improves while the reduced model may momentarily worsen. Indeed, after only 8 iterations, the control fails to contain the epidemic (most stochastic trajectories of the IBM violate the strong constraint $I_{\max}$) even if the reduced model is qualitatively and quantitatively accurate. At the expense of model accuracy, making 8 additional iterations (middle plot) improves the control which now mitigates the peak of the average IBM trajectory (black curve), but not of all individual ones (grey trajectories). However, making

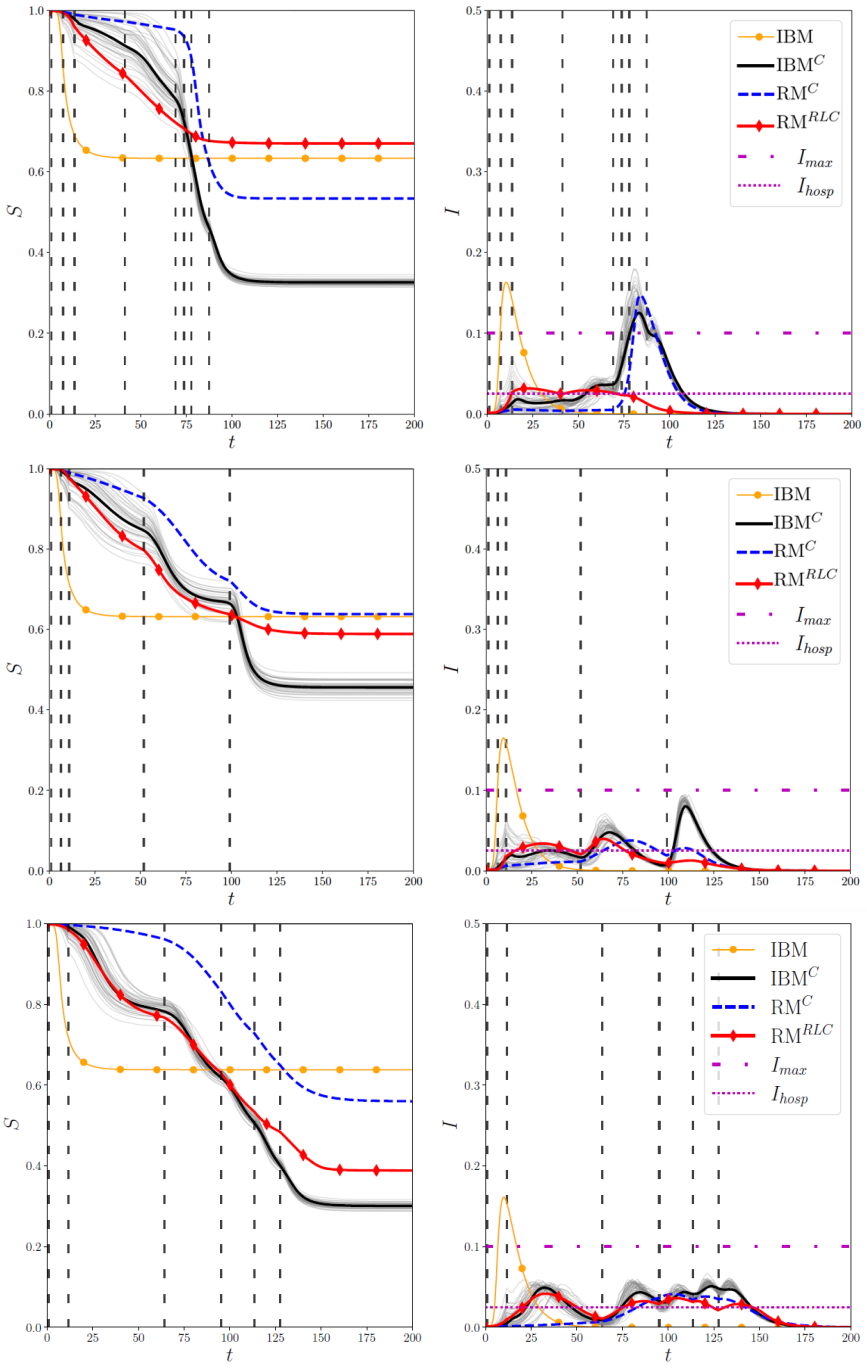


Figure 14: (Variation of the number of iterations of the control algorithm with small dispersion) $n = 0.8$, $\beta_0 = 0.3$, $\kappa_0 = 0.2$. 18 iterations for the top, 34 iterations for the middle and 17 iterations with a (40%) smaller data set $\mathcal{D}_0$ for the bottom. The size and diversity of the initial data set affects the efficiency of the proposed algorithm.

about twice as many iterations (bottom plot, black curve) leads to a control that is acceptable with tolerance $5 \cdot 10^{-3}$ (i.e. $J_{\text{IBM}}[\hat{b}, \hat{k}] \leq 5 \cdot 10^{-3} J^0_{\text{IBM}}$) and to a faithful reduced model (red curve).

These results show that generally, whenever more iterations are made, the reduced model will overall become better. However, this improvement depends on the degree of randomness involved in

