Research paper

The stochastic θ-SEIHRD model: Adding randomness to the COVID-19 spread

Álvaro Leitao *, Carlos Vázquez

CITIC, Universidade da Coruña, Spain
Department of Mathematics, Universidade da Coruña, Spain

ARTICLE INFO

ABSTRACT

In this article we mainly extend a newly introduced deterministic model for the COVID-19 disease to a stochastic setting. More precisely, we incorporated randomness in some coefficients by assuming that they follow a prescribed stochastic dynamics. In this way, the model variables are now represented by stochastic process, that can be simulated by appropriately solving the system of stochastic differential equations. Thus, the model becomes more complete and flexible than the deterministic analogous, as it incorporates additional uncertainties which are present in more realistic situations. In particular, confidence intervals for the main variables and worst case scenarios can be computed.

1. Introduction

The coronavirus disease 2019, renamed as COVID-19, is an infectious disease produced by the virus SARS-CoV-2, which was declared pandemic by the World Health Organization (WHO) in March 11, 2020. This disease has posed many novel scientific challenges, ranging from contagious patterns, medical treatments or vaccine developments, to data analytics, spread modelling or evolution forecast. The research on all of these topics has been intensive in the last few months, as so will be in the near future. Thus, many disciplines from interconnected fields, like bio-sciences, mathematical and statistical modelling or artificial intelligence, will be involved and they will play an important role to rapidly overcome this exceptional emergency situation.

From the mathematical modelling point of view, epidemiological compartmental models have been often used to understand and analyse the behaviour of the contagious diseases, with the article [1] being the pioneering work. These models provide useful tools to make predictions about the future evolution of the epidemic and to control its propagation. In the literature of epidemic modelling, many examples of general-purpose models have been proposed (see [2], for a review), each of them accounting for some specific characteristics of the diseases, like the well-known examples SIR (Susceptible–Infected–Recovered), SEIR (Susceptible–Exposed–Infected–Recovered) or SIS (Susceptible–Infected–Susceptible) models.

---

* Corresponding author at: CITIC, Universidade da Coruña, Spain.
   E-mail addresses: alvaro.leitao@udc.es (Á. Leitao), carlos.vazquez.cendon@udc.es (C. Vázquez).

In addition, compartment models are building blocks of more involved approaches like the *Agent-Based* (AB) models, see [3] for example. This type of models are potentially interesting, as they are designed to include heterogeneity in the modelling, in contrast to the homogeneity that characterizes the compartmental counterparts. Such models are then able to incorporate individual attributes and the interaction among the agents themselves and with the environment, at cost of increasing the model complexity. However, AB models exhibit certain relevant disadvantages. Firstly, they are not suitable candidates at the initial stages of diseases when the lack of high quality data, specially when considering lower (micro) levels, can cause a problematic calibration of model parameters and erroneous predictions. As stated in [4], this fact can produce catastrophic consequences due to the *butterfly effect*. In addition, overparameterized AB models can result in a poor balance between complexity and interpretability, a key aspect for policy makers to decide the control measures and deliver clear messages. Furthermore, AB models are computationally intensive, while the analogous compartmental ones often turn out to be much faster alternatives. Adding extra compartments to the classical SIR-like models to accounting for different groups of interest can be understood as an intermediate strategy, aiming to preserve the advantages of both kind of models, while mitigating its own disadvantages. All in all, here we focus on an advanced compartmental model that includes enough information keeping a sufficient degree of flexibility.

Typically, the compartmental models are originally formulated following a deterministic approach, i.e., in terms of a system of *Ordinary Differential Equations* (ODEs), although, usually, they are readily extended to a stochastic version. There are two common approaches to include stochasticity into a deterministic model, relying on either the *Continuous Time Markov Chain* (CTMC) or *Stochastic Differential Equations* (SDEs), see [5] and the references therein. The stochastic models allow to capture many kinds of circumstances including all types of uncertainty that may influence the compartments dynamics: behavioural effects, public interventions, seasonal patterns, environmental factors, etc. For example, in [6,7], the authors consider a statistical inference-based approach to include random noise arising from different sources/components. Furthermore, while the solution of a deterministic model is given by a set of functions of time uniquely dependent on the initial data, the solution of the stochastic model is a set of stochastic processes, containing much more information than the deterministic analogous. In fact, at each time instant we can exploit the information provided by the probability distribution associated to the underlying random variable. A statistical analysis can be therefore performed, producing useful quantities like expected outcomes, quantiles or worst case scenarios.

In this work we present a stochastic extension of the deterministic compartmental model in [8], that has been proposed as an *ad hoc* model for the COVID-19 disease. However, note that the stochastic extension proposed here is model-independent and it can be exploited in a similar fashion under any compartmental-based approach. Unlike to the classical models (SIR, SEIR, SIS), the selected model is adapted to the specific characteristics of the COVID-19, taking into account, besides the usual factors, the undetected infectious cases, the hospitalized population or the deaths, for example. Thus, the so-called θ-SEIHRD model proposed in [8] was developed as a very general and sophisticated model (based on a previously introduced model [9]) in order to be able to study the spread of the disease worldwide. In the same work, the authors proposed a simplified version which reduces the mathematical complexity towards a more tractable model, but yet preserving the ability to capture the most relevant features. Other recent studies addressing the mathematical or statistical modelling in the context of the COVID-19 include [10–13], for example. As the literature focusing on COVID-19 modelling is increasing a lot during last months and weeks, it results quite difficult for the authors to select a list of the more relevant papers more recently arising in the topic, surely some of them being developed in parallel to this contribution.

Here, we will follow the SDE-based approach where the randomness is typically achieved by incorporating a *Brownian motion*, also known as *Wiener process*, to the ODEs of the deterministic system. There are two common ways of addressing this stochastic extension. On the one hand, an arbitrary random noise can be added to some of the equations in the ODE system, thus transforming them into SDEs. On the other hand, one (or more) of the existing model parameters can be perturbed, meaning that it (they) becomes a random variable or a stochastic process. Although both approaches have been well investigated, recent examples are [5,14,15] for the random noise approach and [16–18] for the parameter perturbation, here we prefer the second alternative. When adding random noise to an equation, it often comes with either an extra parameter to control the level of volatility or a covariance term (usually relying on the Cholesky decomposition as in [5]). Both methodologies might lack biological interpretation, specially when employing a mid-high number of compartments (SDEs), as it is the case of the model considered in this work. Furthermore, in practice, the uncertainty will have impact on a particular model component, typically represented by a model parameter. Therefore, a randomly perturbed parameter can be reasonably explained in terms of the variability produced by the source of the considered uncertainty.