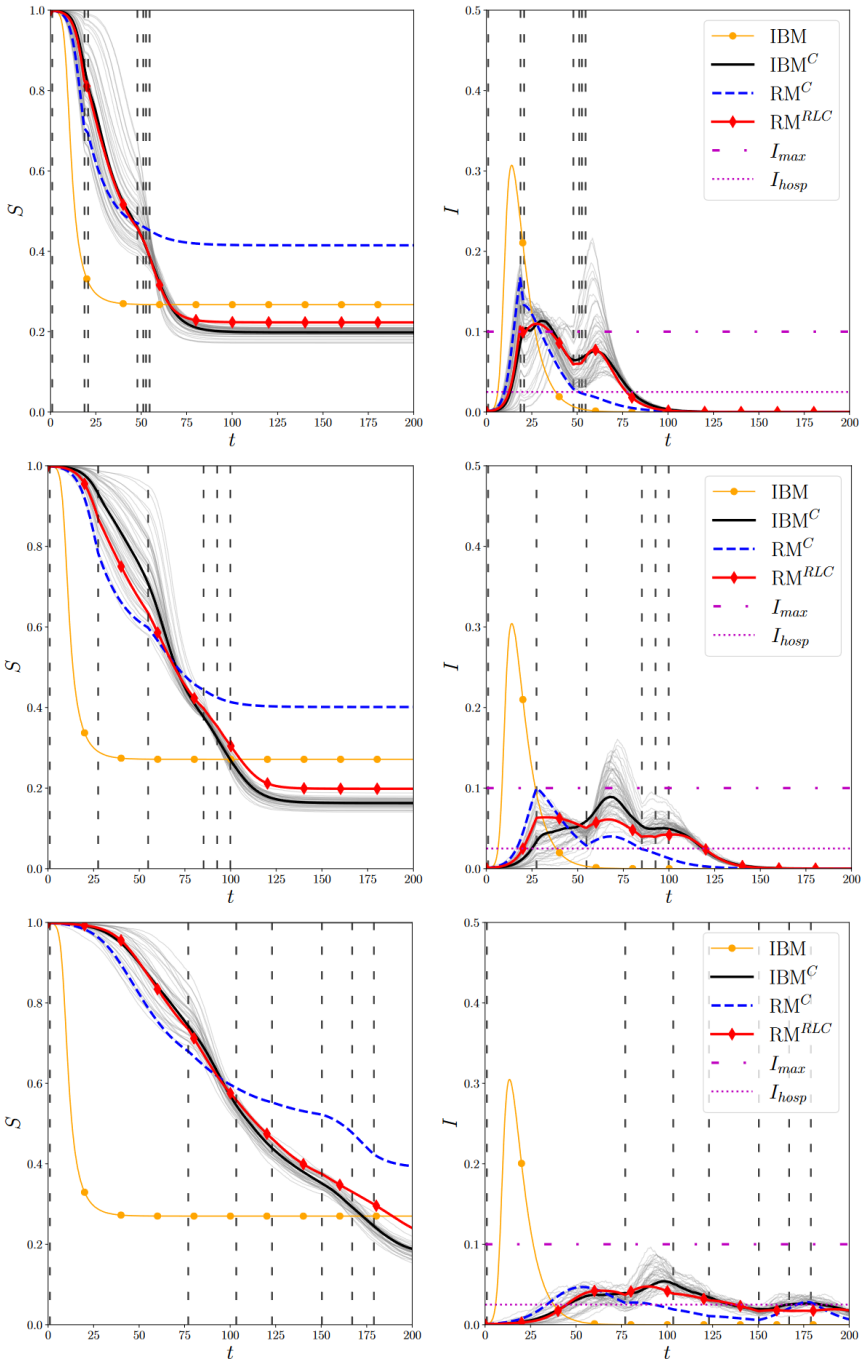


Figure 15: (Variation of the number of iterations of the control algorithm with intermediate dispersion) $n = 0.6$, $\beta_0 = 0.5$, $\kappa_0 = 1$. Respectively 8, 16 and 35 iterations for the top, middle and bottom. The control effectiveness usually improves throughout the iterations, even if the reduced model may momentarily worsen.

the parameter regime at stake and it is not given that an increase in the accuracy of the model will lead to a better control. Indeed, we sometimes observe efficient controls associated with unfaithful reduced models. But, generally speaking, when the model becomes good, so does the control. That is, the "convergence" of the reduced model towards the IBM trajectory seems to guarantee that the

corresponding control is accurate, although sometimes not the best.

4.3 Over-fitting effect and limitations of the proposed algorithm

Our algorithm may fail to both generate an accurate reduced model and an acceptable control. We identified two possible mechanisms at the root of these seemingly rare failures. The first one relates to an *over-fitting-like effect*. More precisely, throughout the reinforcement algorithm iterations, the neural network is generally fed with more and more similar samples. Thus, at a certain point, the network predictions are likely to deteriorate in an attempt to capture the dynamics associated with the mean IBM. For instance, it is seen in Figure 16 that, after 18 iterations, the reduced model is very accurate even if the control is unsatisfactory. Hence, to improve the effectiveness of the control on the IBM, 7 additional reinforcement iterations are made. This strategy ends up paying off, but the improvement comes at the expense of a significant deterioration in the accuracy of the reduced model. Moreover, if we were to continue, the predictions of the reduced model might not improve. Considering early-stopping or a posteriori model selection among saved intermediate models (by tracking performance criteria) may help overcoming these difficulties.

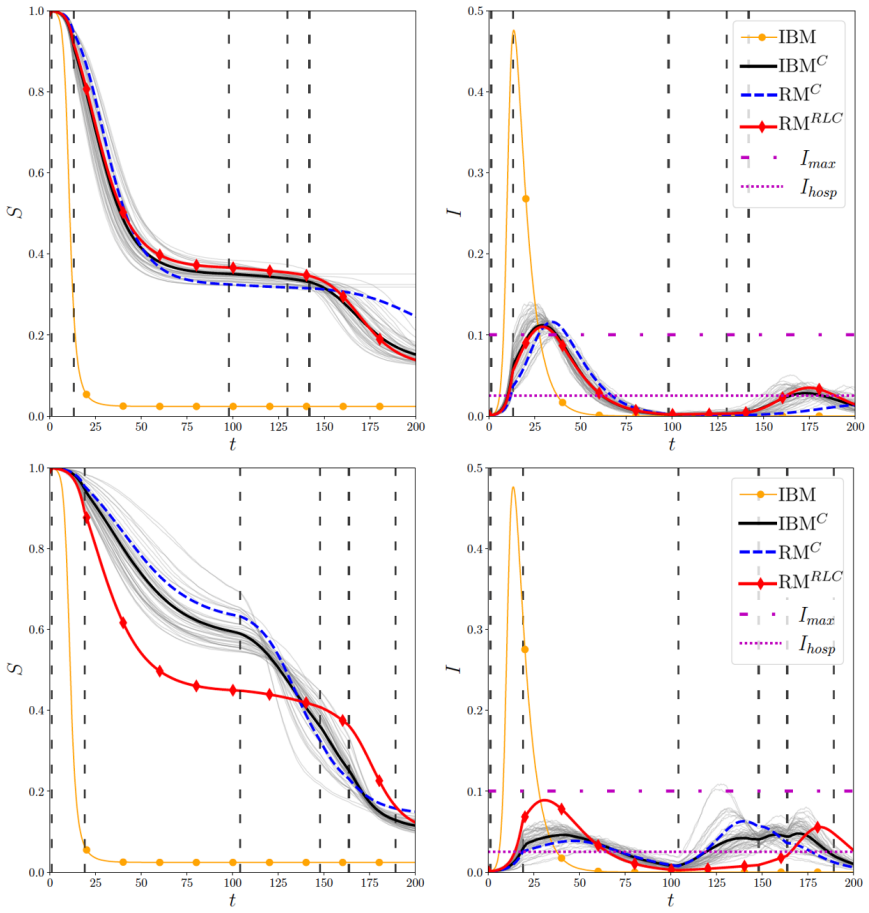


Figure 16: (Possible deterioration of the reduced model with algorithm iterations) $n = 0.95$, $\beta_0 = 0.8$, $\kappa_0 = 9$. Top: 18 iterations. Bottom: 25 iterations. Illustrates the need for a posteriori model selection among saved intermediate reduced model when accuracy matters.

The second mechanism we identified relates to the limitations of approximating a fundamentally stochastic system (the IBM) by means of a deterministic reduced model. In other words, when

the parameter regime is such that randomness is dominating the behaviour of the system (e.g. all three parameters n, β and κ are low), approximating the average trajectory of the IBM is very challenging, as can be seen in Figure 17. Indeed, in this setting, computing the incidence function $F_\theta(S(t), I(t); n, \beta, \kappa)$ is particularly sensitive to errors. In addition, this range of parameters is scarcely represented in the initial data set $\mathcal{D}_0$ (see Table 3). The latter could partially explain the difficulties observed in Figure 17. However, enriching $\mathcal{D}_0$ with more samples may not be recommended since previous examples have shown that the reduced model would probably end up having difficulty specializing around the controlled trajectory.

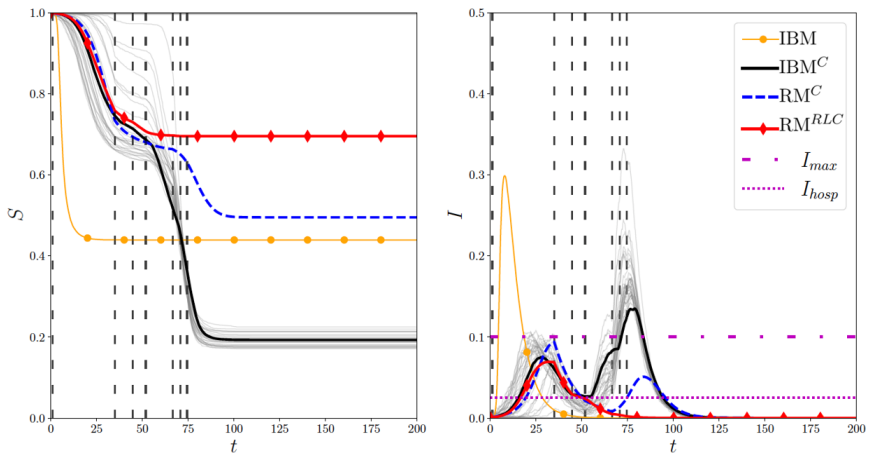


Figure 17: (Difficulties in capturing the reduced dynamics in stochastic regime, with small population size, dispersion and transmission rate) $n = 0.15$, $\beta_0 = 0.7$, $\kappa_0 = 0.4$, 33 iterations. Limitations due to a deterministic approximation of a stochastic system: poor results after numerous iterations.

5 Conclusion and perspectives

In this work, we proposed a method to control a stochastic individual-based epidemic model which takes into account super-spreaders. To this end, we proposed a model-based reinforcement approach which consists in alternating between the learning phase of a reduced model and the control phase. Our approach could be interpreted as a Model Predictive Control (MPC) type method. In the literature on model-based methods, it is common to either use global models for all states, or linear and local models around the current state. Here, we struck a compromise by building a non-linear model valid in a certain sub-region of the admissible values of the controls. To solve the control problem for the reduced model, we then use optimal control approaches for ODEs. The iterative algorithm allows us to build a control for the original IBM based on the one computed for the learned reduced model. The results show the ability of the algorithm to compute efficient controls in classical regimes (low dispersion, large population) as well as in more complicated regimes, as they are generally more stochastic, for which the population size ratio is small and contact heterogeneity large. This algorithm involves building a reduced SIR model, relying on a neural network, which takes into account the effects of small population and dispersion effects associated with super-spreaders. Constructing the latter model also provides tools to study the effects of contact heterogeneity (dispersion) in epidemics. Indeed, since we learn a SIR-type model where the incidence function $F_\theta(S, I; n, \kappa, \beta)$ is differentiable, we could, for instance, derive by analytical means a formula for the basic reproduction ratio $\mathcal{R}_0$. Hence, it is possible to investigate the dependence of the latter on the dispersion and the population size while this is usually difficult to estimate for individual-based

models. More generally, building a reduced SIR-type model with a neural network from heavier simulations may be an interesting way to study phenomena that are not easily understood in large models such as the epidemic threshold, the group epidemic threshold, etc. One of the limitations of our approach is that we only aim at controlling the mean trajectory of the individual-based model, but it could be relevant and interesting to take into account its variance.

Acknowledgements

The last author were partially supported by the ANR Project "TRECOS".

References

[1] S. E. Ahmed, S. Pawar, O. San, A. Rasheed, T. Iliescu, and B. R. Noack. On closures for reduced order models—a spectrum of first-principle to machine-learned avenues. *Physics of Fluids*, 33(9):091301, 2021.

[2] G. An, B. G. Fitzpatrick, S. Christley, P. Federico, A. Kanarek, R. M. Neilan, M. Oremland, R. Salinas, R. Laubenbacher, and S. Lenhart. Optimization and Control of Agent-Based Models in Biology: A Perspective. *Bulletin of Mathematical Biology*, 79(1):63–87, 2017.

[3] M. A. Bhouri, F. S. Costabal, H. Wang, K. Linka, M. Peirlinck, E. Kuhl, and P. Perdikaris. COVID-19 dynamics across the US: A deep learning study of human mobility and social behavior. *Computer Methods in Applied Mechanics and Engineering*, 382:113891, 2021.

[4] P.-A. Bliman and M. Duprez. How best can finite-time social distancing reduce epidemic final size? *J. Theoret. Biol.*, 511:110557, 12, 2021.

[5] P.-A. Bliman, M. Duprez, Y. Privat, and N. Vauchelet. Optimal immunity control and final size minimization by social distancing for the SIR epidemic model. *J. Optim. Theory Appl.*, 189(2):408–436, 2021.

[6] R. Capobianco, V. Kompella, J. Ault, G. Sharon, S. Jong, S. Fox, L. Meyers, P. R. Wurman, and P. Stone. Agent-Based Markov Modeling for Improved COVID-19 Mitigation Policies. *J. Artif. Int. Res.*, 71:953–992, sep 2021.

[7] E. A. Coddington and N. Levinson. *Theory of ordinary differential equations*. McGraw-Hill Book Co., Inc., New York-Toronto-London, 1955.

[8] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to derivative-free optimization*. SIAM, 2009.

[9] O. Diekmann, J. Heesterbeek, and J. Metz. On the Definition and the Computation of the Basic Reproduction Ratio R0 in Models for Infectious Diseases in Heterogeneous Populations. *Journal of Mathematical Biology*, 28(4), 1990.

[10] I. Ekeland and R. Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999. Translated from the French.

[11] B. Elie, C. Selinger, and S. Alizon. The source of individual heterogeneity shapes infectious disease outbreaks. *Proceedings of the Royal Society B: Biological Sciences*, 289(1974):20220232, 2022.

[12] F. Chollet and others. Keras, 2015.

[13] R. Fujie and T. Odagaki. Effects of superspreaders in spread of epidemic. *Physica A: Statistical Mechanics and its Applications*, 374(2):843–852, 2007.

[14] T. Garske and C. Rhodes. The effect of superspreading on epidemic outbreak size distributions. *Journal of Theoretical Biology*, 253(2):228–237, 2008.

[15] D. 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.

[16] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Number vol. 80 in Monographs in Mathematics. Birkhäuser, 1984.

[17] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. 2016.

[18] V. Grimm, A. Heinlein, A. Klawonn, M. Lanser, and J. Weber. Estimating the time-dependent contact rate of SIR and SEIR models in mathematical epidemiology using physics-informed neural networks. Technical Report, Universität zu Köln. 5, 2020.

[19] A. Hagberg, P. Swart, and D. S Chult. Exploring network structure, dynamics, and function using NetworkX. 2008.

[20] M. Hintermüller, M. Holler, and K. Papafitsoros. A function space framework for structural total variation regularization with applications in inverse problems. *Inverse Problems*, 34(6):064002, 2018.

[21] Y. Kim, H. Ryu, and S. Lee. Agent-Based Modeling for Super-Spreading Events: A Case Study of MERS-CoV Transmission Dynamics in the Republic of Korea. *International Journal of Environmental Research and Public Health*, 15(11):2369, 2018.

[22] I. Z. Kiss, D. M. Green, and R. R. Kao. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Mathematical Biosciences*, 203(1):124–136, 2006.

[23] I. Z. Kiss, J. C. Miller, and P. L. Simon. *Mathematics of Epidemics on Networks*. Springer International Publishing, 2017.

[24] S. Koziel and L. Leifsson. *Surrogate-based modeling and optimization*. Springer, 2013.

[25] E. B. Lee and L. Markus. *Foundations of optimal control theory*. Wiley New York, 1967.

[26] S. Levine. Model Based Reinforcement Learning. Lecture at Berkeley University.

[27] C. Liu. A microscopic epidemic model and pandemic prediction using multi-agent reinforcement learning. arXiv:2004.12959, 2020.

[28] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[29] J. Long, A. Khaliq, and K. Furati. Identification and prediction of time-varying parameters of covid-19 model: a data-driven deep learning approach. arXiv :2103.09949, 2021.

[30] M. Martcheva. *An introduction to mathematical epidemiology*. Number 61 in Texts in applied mathematics. Springer, 2015.

[31] S. T. McQuade, R. Weightman, N. J. Merrill, A. Yadav, E. Trélat, S. R. Allred, and B. Piccoli. Control of COVID-19 outbreak using an extended SEIR model. *Math. Models Methods Appl. Sci.*, 31(12):2399–2424, 2021.