In this work, we will employ a randomly perturbed disease contact rates. The approach proposed here is however conceptually different than the ones typically found in the literature, aiming to deal with some of the observed inconsistencies in the COVID-19 context. As a natural choice, many stochastic SIR-like models rely on a normally distributed perturbation for the disease contact rate parameters (see [16,18], for example). However, this approach allows negative values for the rates, which is biologically non-sensical, potentially appearing when the rate is close to zero (a pattern observed in practice for the COVID-19, see [19]). This issue is overcome in [17], by employing an exponential Ornstein–Uhlenbeck (OU) process to model the contact rate. Although this process indeed ensures positiveness, it presents another undesirable feature: an increasing variance in time. When analysing the disease transmission dynamics, there is not an objective evidence of more variability in the disease contact rates in the long-term [20]. Actually, one can expect more

control of the disease spread patterns over the course of the epidemic. For all the reasons mentioned above, we propose to model the contact disease rates by means of the so-called *CIR process*, named after its authors Cox, Ingersoll and Ross [21]. The CIR process is widely employed to simulate the evolution of interest rates in the quantitative finance framework. In some sense, the interest rates in computational finance and the disease contact rates in epidemiology present a rather similar behaviour: positiveness, controlled variability and long-term stability.

The article is organized as follows. In Section 2 we introduce the proposed mathematical model. Section 3 describes the numerical methods for solving the model and compares it with a simpler alternative. Section 4 contains the obtained results and their numerical and statistical analysis. Finally, Section 5 points out some conclusions and possible future research lines.

2. A stochastic compartmental model for the COVID-19

As mentioned in the previous introduction, we base our approach on the the so called θ-SEIHRD model first proposed in [8], which provides an advanced extension of the classical SIR-like models and adds to the usual compartments in the SEIR model the very specific ones for COVID-19: undetected infectious, hospitalized and deaths compartments. In fact, it results in a very sophisticated, well motivated and general model, with a significant number of free parameters, allowing an enormous flexibility (including several territories). In order to make this model more practically tractable, the authors also proposed a simplified version by imposing several assumptions: single territory, regular natality/mortality is neglected, and no imported/exported cases. Furthermore, some predefined forms for the open parameters of the model are adopted.

For the sake of completeness, we devote the following section to briefly describe the most relevant components of the simplified θ-SEIHRD model, before presenting its extension with randomly perturbed parameters towards the *stochastic θ-SEIHRD model* here proposed.

2.1. Original model

The original model consists of nine compartments, including classical ones like susceptible, exposed or infected (although some have a slightly different interpretation in comparison with the usual meaning) and other ones that are *COVID-related*, i.e., they account for the particular features of the disease. As it is common in this kind of models, an individual stays in a compartment a period of time and then moves to another compartment according to some transition rates. More precisely, the simplified θ-SEIHRD model in [8] is given by the following system of differential equations:

$$
\begin{aligned}
\frac{dS}{dt}(t) &= -\frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\beta_I I(t) + m_{I_u}(t)\beta_{I_u}(\theta(t))I_u(t)\right) \\
&\quad - \frac{S(t)}{N}\left(m_{H_R}(t)\beta_{H_R}(t)H_R(t) + m_{H_D}(t)\beta_{H_D}(t)H_D(t)\right), \\
\frac{dE}{dt}(t) &= \frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\beta_I I(t) + m_{I_u}(t)\beta_{I_u}(\theta(t))I_u(t)\right) \\
&\quad + \frac{S(t)}{N}\left(m_{H_R}(t)\beta_{H_R}(t)H_R(t) + m_{H_D}(t)\beta_{H_D}(t)H_D(t)\right) - \gamma_E E(t), \\
\frac{dI}{dt}(t) &= \gamma_E E(t) - \gamma_I(t)I(t), \\
\frac{dI_u}{dt}(t) &= (1-\theta(t))\gamma_I(t)I(t) - \gamma_{I_u}(t)I_u(t), \\
\frac{dH_R}{dt}(t) &= \theta(t)\left(1 - \frac{\omega(t)}{\theta(t)}\right)\gamma_I(t)I(t) - \gamma_{H_R}(t)H_R(t), \\
\frac{dH_D}{dt}(t) &= \omega(t)\gamma_I(t)I(t) - \gamma_{H_D}(t)H_D(t), \\
\frac{dR_d}{dt}(t) &= \gamma_{H_R}(t)H_R(t), \\
\frac{dR_u}{dt}(t) &= \gamma_{I_u}(t)I_u(t), \\
\frac{dD}{dt}(t) &= \gamma_{H_D}(t)H_D(t),
\end{aligned}
\tag{1}
$$

where the first two compartments, $S(t)$ and $E(t)$, denote the persons not affected by the disease pathogen (Susceptible) and the ones in incubation after being infected without clinical signs (Exposed), respectively. The compartment $I(t)$ (Infectious) includes the persons in the very preliminary stage of the infection, where nobody is expected to be detected yet, although they may infect other people after finishing the incubation period. After this period, they can either remain undetected and enter in the compartment $I_u(t)$ (Undetected Infectious), or be taken in charge by sanitary authorities

and enter in the hospitalized/quarantined group. People in compartment I_u can infect others and develop the disease, although they are not reported by sanitary authorities. Moreover, it is assumed that people in compartment I_u present no (or very weak) symptoms, so that they directly pass to recovered state after an infectious period of time. Among hospitalized persons (which include those ones in quarantine at home), we distinguish between those ones who will recover, placed in compartment $H_R(t)$, and others who will die, placed in compartment $H_D(t)$. Note that people in hospital can still infect others. Among recovered people, we distinguish between persons who were previously detected, $R_d(t)$, from those ones who were undetected, $R_u(t)$. Individuals in recovered compartments are not contagious anymore and have developed immunity. Finally, $D(t)$ denotes the persons who did not overcome the disease and died from COVID-19. As every compartment presents a dependence in time, we incorporate this time dependence by referring to each compartment's situation at instant t. More detailed explanations about this deterministic model can be found in [8].