[32] J. Miller and T. Ting. EoN (Epidemics on Networks): A fast, flexible Python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software*, 4(44):1731, 2019.

[33] T. Mkhatshwa and A. Mummert. Modeling Super-spreading Events for Infectious Diseases: Case Study SARS. arXiv :1007.0908, 2010.

[34] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, probability and computing*, 7(3):295–305, 1998.

[35] C. Nowzari, V. M. Preciado, and G. J. Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Transactions on Control of Network Systems*, 4(2):159–169, 2015.

[36] C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.

[37] R. Pastor-Satorras, C. Castellano, P. Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87, 2014.

[38] V. M. Preciado and M. Zargham. Traffic optimization to control epidemic outbreaks in metapopulation models. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 847–850. IEEE, 2013.

[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.

[40] M. d. V. Rafo and J. P. Aparicio. Simple epidemic network model for highly heterogeneous populations. *Journal of Theoretical Biology*, 486:110056, 2020.

[41] C. Ramirez, R. Sanchez, V. Kreinovich, and M. Argaez. $\sqrt{x^2+\mu}$ is the most computationally efficient smooth approximation to $|x|$: a proof. 2013.

[42] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[43] P. van den Driessche and J. Watmough. Further Notes on the Basic Reproduction Number. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, volume 1945 of *Lecture Notes in Mathematics*, pages 159–178. Springer Berlin Heidelberg, 2008.

[44] P. Van Mieghem and J. Omic. In-homogeneous virus spread in networks. *arXiv:1306.2588*, 2013.

[45] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Transactions On Networking*, 17(1):1–14, 2008.

[46] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola. Dive into Deep Learning, 2021.

[47] S. N. Zisad, M. S. Hossain, M. S. Hossain, and K. Andersson. An Integrated Neural Network and SEIR Model to Predict COVID-19. *Algorithms*, 14(3), 2021.

Appendices

Appendix A Averaging the IBM output

Suppose we are interested in averaging $P \in \mathbb{N}^*$ runs of the IBM over the time interval $[0, T]$ and let $M \geq 2$ be the number of points of a regular subdivision of this time interval with time-step Δt.

For each trajectory $p \in \{1, 2, \ldots, P\}$, run the IBM and refer to the resulting discrete output as S_p and I_p, both of which are elements of $\mathbb{R}^M$. Their values at time $m\Delta t$ are respectively denoted S_p^m and I_p^m for $m \in \{1, 2, \ldots, M\}$. Note that we have $R_p = 1 - S_p - I_p$.

The p-th trajectory is considered an outlier whenever the size of recovered population ends up being underestimated in the following sense

$$R_p^M - R_p^0 \leq 0.8 \times R_{\max},$$

where $R_{\max} = \max\{R_p^M,\, p = 1, \ldots, P\}$. In other words, all trajectories leading to immediate extinction are excluded since they would otherwise pull down the pointwise values of the mean trajectory.

The next step is to find the average time of the first epidemic onset. For each p, let

$$\tau_p = \min\{m\Delta t : I_p^m - I_p^0 > 10^{-3},\, m = 1, \ldots, M\}, \tag{11}$$

if the involved set is non-empty and zero otherwise. Denote by P' the number of trajectories for which $\tau_p > 0$. Based on these values, compute the mean time

$$\bar{\tau} = \frac{1}{P'} \sum_{p=1}^{P'} \tau_p,$$

and find for each trajectory p the number of time-steps $d_p \in \mathbb{Z}$ by which the outbreak time is delayed (or in advance) with respect to the mean value $\bar{\tau}$, that is

$$\forall p, \qquad d_p = \left\lfloor \frac{\tau_p - \bar{\tau}}{\Delta t} \right\rfloor.$$

Before averaging the trajectories, time-translate each trajectory p by d_p time-steps Δt and denote $\tilde{S}_p, \tilde{I}_p$ the resulting vectors. To keep vectors of the same size, we extend the vector by constant values on the left or right depending on the sign of the translation:

for $d_p > 0$,

$$\tilde{S}^m = \begin{cases} S^{d_p+m} & \text{if } 1 \leq m \leq M - d_p, \\ S^M & \text{if } M - d_p + 1 \leq m \leq M, \end{cases}$$

for $d_p < 0$,

$$\tilde{S}^m = \begin{cases} S^0 & \text{if } 1 \leq m \leq |d_p|, \\ S^{|d_p|+m} & \text{if } |d_p| + 1 \leq m \leq M. \end{cases}$$

We therefore implicitly assume that the trajectories do not vary too much at the beginning and end of the simulations on a time scale $|d_p|\Delta t$. Lastly, we compute the point-wise average according to

$$\forall m \in \{1, 2, \ldots, M\}, \qquad \bar{S}^m = \frac{1}{P'} \sum_{p=1}^{P'} \tilde{S}_p^m, \quad \text{and} \quad \bar{I}^m = \frac{1}{P'} \sum_{p=1}^{P'} \tilde{I}_p^m.$$

Note that in Definition (11), the threshold 10^{-3} offers a decent compromise. Indeed, the higher this value is, the more accurate the estimation of the family $(\tau_p)_p$ is, but at the same time, the higher

the risk that some very rare trajectories reach the threshold much later (or earlier) than the others, resulting in a biased mean value $\bar{\tau}$.

We wish to draw the reader's attention to the following observation: when we run the IBM with piece-wise constant parameters β and κ, epidemic rebounds may occur several times in a given simulation, e.g. in Figures 5a and 12. Nevertheless, since it is in the early stage of the epidemic that immediate extinctions are the most likely (due to stochasticity and very low proportions of infected people), translating the individual trajectories solely based on the time of the first epidemic onset remains *a priori* a reasonable assumption.

Appendix B Byproduct of our approach: estimating key epidemiological quantities

We trained the population-based model (RM_θ) which is expected to faithfully capture not only the global dynamics arising from individual variation, but also the impact of super-spreaders on an epidemic. In this subsection, we are working towards defining a threshold number via the so-called next-generation matrix theory and estimating the size of an epidemic. The goal is not so much about getting precise quantitative results regarding those epidemiological indicators, but rather to make qualitative statements and gain insight into how the parameters at stake, namely n, β and κ, interact and influence them. Indeed some quantities like the epidemic threshold are very useful for the epidemiologists but, contrary to what is the case for the classical homogeneous models, not easy to calculate when super-spreaders are taken into account [28]. In order to obtain smoother results, we constructed another reduced model dedicated to the calculation of these epidemiological indicators. This model involves a larger neural network which was trained on a data set containing many additional parameter configurations.

B.1 Estimating a threshold number.

In the case of population-based models and under suitable hypothesis, the *next-generation matrix* theory introduced by Diekmann *et al.* [9] offers a systematic framework to define a threshold number whose properties are identical to the well known $\mathcal{R}_0$ in the traditional SIR model. By analogy, this threshold number will hereafter be called $\mathcal{R}_0$. Informally, it provides information about the stability of the disease-free equilibrium (DFE) $(S, I) = (1, 0)$ in the population-based model (RM_θ). That is, if $\mathcal{R}_0 < 1$, then the DFE is locally asymptotically stable and the epidemic dies out; if not, then it is unstable and an outbreak occurs [30, 43].

Having numerically checked that the learned incidence function F_θ satisfies all the hypotheses of the next-generation matrix theory (mainly dealing with positivity) [9], the calculation is straightforward and unambiguous. The threshold number $\mathcal{R}_0$ can be seen as a scalar function of the three parameters (n, β, κ) given by:

$$\mathcal{R}_0(n, \beta, \kappa) = \left. \frac{\partial_I F_\theta}{\gamma} \right|_{(S=1, I=0; n, \beta, \kappa)}. \tag{12}$$

We refer to Appendix B.3 for a detailed derivation of this formula. Equation (12) suggests that, in the early stage of an epidemic, an outbreak is all the more likely to occur as the rate of secondary infections is sensitive to increases in infected individuals. Since the parametric function F_θ is a neural network, its partial derivative can be easily computed using automatic differentiation implemented in libraries such as Keras [12].

To visualize the dependence of the threshold number on its parameters, we plot in Figure 18, for many values of population size ratios and dispersion coefficient (n, κ), the ratio $\beta_c(n, \kappa)/\gamma$ where

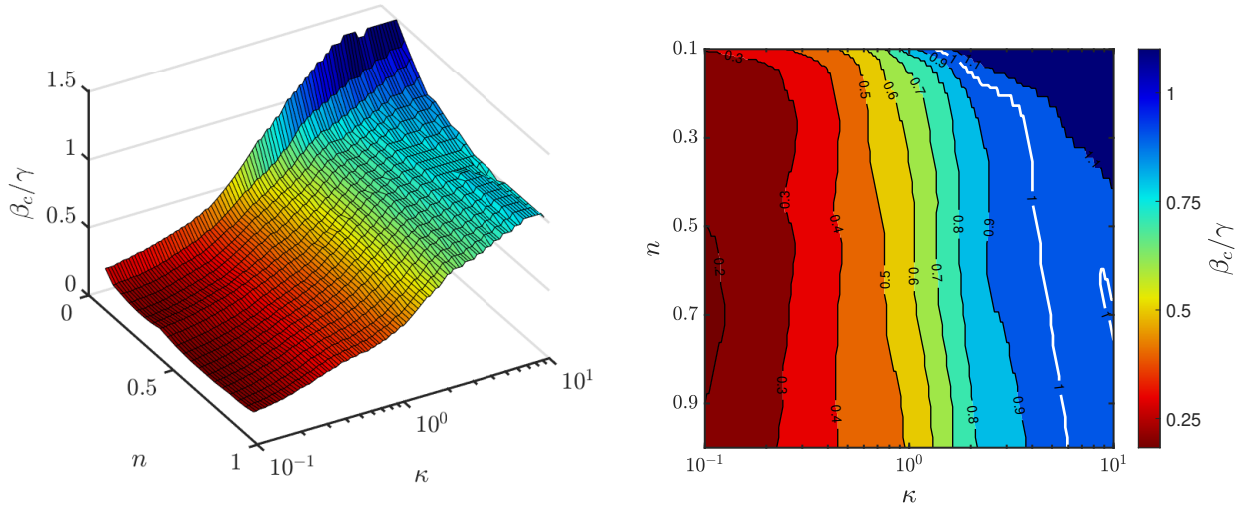


Figure 18: Left: We represent the critical value β_c above which the DFE is unstable. It is normalized by γ to obtain a dimensionless number (similar to $\mathcal{R}_0$ in the classical SIR model). The grid used contains 50 points for n, 200 for κ and 200 for β. Right: Same plot in 2D. The white contour corresponds to the case where the threshold is exactly equal to one.

$\beta_c(n, \kappa)$ denotes the critical transmission rate value above which the DFE becomes unstable. More precisely, $\beta_c = \beta_c(n, \kappa)$ is the smallest value of the transmission rate such that the following inequality holds: $\mathcal{R}_0\left(n, \beta_c(n, \kappa), \kappa\right) \geqslant 1$.

In Figure 18, we observe that the critical transmission rate β_c is an increasing function of the dispersion coefficient κ. Consequently, with small dispersion coefficients (and thus super-spreaders), low transmission rates may be more likely to lead to the development of epidemics. In other words, low-dispersion diseases have a high risk of developing into epidemics. This is the kind of tendency we would expect, as suggested in the work [28]. We also note that the critical transmission rate seems not to be significantly dependent on the population size, except when the contact distribution is almost homogeneous.

Moreover, the results show that the model captures epidemic outbreak well. For this, we plot in Figure 19 the critical transmission surface projected in 2D and compare it with the individual-based simulations. Green dots refers to IBM simulations without outbreak, while red ones refer to simulations with outbreak. If the model were perfect, the green dots would all be under the surface and the red ones above the surface. Overall, this is the kind of behaviour we observe and see that the trend described by the surface corresponds to the one found empirically.

B.2 Estimating the epidemic size.

The epidemic size, hereafter denoted R_∞, is defined as the total number of people who caught the disease. The epidemic size thus corresponds to the number of recovered individuals in large time:

$$R_\infty = \lim_{t \to +\infty} R(t). \tag{13}$$

This quantity can be seen as a function of the parameters (n, β, κ) as the dynamics of R is depending on them. In practice, the epidemic size is found by running the population-based model (RM_θ) on a sufficiently long time interval, namely T equal to 200, which is large enough so that the system

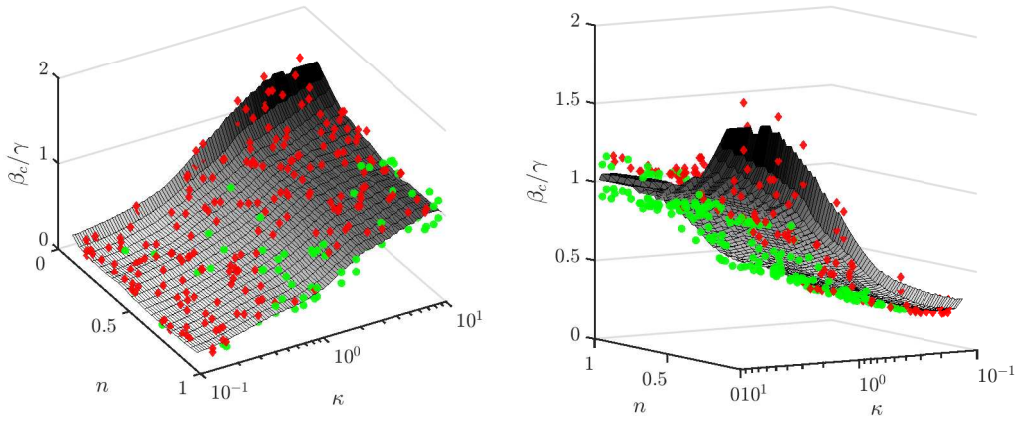


Figure 19: Green dots: stable IBM simulations. Red dots: IBM simulations with outbreak epidemic. In grey the projection of the critical transmission rate surface of Figure 18. Out of 250 simulations (dots), more than half match the theory (40 green dots should have been red while 70 red dots should have been green).