The right hand sides of equations in system (1) contain the balances between the entries and exits of persons in each compartment. In order to properly pose the initial value problem, the system of equations is completed with the initial data $S(t_0)$, $E(t_0)$, $I(t_0)$, $I_u(t_0)$, $H_R(t_0)$, $H_D(t_0)$, $R_d(t_0)$, $R_u(t_0)$ and $D(t_0)$, representing the compartment's composition at inception time, t_0. Note that the last three equations of the system (1) are uncoupled with the six previous ones, so that the expression of their solution can be obtained and is given by

$$R_d(t) = R_d(t_0) + \int_{t_0}^{t} \gamma_{H_R}(s) H_R(s)\mathrm{d}s,$$

$$R_u(t) = R_u(t_0) + \int_{t_0}^{t} \gamma_{I_u}(s) I_u(s)\mathrm{d}s,$$

$$D(t) = D(t_0) + \int_{t_0}^{t} \gamma_{H_D}(s) H_D(s)\mathrm{d}s,$$

once the first six ones have been solved in a first stage. Thus, only the first six coupled equations need to be solved.

The model includes a number of open parameters, which are determined either from the available data/literature or by calibration. In the following, a detailed explanation of each parameter is provided.

- Efficiency of the control measures: $m_E, m_I, m_{I_u}, m_{H_R}, m_{H_D} \in [0, 1]$. Here, only one control measure is assumed (mobility restrictions, for example), implemented at date λ_1 and represented by the following decaying function:

$$m_E(t) = m_I(t) = m_{I_u}(t) = m_{H_R}(t) = m_{H_D}(t) = \begin{cases} 1, & \text{if } t \in [0, \lambda_1], \\ \exp\left(-\kappa_1(t - \lambda_1)\right), & \text{if } t \in [\lambda_1, T], \end{cases}$$

  with the parameter $\kappa_1 \in [0, 0.2]$ entering in the calibration procedure. The generalization to more/individualized control measures is straightforward.

- The fatality rate: $\omega(t) \in [\underline{\omega}, \overline{\omega}] \subset [0, 1]$. The following form is proposed:

$$\omega(t) = m_I(t)\overline{\omega} + (1 - m_I(t))\underline{\omega},$$

  with $\underline{\omega}$ and $\overline{\omega}$ being the fatality rate limits with and without control measures, respectively. As $\overline{\omega} \geq \underline{\omega}$, the value of $\overline{\omega}$ is defined in terms of its difference w.r.t the lower limit, i.e., $\overline{\omega} = \underline{\omega} + \delta_\omega$, whose values are calibrated.

- The fraction of infected individuals which are detected and documented by the authorities: θ. We assume that all deaths are detected, so that $\theta \in [\overline{\omega}, 1]$, and that it changes in time. More precisely, we consider the expression

$$\theta(t) = \begin{cases} \underline{\theta}, & \text{if } t \in [t, \lambda_1], \\ \text{linear continuous}, & \text{if } t \in [\lambda_1, \lambda_2], \\ \overline{\theta}, & \text{if } t \in [\lambda_2, T], \end{cases}$$

  with $\underline{\theta}, \overline{\theta}, \lambda_1$, and λ_2 inferred from the data.

- The compartment transition rates: $\gamma_E, \gamma_I, \gamma_{I_u}, \gamma_{H_R}, \gamma_{H_D} \in (0, +\infty)$. These parameters are constructed in accordance to the recent literature. They are based on the average duration (in days) of an individual in each infectious compartment, denoted by $d_E, d_I, d_{I_u}, d_{H_R}$ and d_{H_D}. Moreover, we assume that $d_{I_u} = d_{H_R}$ and $d_{H_D} = d_{H_R} + \delta_R$, $\delta_R > 0$. Thus,

$$\gamma_I = \frac{1}{d_E}, \quad \gamma_I(t) = \frac{1}{d_I - g(t)}, \quad \gamma_{I_u}(t) = \gamma_{H_R}(t) = \frac{1}{d_{I_u} + g(t)}, \quad \gamma_{H_D}(t) = \frac{1}{d_{I_u} + g(t) + \delta_R},$$

  where $g(t) = d_g(1 - m_I(t))$ represents the decrease of the duration of d_I due to the application of control measures at time t, with d_g being the maximum number of days that d_I can be decreased due to the control measures. The parameter δ_R is included in the calibration.

- The disease contact rates: $\beta_E, \beta_I, \beta_{I_u}, \beta_{H_R}, \beta_{H_D} \in \mathbb{R}^+$. The parameter β_I is assumed to be given, after calibration to the available data. Furthermore, the following relation between β_I and the rest of the contact rates is considered:

$$\beta_E = C_E \beta_I, \quad \beta_{I_u}(t) = \underline{\beta}_I + \frac{\beta_I - \underline{\beta}_I}{1 - \omega(t)}(1 - \theta(t)), \quad \beta_{H_R} = \beta_{H_D} = C_H(t)\beta_I, \tag{2}$$

where $\underline{\beta}_I = C_u\beta_I$, with C_E, $C_H(t)$ and $C_u \in [0, 1]$. Parameters C_E and C_u are also obtained by calibration, while $C_H(t)$ is determined by the following expression:

$$C_H(t) = \frac{\alpha_H \left(\frac{\beta_I}{\gamma_I(t)} + \frac{\beta_E}{\gamma_E(t)} + (1 - \theta(t))\frac{\beta_{I_u}(t)}{\gamma_{I_u}(t)}\right)}{(1 - \alpha_H)\beta_I\theta(t)\left(\left(1 - \frac{\omega(t)}{\theta(t)}\right)\frac{1}{\gamma_{H_R}(t)} + \frac{\omega(t)}{\theta(t)}\frac{1}{\gamma_{H_R}(t)}\right)}, \tag{3}$$

with α_H being the percentage of cases (healthcare workers) infected by individuals in compartments H_R or H_D. Expression (3) results from the use of the available data for infection transmissions within hospitals, which is a trustworthy and more easily gathered information, specially at the beginning of the pandemic. The term incorporates a relation with the other disease contact and transition rates (including a temporal dependency).