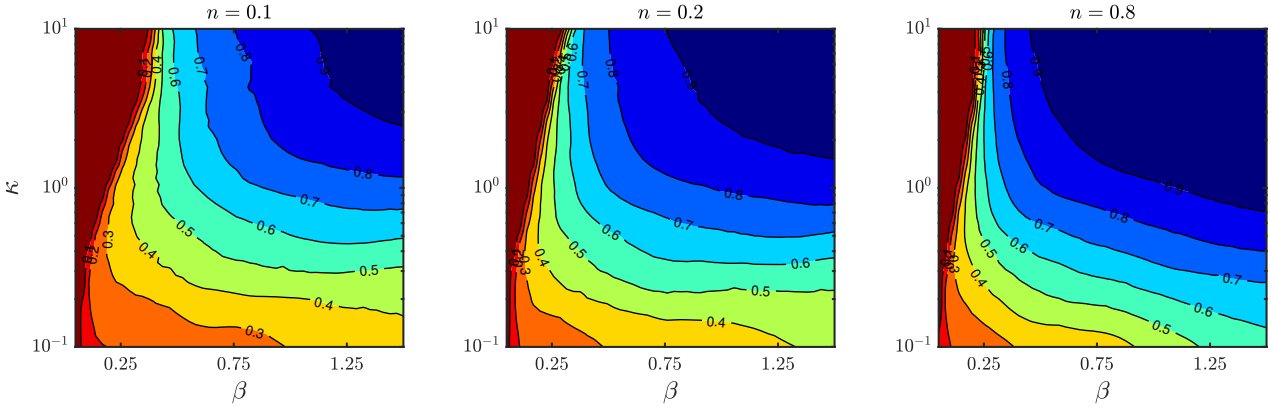


Figure 20: R_∞ as a function of (β, κ) for different population size ratios n. The grid used contains 50 points for κ, 30 points for β.

dynamics approach a stationary state. Figure 20 shows contour plots of R_∞ based on the outcome of numerical simulations with, for each population size ratio $n = 0.1, 0.2$ and 0.8, many parameters (β, κ). This suggests that:

(i) R_∞ is less dependent on the infection rate β whenever $\kappa \leq 0.5$,

(ii) dependency of R_∞ on n is more sensitive in the large dispersion coefficient case ($\kappa > 1$) and for large values of the transmission rate β,

(iii) R_∞ seems to decrease with κ; a possible interpretation of this observation is that if epidemics are more intense but shorter, the total number of infected people may be less than if the epidemic is less intense but extends over longer periods of time.

B.3 More careful derivation of $\mathcal{R}_0$

The derivation proposed in Section B.1 is incomplete because one of the hypotheses required for applying the next-generation matrix theory is not satisfied by the reduced model (RM_θ). The fifth assumption stated in [43] is lacking: in our case, it states that the ODE should have, provided no infected individuals ($I \equiv 0$), a unique asymptotically stable equilibrium point, the so-called disease-free equilibrium (DFE). However, in the model (RM_θ), there exists an infinite number of equilibrium points of the form $(S^*, 0)$ for any $S^* \in \mathbb{R}$ and none of them is asymptotically stable. Nevertheless, in order to fit to the theoretical framework, we can add demographic dynamics, through birth and death rates, leading to a stabilizing population size ratio and such that the only disease-free equilibrium point is $(S^*, I^*) = (1, 0)$. Given the uncertainty about the long term accuracy of the reduced model, the order of magnitude of the time horizons up to which the model is to be run is about 100 days. Moreover, since the demographic dynamics occur on a much larger time scale (years if not decades), this modelling assumption seems reasonable and will not strongly affect the dynamics arising from the reduced model (RM_θ).

Therefore, we insert birth and death dynamics into the model:

$$\begin{array}{rcl} S' &=& -f_\theta(S, I; n, \beta, \kappa)SI + \mu - \mu S, \\ I' &=& f_\theta(S, I; n, \beta, \kappa)SI - \gamma I - \mu I, \end{array} \tag{14}$$

where $\mu > 0$ stands for the population constant birth and death rates. First, observe that $(S^*, I^*) = (1, 0)$ is indeed the only disease-free state value making the dynamics of (14) stationary. Moreover, any solution with initial condition $(S_{\mathrm{in}}, 0)$, with $S_{\mathrm{in}} \in \mathbb{R}$ converges to the unique equilibrium point $(1, 0)$.

Then for any $\mu > 0$, the next-generation matrix theory, can be applied to model (14), leading to an expression of the threshold number, say $\mathcal{R}_0^\dagger = \mathcal{R}_0^\dagger(\mu)$. Indeed, let $\mathcal{F}$ denote the rate at which secondary infections increase in the infected compartment and $\mathcal{V}$ the sum of the rates at which the disease progresses and infected individuals die. We have that

$$I' = \mathcal{F}(S, I) - \mathcal{V}(I), \qquad \text{with} \quad \left\{ \begin{array}{rcl} \mathcal{F}(S, I) &=& f_\theta(S, I; n, \beta, \kappa)SI, \\ \mathcal{V}(I) &=& (\gamma + \mu)I. \end{array} \right.$$

Among the four other assumptions stated in [43], three of them are straightforward to verify. The last one states that the rate of secondary infections be positive or zero whenever susceptible or infected individuals remain ($\mathcal{F}(S, I) \geq 0$ for any $S, I \geq 0$). The latter was checked numerically for more than 10,000 randomly chosen different combinations of positive values. The requirement did not fail to hold and we can thus define the threshold number.

In the particular case of model (14), the next-generation matrix is actually a scalar and coincides with the threshold number:

$$\mathcal{R}_0^\dagger(\mu) = \left. \frac{\partial_I F_\theta}{\gamma + \mu} \right|_{(S=1, I=0, n, \beta, \kappa)}.$$

As $\mu \to 0$, we recover the expression of $\mathcal{R}_0$ given by Eq. (12).

Appendix C Properties of the controlled model

Well-posedness and qualitative properties

It is notable that if F_θ is assumed to be locally Lipschitz with respect to (S, I), continuous with respect to its other variables, then System (7) is well-posed according to the Carathéodory's existence theorem [7, Theorem 1.1 of Chapter 2]: it has a unique solution that belongs to $W^{1,\infty}(T_c, T; \mathbb{R}^3)$.

Since we are interested in implementing an algorithm based on gradient like iterations, and in particular at deriving first order optimality conditions, we will further assume in what follows that F_θ satisfies $(\mathcal{H}_{F_\theta})$.

The qualitative properties on (S, I) follow from the uniqueness property above and the fact that F_θ is of the particular form (6). Indeed, since $F_\theta(0, \cdot, \cdot, \cdot, \cdot) = F_\theta(\cdot, 0, \cdot, \cdot, \cdot) = 0$, the semi-axis $\{S = 0,\ I \geq 0\}$ and $\{S \geq 0,\ I = 0\}$ correspond to particular orbits of System (7). Therefore, a component of the solution (S, I) associated with positive initial data cannot vanish. Furthermore, since initial data have been chosen in such a way that $S + I + R = 1$ at every time, and since R is obviously non-negative, we infer that $\max\{S, I\} \leq 1 - R \leq 1$ at every time.

Analysis of the optimal control problem (OCP_δ)

Before stating the first order optimality conditions for Problem (OCP_δ), let us first investigate existence properties for Problem (OCP_δ).

Lemma C.1. *Let $\delta > 0$. Problem (OCP_δ) has a solution (b_δ, k_δ).*

Proof. Let $(b_p, k_p)_{p\in\mathbb{N}}$ denote a minimizing sequence for Problem (OCP_δ). Since all terms of the cost functional are non-negative, the sequence $(\mathrm{TV}(b_p) + \mathrm{TV}(k_p))_{p\in\mathbb{N}}$ is bounded. Since $(b_p)_{p\in\mathbb{N}}$ and $(k_p)_{p\in\mathbb{N}}$ are uniformly bounded in $L^\infty(T_c, T)$, it follows that $(\|(b_p, k_p)\|_{\mathrm{BV}(T_c,T)})_{p\in\mathbb{N}}$ is bounded, and we infer that $(b_p, k_p)_{p\in\mathbb{N}}$ converges up to a subsequence to some element $(b_\delta, k_\delta) \in \mathrm{BV}(T_c, T)$ in $L^1(T_c, T)$ and in particular pointwisely. In what follows, when there is no ambiguity, we will denote similarly a sequence and a subsequence with a slight abuse of notation. The pointwise convergence implies that $b_{\min} \leq b_\delta(\cdot) \leq 1$ and $1 \leq k_\delta(\cdot) \leq k_{\max}$ a.e in (T_c, T) which yields that (b_δ, k_δ) belongs to $\mathcal{U}$.

Let us denote by (S_p, I_p, R_p) the solution to (7) for the control choice $(b, k) = (b_p, k_p)$. Since $S_p + I_p + R_p$ is constant in $[T_c, T]$ and since S_p and I_p are non-negative because of the particular form of F_θ given by (6) and according to $(\mathcal{H}_{F_\theta})$, it follows that $(S_p)_{p\in\mathbb{N}}$ and $(I_p)_{p\in\mathbb{N}}$ are uniformly bounded in $L^\infty(T_c, T)$. Since f_θ is assumed to be continuous with respect to each variable, it follows from (7) that $(S_p)_{p\in\mathbb{N}}$ and $(I_p)_{p\in\mathbb{N}}$ are uniformly bounded in $W^{1,\infty}(T_c, T)$. According to the Ascoli theorem, up to a subsequence, $(S_p)_{p\in\mathbb{N}}$ and $(I_p)_{p\in\mathbb{N}}$ converge in $C^0([T_c, T])$ to some element $(S_\delta, I_\delta) \in W^{1,\infty}(T_c, T)$.

Now, let us recast (7) as

$$\left\{\begin{array}{rcl} S_p(t) &=& S_c - \int_{T_c}^t F_\theta(S_p, I_p; n, \beta_0 b_p v(k_p), \kappa_0 k_p), \\ I_p(t) &=& I_c + \int_{T_c}^t F_\theta(S_p, I_p; n, \beta_0 b_p v(k_p), \kappa_0 k_p) - \gamma I_p, \end{array}\right. \tag{15}$$

for all $t \in [T_c, T]$. Since F_θ is assumed to be (at least) continuous with respect to any of its variable, passing to the limit in these equation follows straightforwardly from the Lebesgue dominated convergence theorem. We get

$$\left\{\begin{array}{rcl} S_\delta(t) &=& S_c - \int_{T_c}^t F_\theta(S_\delta, I_\delta; n, \beta_0 b_\delta v(k_\delta), \kappa_0 k_\delta), \\ I_\delta(t) &=& I_c + \int_{T_c}^t F_\theta(S_\delta, I_\delta; n, \beta_0 b_\delta v(k_\delta), \kappa_0 k_\delta) - \gamma I_\delta, \end{array}\right. \tag{16}$$

for all $t \in [T_c, T]$, yielding that (S_δ, I_δ) satisfies (7). We conclude by noting that, according to the Lebesgue dominated convergence theorem, one has

$$\lim_{p\to+\infty} J(b_p, k_p) = J(b_\delta, k_\delta).$$

By semicontinuity of the TV seminorm for the L^1 convergence, one has

$$\mathrm{TV}[b_\delta] + \mathrm{TV}[k_\delta] \leq \liminf_{p\to+\infty} (\mathrm{TV}[b_p] + \mathrm{TV}[k_p]),$$

leading to

$$J_\delta[b_\delta, k_\delta] \leq \liminf_{p\to+\infty} J_\delta[b_p, k_p] = \inf_{(b,k)\in\mathcal{U}} J_\delta[b,k].$$

This concludes the proof. □

Computation of the differential of J and first order optimality conditions

Proof of Theorem 3.1. The first claim of the statement, related to the differentiability of S and I with respect to $[b,k]$ follows directly from the so-called Pontryagin maximum principle (see e.g. [25]). The differentiability of J is hence straightforward. Let $[b,k]\in\mathcal{U}$ and $[h_1,h_2]$ be an admissible perturbation. It remains to compute $\langle dJ[b,k],[h_1,h_2]\rangle$.

Let us introduce $\widehat{S}_1$, $\widehat{I}_1$ (resp. $\widehat{S}_2$, $\widehat{I}_2$) as the differentials of the mappings $b\mapsto S$, $b\mapsto I$ at b in the direction h_1 (resp. the differentials of $k\mapsto S$, $k\mapsto I$ at k in the direction h_2). In what follows, we will temporarily drop the variables $(S,I;n,\beta_0 bv(k),\kappa_0 k)$ in the quantities involving $F_\theta(S,I;n,\beta_0 bv(k),\kappa_0 k)$ in order to alleviate notations.

These functions solve the following ODE system:

$$\frac{d}{dt}\begin{pmatrix}\widehat{S}_1\\ \widehat{I}_1\end{pmatrix} = M_\theta\begin{pmatrix}\widehat{S}_1\\ \widehat{I}_1\end{pmatrix} + h_1\begin{pmatrix}-\beta_0 v(k)\partial_4 F_\theta\\ \beta_0 v(k)\partial_4 F_\theta\end{pmatrix} \tag{17}$$

and

$$\frac{d}{dt}\begin{pmatrix}\widehat{S}_2\\ \widehat{I}_2\end{pmatrix} = M_\theta\begin{pmatrix}\widehat{S}_2\\ \widehat{I}_2\end{pmatrix} + h_2\begin{pmatrix}-\beta_0 bv'(k)\partial_4 F_\theta - \kappa_0\partial_5 F_\theta\\ \beta_0 bv'(k)\partial_4 F_\theta + \kappa_0\partial_5 F_\theta\end{pmatrix} \tag{18}$$

completed with the initial conditions

$$\widehat{S}_i(T_c) = \widehat{I}_i(T_c) = 0, \quad i=1,2,$$

where M_θ is given by (8).

By using standard differentiation rules, one gets

$$\begin{aligned}\langle dJ[b,k],[h_1,h_2]\rangle &= \int_{T_c}^{T} \omega_\beta h_1\,(b-1) + \omega_\kappa h_2\,(k-1)\\ &\quad + \int_{T_c}^{T} \frac{\omega_{\mathrm{hosp}}}{I_{\mathrm{hosp}}}(\widehat{I}_1+\widehat{I}_2)\left(\frac{I}{I_{\mathrm{hosp}}}-1\right)_+ + \frac{1}{I_{\max}\varepsilon}(\widehat{I}_1+\widehat{I}_2)\left(\frac{I}{I_{\max}}-1\right)_+.\end{aligned}$$

Now, let us multiply System (17) by (p_1,q_1) and System (18) by (p_2,q_2) in the sense of the inner product. By integrating by parts, one gets successively

$$\begin{pmatrix}p_1\\ q_1\end{pmatrix}\cdot\begin{pmatrix}\widehat{S}_1\\ \widehat{I}_1\end{pmatrix}\Big|_{t=T} + \int_{T_c}^{T}\left(-\frac{d}{dt}\begin{pmatrix}p_1\\ q_1\end{pmatrix} - M_\theta^\top\begin{pmatrix}p_1\\ q_1\end{pmatrix}\right)\cdot\begin{pmatrix}\widehat{S}_1\\ \widehat{I}_1\end{pmatrix} = \int_{T_c}^{T} h_1\begin{pmatrix}p_1\\ q_1\end{pmatrix}\cdot\begin{pmatrix}-\beta_0 v(k)\partial_4 F_\theta\\ \beta_0 v(k)\partial_4 F_\theta\end{pmatrix}$$

and

$$\begin{pmatrix}p_2\\ q_2\end{pmatrix}\cdot\begin{pmatrix}\widehat{S}_2\\ \widehat{I}_2\end{pmatrix}\Big|_{t=T} + \int_{T_c}^{T}\left(-\frac{d}{dt}\begin{pmatrix}p_2\\ q_2\end{pmatrix} - M_\theta^\top\begin{pmatrix}p_2\\ q_2\end{pmatrix}\right)\cdot\begin{pmatrix}\widehat{S}_2\\ \widehat{I}_2\end{pmatrix} = \int_{T_c}^{T} h_2\begin{pmatrix}p_2\\ q_2\end{pmatrix}\cdot\begin{pmatrix}-\beta_0 bv'(k)v(k)\partial_4 F_\theta - \kappa_0\partial_5 F_\theta\\ \beta_0 bv'(k)v(k)\partial_4 F_\theta + \kappa_0\partial_5 F_\theta\end{pmatrix}.$$