2.2. Stochastic extension

Our aim is to introduce stochasticity in the simplified θ-SEIHRD model previously described. Moreover, in order to keep the interpretability of the proposed stochastic model, we will add some randomness to a set of parameters. More precisely, we will add randomness on the disease contact rates, β's. Note that in the simplified deterministic version proposed in [8] all the β's depend on β_I, as indicated in Eq. (2). Therefore, we propose to add uncertainty in β_I by turning it into a stochastic process, i.e., a random variable at each time instant.

Then, we start by writing the model in terms of β_I. From Eq. (2), we have

$$\beta_E = \beta_I A_E, \quad \beta_{I_u} = \beta_I A_{I_u}, \quad \beta_{H_R} = \beta_I A_{H_R}, \quad \beta_{H_D} = \beta_I A_{H_D},$$

where

$$A_E = C_E,$$

$$A_{I_u}(t) = C_u + \frac{(1 - C_u)(1 - \theta(t))}{1 - \omega(t)},$$

$$A_{H_R}(t) = A_{H_D}(t) = \frac{\alpha_H \left(\frac{1}{\gamma_I(t)} + \frac{A_E}{\gamma_E} + (1 - \theta(t))\frac{A_{I_u}(t)}{\gamma_{I_u}(t)}\right)}{(1 - \alpha_H)\theta(t)\left((1 - \frac{\omega(t)}{\theta(t)})\frac{1}{\gamma_{H_R}(t)} + \frac{\omega(t)}{\theta(t)}\frac{1}{\gamma_{H_D}(t)}\right)}.$$

By using the previous notation, the simplified θ-SEIHRD model (1) can be rewritten as

$$\frac{dS}{dt}(t) = -\beta_I\frac{S(t)M(t)}{N},$$

$$\frac{dE}{dt}(t) = \beta_I\frac{S(t)M(t)}{N} - \gamma_E E(t),$$

$$\frac{dI}{dt}(t) = \gamma_E E(t) - \gamma_I I(t),$$

$$\frac{dI_u}{dt}(t) = (1 - \theta(t))\gamma_I I(t) - \gamma_{I_u}I_u(t),$$

$$\frac{dH_R}{dt}(t) = \theta(t)\left(1 - \frac{\omega(t)}{\theta(t)}\right)\gamma_I I(t) - \gamma_{H_R}H_R(t), \tag{4}$$

$$\frac{dH_D}{dt}(t) = \omega(t)\gamma_I I(t) - \gamma_{H_D}H_D(t),$$

$$\frac{dR_d}{dt}(t) = \gamma_{H_R}(t)H_R(t),$$

$$\frac{dR_u}{dt}(t) = \gamma_{I_u}(t)I_u(t),$$

$$\frac{dD}{dt}(t) = \gamma_{H_D}(t)H_D(t),$$

where

$$M(t) = m_E A_E E(t) + m_I I(t) + m_{I_u}A_{I_u}I_u(t) + m_{H_R}A_{H_R}H_R(t) + m_{H_D}A_{H_D}H_D(t). \tag{5}$$

As previously indicated, we incorporate a stochastic component into the model by replacing the parameter β_I of the deterministic model with a random variable. Instead of considering a constant value for β_I, the disease contact rate in compartment I follows a newly introduced stochastic process $\tilde{\beta}_I(t)$. In order to preserve the positiveness in the parameter definition (imposed by its biological interpretation), we choose the well-known CIR process [21]. The main advantage of the CIR process is that it theoretically ensures the spacial states to be non-negative. Moreover, as it is a mean-reverting process, the dynamics of the CIR process tends to a prescribed value in the long term. This property can have the following

biological interpretation: at the early stages of the disease, the contacts between people are less controlled (so more volatile), while in the mid-long term the individual contacts and the disease spread patterns are more studied and the ranges of variability are more reduced. Accordingly, the dynamics of $\tilde{\beta}_I$ satisfies the following stochastic differential equation:

$$d\tilde{\beta}_I(t) = \nu_{\beta_I}(\mu_{\beta_I} - \tilde{\beta}_I(t))dt + \sigma_{\beta_I}\sqrt{\tilde{\beta}(t)}dW(t), \tag{6}$$

where ν_{β_I} is the mean reverting speed, μ_{β_I} is the long-term average, σ_{β_I} is the volatility, and $dW(t)$ is a Brownian motion increment. Therefore, the stochastic θ-SEIHRD model is governed by the following system of Random Ordinary Differential Equations (RODEs):

$$
\begin{aligned}
dS(t) &= \tilde{\beta}_I(t)\frac{S(t)M(t)}{N}dt, \\
dE(t) &= \left(\tilde{\beta}_I(t)\frac{S(t)M(t)}{N} - \gamma_E E(t)\right)dt, \\
dI(t) &= (\gamma_E E(t) - \gamma_I I(t))\,dt, \\
dI_u(t) &= \left((1 - \theta(t))\gamma_I I(t) - \gamma_{I_u} I_u(t)\right)dt, \\
dH_R(t) &= \left(\theta(t)\left(1 - \frac{\omega(t)}{\theta(t)}\right)\gamma_I I(t) - \gamma_{H_R} H_R(t)\right)dt, \\
dH_D(t) &= \left(\omega(t)\gamma_I I(t) - \gamma_{H_D} H_D(t)\right)dt, \\
dR_d(t) &= \gamma_{H_R}(t)H_R(t)dt, \\
dR_u(t) &= \gamma_{I_u}(t)I_u(t)dt, \\
dD(t) &= \gamma_{H_D}(t)H_D(t)dt, \\
d\tilde{\beta}_I(t) &= \nu_{\beta_I}(\mu_{\beta_I} - \tilde{\beta}_I(t))dt + \sigma_{\beta_I}\sqrt{\tilde{\beta}_I(t)}dW(t),
\end{aligned}
\tag{7}
$$

with $M(t)$ as defined in Eq. (5). Although only one source of randomness is introduced, the solution of the whole system becomes a set of stochastic processes, due to the dependence of the remaining equations on the first equations for S and E, and their own dependence on $\tilde{\beta}_I$. Note also that we could add time dependency to $\tilde{\beta}_I$ both in the deterministic model (which is not the case in [8]) as well as in the stochastic version (by considering either ν_{β_I}, μ_{β_I} or σ_{β_I} time dependent).

Remark. Note that our approach could be generalized to a setting where some of the parameters were independent, in this case we would consider each one as a Gaussian random variable with possible correlations between (some or) all of them. In this more general setting, a certain number of different (possibly correlated) Brownian motion processes would come into place.

3. On the stochastic θ-SEIHRD model

In this section, we provide a detailed analysis of the stochastic θ-SEIHRD model defined in (7).

3.1. Existence and uniqueness of solution

First, note that (7) is a system of RODEs driven by an Itô process, which defines the CIR model. Since the paths of the Itô process are at most Hölder continuous, the solution of the system of RODEs is at most continuously differentiable independently of the smoothness of the functions that define the right hand side of the system of ODEs [14]. Therefore, the usual arguments based on Taylor expansions to obtain the order of convergence for the classical numerical methods does not apply. For example, classical Runge–Kutta methods do not achieve their order of convergence in ODEs when applied to RODEs (see [22] and the references therein).

Secondly, in order to establish the existence of solution for the system of RODEs (7) driven by the CIR process, we introduce the notation

$$X(t) = (S(t),\ E(t),\ I(t),\ I_u(t),\ H_R(t),\ H_D(t),\ R_d(t),\ R_u(t),\ D(t))^T,$$

where the super-index $(\cdot)^T$ denotes the transpose operator. Moreover, we introduce the following notation for the coefficients of the CIR process,

$$a(\tilde{\beta}_I) = \nu_{\beta_I}(\mu_{\beta_I} - \tilde{\beta}_I), \quad b(\tilde{\beta}_I) = \sigma_{\beta_I}\sqrt{\tilde{\beta}_I}.$$

By using the previous notation, the system (7) can be equivalently written as

$$
\begin{aligned}
dX(t) &= F(t, X(t))dt, \\
d\tilde{\beta}_I(t) &= a(\tilde{\beta}_I(t))dt + b(\tilde{\beta}_I(t))dW(t),
\end{aligned}
\tag{8}
$$

where F is the function that defines the right hand side of the system of RODEs. Note that (8) can be understood as a system of SDEs where the diffusion coefficients for all equations are equal to zero except for the last equation which is equal to $b(\tilde{\beta}_I(t))$. For a set of constant initial data $S(0)$, $E(0)$, $I(0)$, $I_u(0)$, $H_R(0)$, $H_D(0)$, $R_d(0)$, $R_u(0)$, $D(0)$, and $\tilde{\beta}_I(0)$, the system (8) has a unique strong solution. This follows from the fact that the coefficients of the system (8) of SDEs are locally Lipschitz continuous and the initial condition is a constant value (see [23], for example). Therefore, the existence and uniqueness of solution for the θ-SEIHRD model defined by (7) follows.

3.2. Numerical solution

As the system (7) is nonlinear, it is not possible to obtain a closed-form expression for the solution, as it also happens with the corresponding deterministic version. Therefore, the use of numerical methods for solving (7) becomes mandatory. Here, we adopt the following strategy. First, we perform a simulation of the dynamics of $\tilde{\beta}_I(t)$, in accordance with the CIR process. After, we solve the resulting ODE system for each simulated path of $\tilde{\beta}_I(t)$. In this way, we obtain a set of random walks for each stochastic process representing a model variable.

Remark. Although in the stochastic θ-SEIHRD model (7) the solution is a finite set of stochastic processes, we keep the same notation we used for the finite set of real valued functions representing the solution of the deterministic θ-SEIHRD model.

The CIR process is a well-studied dynamics often employed in computational finance (see [24] for example), which satisfies the SDE in Eq. (6). From the mathematical point of view, one of its relevant features is that the underlying distribution is known analytically, relying on the *non-central chi-squared* distribution. Thus, for $s < t$, the conditional distribution of $\tilde{\beta}_I$ defined in Eq. (6) reads

$$\tilde{\beta}_I(t)|\tilde{\beta}_I(s) \sim c(t,s) \cdot \chi^2 \left(d, \frac{e^{-\nu_{\beta_I}(t-s)}}{c(t,s)} \tilde{\beta}_I(s) \right),$$

where

$$c(t,s) = \frac{\sigma_{\beta_I}^2}{4\mu_{\beta_I}} \left(1 - e^{-\nu_{\beta_I}(t-s)} \right), \quad d = \frac{4\nu_{\beta_I}\mu_{\beta_I}}{\sigma_{\beta_I}^2},$$

and $\chi^2(a,b)$ is the non-central chi-squared distribution with a degrees of freedom and non-centrality parameter b. This forms the basis for an *exact simulation* scheme, which can be used to obtain realizations of $\tilde{\beta}_I$. Given a set of $m+1$ time points, $\{t_i\}_0^m$, where the solution will be computed, we have

$$c(t_{i+1}, t_i) = \frac{\sigma_{\beta_I}^2}{4\mu_{\beta_I}} \left(1 - e^{-\nu_{\beta_I}(t_{i+1}-t_i)} \right),$$

$$\tilde{\beta}_I(t_{i+1}) = c(t_{i+1}, t_i)\chi^2 \left(d, \frac{e^{-\nu_{\beta_I}(t_{i+1}-t_i)}}{c(t_{i+1}-t_i)} \tilde{\beta}_I(t_i) \right),$$

with the constant parameter $d = \frac{4\nu_{\beta_I}\mu_{\beta_I}}{\sigma_{\beta_I}^2}$ and some initial value $\tilde{\beta}_I(t_0) = \tilde{\beta}_I(0)$. By employing this scheme, we can simulate n discrete sample paths of $\tilde{\beta}_I$.

Once the sample paths of $\tilde{\beta}_I$ have been obtained, we numerically solve n ODE systems, one for each of these paths. For this purpose, we choose the explicit Runge–Kutta method of order 5(4) when applied to deterministic ODE systems, known as *RK45*, *RKDP* or Dormand–Prince method, see [25]. Its practical implementation requires that the time discretization mesh includes the time points used in the simulation of $\tilde{\beta}_I$. As previously indicated, the lack of enough smoothness of the sample paths of the CIR process reduce the regularity of the solution and also the order of convergence of Runge–Kutta method obtained when solving deterministic ODE systems.