Using that (p_1, q_1, p_2, q_2) solves the linear system (9) yields

$$\int_{T_c}^{T} \frac{\omega_{\text{hosp}}}{I_{\text{hosp}}}(\widehat{I}_1 + \widehat{I}_2)\left(\frac{I}{I_{\text{hosp}}} - 1\right)_+ + \frac{1}{I_{\max}\varepsilon}(\widehat{I}_1 + \widehat{I}_2)\left(\frac{I}{I_{\max}} - 1\right)_+ =$$
$$\int_{T_c}^{T} h_1 \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} \cdot \begin{pmatrix} -\beta_0 v(k)\partial_4 F_\theta \\ \beta_0 v(k)\partial_4 F_\theta \end{pmatrix} + \int_{T_c}^{T} h_2 \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} \cdot \begin{pmatrix} -\beta_0 b v'(k)\partial_4 F_\theta - \kappa_0 \partial_5 F_\theta \\ \beta_0 b v'(k)\partial_4 F_\theta + \kappa_0 \partial_5 F_\theta \end{pmatrix},$$

whence the desired expression of the differential.

Let us now prove the last statement of this theorem. Let $[b, k]$ denote a solution to Problem (OCP_δ). Let us introduce the so-called indicator function $\iota_{\mathcal{U}}$ to the set $\mathcal{U}$, given by

$$\iota_{\mathcal{U}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{U} \\ +\infty & \text{else.} \end{cases}$$

The functional J_δ is not differentiable in a standard sense because of the TV terms. For this reason, we will use subdifferentials to derive first order optimality conditions. We first claim that Problem (OCP_δ) is equivalent to the optimization problem

$$\inf_{(b,k)\in L^\infty(T_c,T)} J[b,k] + \delta(\text{TV}[b] + \text{TV}[k]) + \iota_{\mathcal{U}}((b,k)).$$

The standard first order optimality condition reads

$$0 \in \partial\left(J[b,k] + \delta(\text{TV}[b] + \text{TV}[k]) + \iota_{\mathcal{U}}((b,k))\right).$$

which rewrites

$$\begin{cases} -\partial_b J \in \partial\,\text{TV}(b) + \partial\iota_{[b_{\min},1]} \\ -\partial_k J \in \partial\,\text{TV}(k) + \partial\iota_{[1,k_{\max}]} \end{cases}$$

and therefore, there exist $T_b \in \partial\,\text{TV}(b)$ and $T_k \in \partial\,\text{TV}(k)$ such that the Euler inequation

$$\forall (B,K) \in L^\infty(T_c,T;[b_{\min},1]) \times L^\infty(T_c,T;[1,k_{\max}]), \quad \begin{cases} \langle \partial_b J - T_b, B - b\rangle_{L^2(T_c,T)} \geq 0 \\ \langle \partial_k J - T_k, K - k\rangle_{L^2(T_c,T)} \geq 0 \end{cases}$$

holds true and the expected conclusion follows. Note that, since it is not our main purpose, we do not provide details in this article but refer for instance to [20] for explicit characterizations of such sets. □

Appendix D Numerical implementation

Since the adjoint system is state-dependent, given an initial estimate of the control, we first determine the set of $\{(S,I)(t),\ t \in \mathcal{S}\}$ (e.g., using an explicit fourth-order Runge-Kutta numerical scheme), and then $\{(p_1,q_1,p_2,q_2)(t),\ t \in \mathcal{S}\}$.

The regularization term appearing in the problem (OCP_δ) is computed using the following approximation of the total variation (see [16, §1.30]):

$$\text{TV}[q] \simeq \sum_m \left|q(t^m) - q(t^{m-1})\right|,$$

where the family of points (t^m) belongs to $\mathcal{S}$. Note that, in the expression of the total variation, the partition used for the calculation is not arbitrary.

In addition, notice that the resulting regularisation term involves the non-differentiable absolute value function. Since this term contributes to the gradient of the cost functional J_δ, we use the following computationally efficient smooth approximation [41]: $|x| \simeq \sqrt{\eta + x^2}$, for all $x \in \mathbb{R}$, where η is taken equal to 10^{-6}.

A concise description of the optimal projected gradient approach is given in Algorithm 1 where $u \in \mathbb{R}^{2\times|\mathcal{S}|}$ denotes[6] the current (discrete) control approximation to the vector-valued solution of the problem (OCP_δ). In addition, the input u_0 is an initial guess for the control values whereas the parameters N_g and τ_g are respectively the maximum number of iterations of the algorithm and a tolerance. Lastly, $\mathcal{P}_\mathcal{U}$ refers to the map which projects the discrete components $b \in \mathbb{R}^{1\times|\mathcal{S}|}$ and $k \in \mathbb{R}^{1\times|\mathcal{S}|}$ of u according to

$$\mathcal{P}_\mathcal{U} : u = \begin{pmatrix} b_1, b_2, \ldots, b_{|\mathcal{S}|} \\ k_1, k_2, \ldots, k_{|\mathcal{S}|} \end{pmatrix} \mapsto \begin{pmatrix} P_b(b_1), \ldots, P_b(b_{|\mathcal{S}|}) \\ P_k(k_1), \ldots, P_k(k_{|\mathcal{S}|}) \end{pmatrix},$$

where $P_b : x \mapsto \min\left(1, \max\left(b_{\min}, x\right)\right)$ and $P_k : x \mapsto \min\left(k_{\max}, \max\left(1, x\right)\right)$ are defined on $\mathbb{R}$.

Algorithm 1: Optimal step size projected gradient

Require: Partition $\mathcal{S}$ of $[T_c, T]$.
Inputs: $u_0 \in \mathbb{R}^{2\times|\mathcal{S}|}$, $N_\mathrm{g} \in \mathbb{N}^*$, $\tau_\mathrm{g} > 0$.

$$\begin{aligned} p &\leftarrow 0, \\ u &\leftarrow u_0, \\ j_\mathrm{old} &\leftarrow +\infty, \\ j_\mathrm{new}, j_0 &\leftarrow J_\delta[u_0], \\ \nabla j &\leftarrow \nabla J_\delta[u_0]. \end{aligned}$$

while $p < N_\mathrm{g}$ and $(j_\mathrm{old} - j_\mathrm{new}) > \tau_\mathrm{g} j_0$ **do**

Golden-section search to find ρ^* a local minimizer of $\rho \mapsto J_\delta\left[\mathcal{P}_\mathcal{U}\left(u - \rho\nabla j\right)\right]$,
Update :

$$\begin{aligned} u &\leftarrow \mathcal{P}_\mathcal{U}\left(u - \rho^*\nabla j\right), \\ j_\mathrm{old} &\leftarrow j_\mathrm{new}, \\ j_\mathrm{new} &\leftarrow J_\delta[u], \\ \nabla j &\leftarrow \nabla J_\delta[u]. \end{aligned}$$

if $\nabla j \neq 0$ **then**

Normalise $\nabla j \leftarrow \nabla j / \left|\left|\nabla j\right|\right|$,

$p \leftarrow p + 1$.

Return: u.

[6] $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$.