Remark. Alternative numerical methods for solving (7) could be based on the use of classical numerical stochastic methods for solving the equivalent system of SDEs (8). For example, applying the Euler–Maruyama method to (8) would be equivalent to use the explicit Euler scheme to solve the eight ODEs with random coefficients combined with Euler–Maruyama scheme to approximate the paths of the CIR process. Also note that using a stochastic Runge–Kutta scheme for the system of SDEs (8) would be equivalent to use a Runge–Kutta scheme for the ODEs in (7) combined with a stochastic Runge–Kutta method to approximate the paths of the CIR process. Both previous alternatives, by Euler–Maruyama and Runge–Kutta stochastic numerical methods for SODEs, have been numerical analysed in the classical book [23]. By using the equivalence between a RODE driven by an Itô process and the corresponding Stochastic ODE, numerical methods for this kind of RODEs are analysed in [22].

In our approach, instead of using numerical methods to approximate the sample paths of the CIR process, we employ what is known as an exact simulation scheme since we exactly sample the distribution to generate each sample path.

This exact simulation is combined with a Runge–Kutta method to integrate the resulting ODE for each path. In this work we do not address the numerical analysis of the proposed numerical strategy, but only use it to simulate the proposed stochastic model.

In order to illustrate the potential of the stochastic version of θ-SEIHRD model, in Figs. 1, 2 and 3 we present the solution of the deterministic model versus a number of possible *scenarios* (simulations) obtained by the stochastic model. Further, we include the so-called "Average scenario", which is nothing else than an estimation of the expected value of the model variables in time. We can clearly observe that the outcomes produced by the stochastic model provide much more information about the future evolution of each model compartment, while the deterministic one seems to be somehow an *averaged* version of the stochastic formulation. This gives just an initial insight of the potential of the newly introduced model, which will be completed in Section 4 with a more detailed analysis.

3.3. Comparison with a stochastic SEIR model

In this section, we argue the reasons why we choose to extend the θ-SEIHRD model to a stochastic setting, instead of any other simpler alternative. For this purpose, let us consider the well known SEIR model. The θ-SEIHRD can be easily transformed into a SEIR-like model by smartly selecting the values of some of the model parameters. By doing so, we can still use the rest of the calibrated parameters as SEIR model inputs. Thus, if we take $m_{I_u}(t) = m_{H_R}(t) = m_{H_D}(t) = 0$ and $\theta = 1$ in Eq. (1), and introduce the new compartment R (removed), we obtain

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t}(t) &= -\frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\hat{\beta}_I I(t)\right), \\
\frac{\mathrm{d}E}{\mathrm{d}t}(t) &= \frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\hat{\beta}_I I(t)\right), \\
\frac{\mathrm{d}I}{\mathrm{d}t}(t) &= \gamma_E E(t) - \gamma_I(t) I(t), \\
\frac{\mathrm{d}R}{\mathrm{d}t}(t) &= \gamma_I(t) I(t),
\end{aligned}
\tag{9}
$$

where all the model parameters have the same interpretation as in Section 2, while we incorporate a specific disease contact rate, $\hat{\beta}_I$, for compartment I. Since this newly introduced SEIR-like model does not distinguish between detected and undetected (actually it considers all the cases detected, i.e., $\theta = 1$), the disease contact rate of compartment I needs to be compensated in order to take into account this fact and allow to compare the results provided by both models. Therefore, we adopt the following natural formulation, in which we assume that the new disease contact rate of compartment I is the addition of the two rates of the original compartments for detected and undetected infections, i.e.,

$$
\beta_I^* = \beta_I + \beta_{I_u} = \beta_I + \underline{\beta}_I,
$$

where the second equality comes from Eq. (2) and $\theta = 1$. The advantage of this approach is that it enables the use of the previously defined parameters, which are already calibrated to the data.

Next, we compare the outcomes produced by the models defined by Eqs. (1) and (9). In Fig. 4a, we observe that the curve of infected cases resulting from the SEIR model is very similar to the one obtained with the θ-SEIHRD model (see Fig. 1e). Therefore, we can conclude that both models provide equivalent outcomes in terms of the infected cases.

Let us now consider a *stochastic SEIR-like model* that results from simplifying the stochastic θ-SEIHRD of (7) in a similar way as for the deterministic case, i.e., taking $m_{I_u}(t) = m_{H_R}(t) = m_{H_D}(t) = 0$ and $\theta = 1$, incorporating the "removed cases" compartment, R. Then, we have the dynamics

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t}(t) &= -\frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\tilde{\beta}_I^* I(t)\right), \\
\frac{\mathrm{d}E}{\mathrm{d}t}(t) &= \frac{S(t)}{N}\left(m_E(t)\beta_E E(t) + m_I(t)\tilde{\beta}_I^* I(t)\right), \\
\frac{\mathrm{d}I}{\mathrm{d}t}(t) &= \gamma_E E(t) - \gamma_I(t) I(t), \\
\frac{\mathrm{d}R}{\mathrm{d}t}(t) &= \gamma_I(t) I(t), \\
\mathrm{d}\tilde{\beta}_I^*(t) &= \nu_{\beta_I}(\mu_{\beta_I} - \tilde{\beta}_I^*(t))\mathrm{d}t + \sigma_{\beta_I}\sqrt{\tilde{\beta}_I^*(t)}\mathrm{d}W(t),
\end{aligned}
\tag{10}
$$

where a new stochastic process, $\tilde{\beta}_I^*(t)$, following a CIR dynamics is considered. By using the stochastic SEIR model defined by Eq. (10), we can now generate a number of scenarios and compute an estimation of the expected infected cases, $\mathbb{E}[I(t)]$, as we have done in the previous subsection. In Fig. 4b, several of those Monte Carlo scenarios and the expected value of $I(t)$ (labelled as "Average scenario") are depicted. Again, we can easily conclude that this stochastic extension of the SEIR model reproduces the observed behaviour for the θ-SEIHRD model (see Figs. 1, 2 and 3), since the deterministic version can be seen as an averaging scenario of the stochastic generalization.
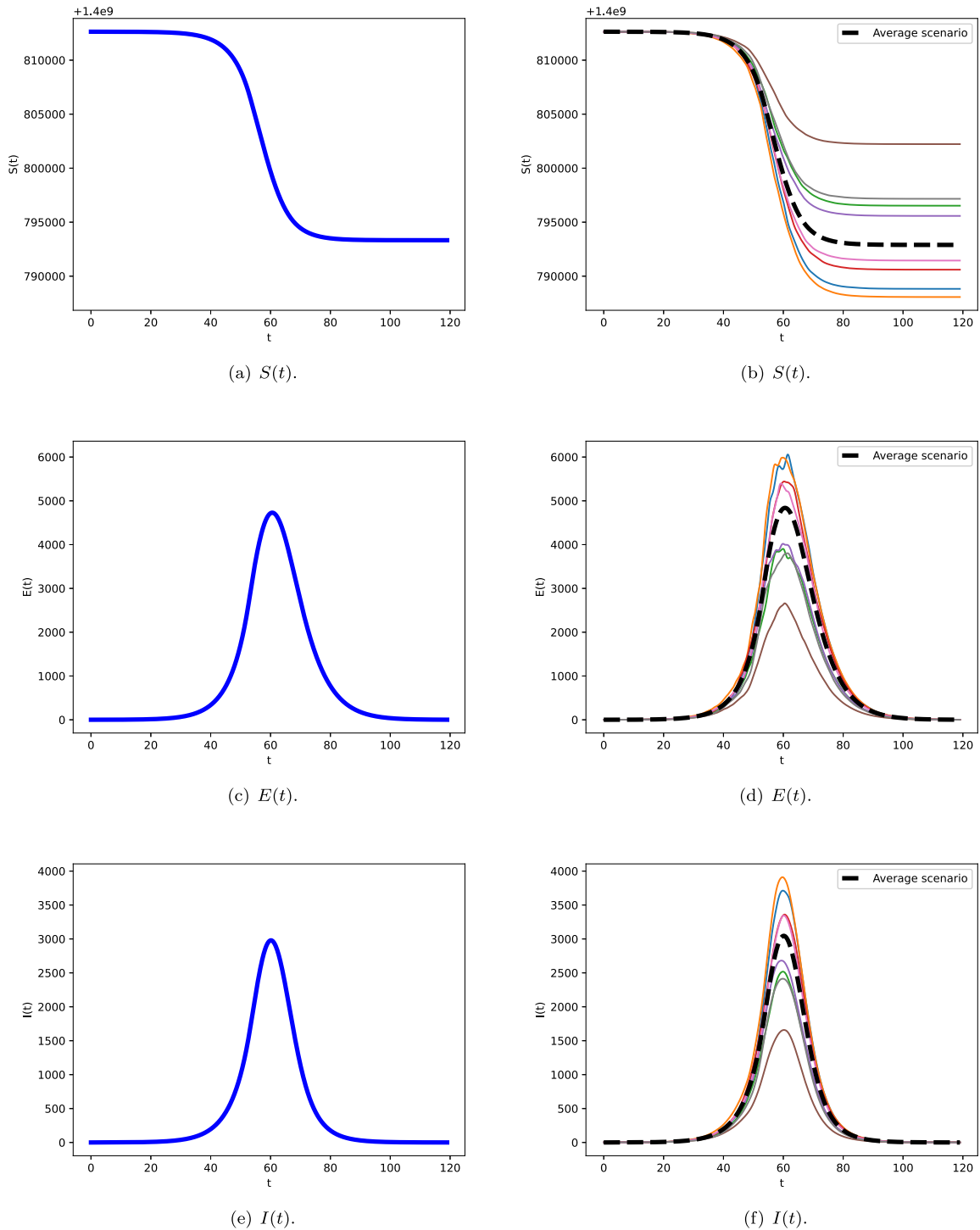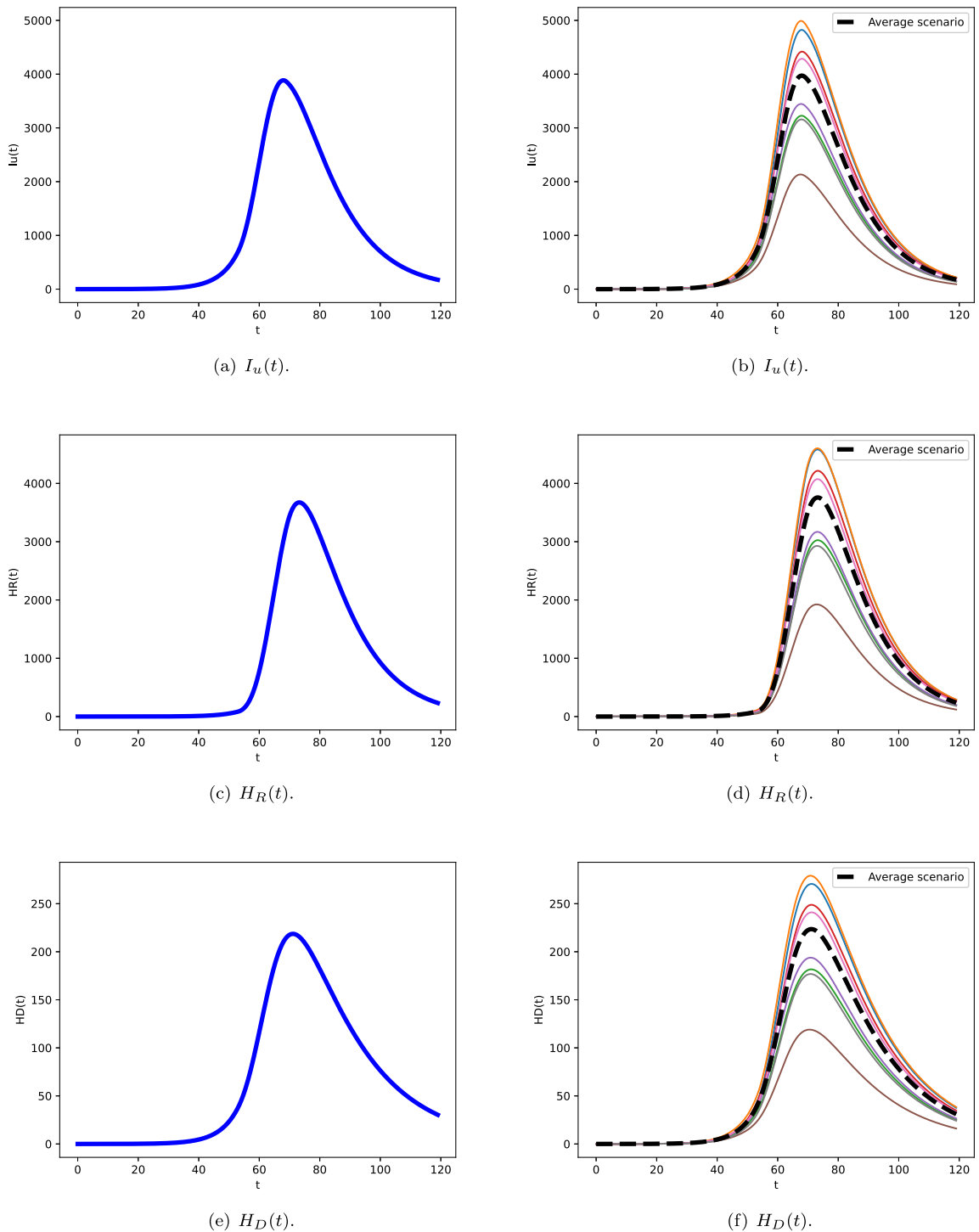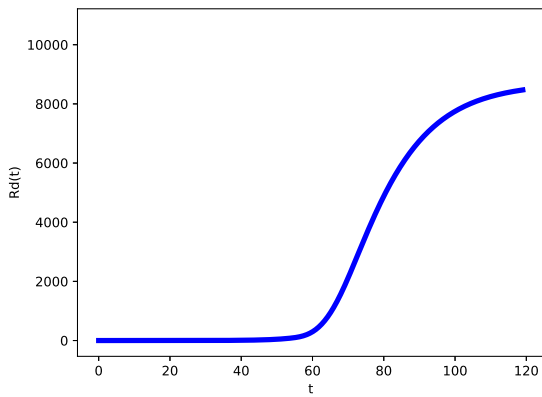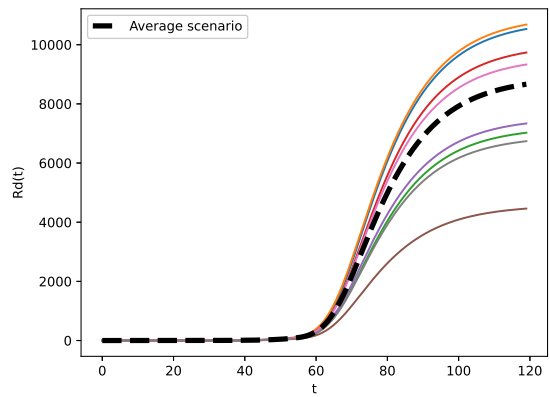
Fig. 1. Deterministic vs. Stochastic: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $\sigma_{\beta_I} = 0.1$, with $n = 2^{15}$ Monte Carlo simulations (only 8 simulations depicted).

Next, the total number of active cases predicted by both models, θ-SEIHRD and SEIR, and their corresponding stochastic versions are considered. The obtained results are presented in Fig. 5. We can readily observe that the SEIR model tends to significantly underestimate the number of active cases, an issue that is highly undesirable. This effect is even more pronounced when the stochastic models are employed, thus obtaining a gap of around 3000 cases in average (around 8000 cases in the worst case scenario).

(a) $I_u(t)$.

(b) $I_u(t)$.

(c) $H_R(t)$.

(d) $H_R(t)$.

(e) $H_D(t)$.

(f) $H_D(t)$.

Fig. 2. Deterministic vs. Stochastic: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $\sigma_{\beta_I} = 0.1$, with $n = 2^{15}$ Monte Carlo simulations (only 8 simulations depicted).

The undesired behaviour produced by the simplified SEIR alternative can be very problematic. Although this simpler model is able to capture the infected cases (see Fig. 4), an erratic estimation of the total active cases is observed. These poor predictions of the disease evolution generated with the SEIR model are caused by the concentration of the cases in a single compartment and the absence of intermediate ones (hospitalized). In addition, the impact of this effect on the stochastic model outcomes is significantly more important, which makes the introduced SEIR model an inappropriate candidate to be stochastically extended in the context of the COVID-19 disease.
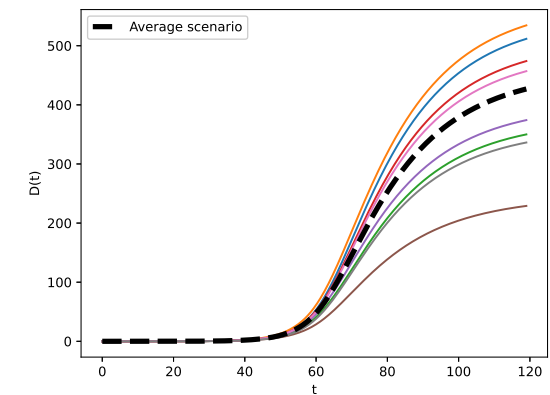
(a) $R_d(t)$.

(b) $R_d(t)$.
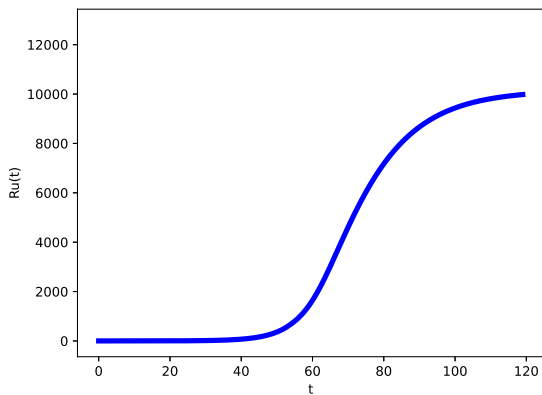
(c) $R_u(t)$.
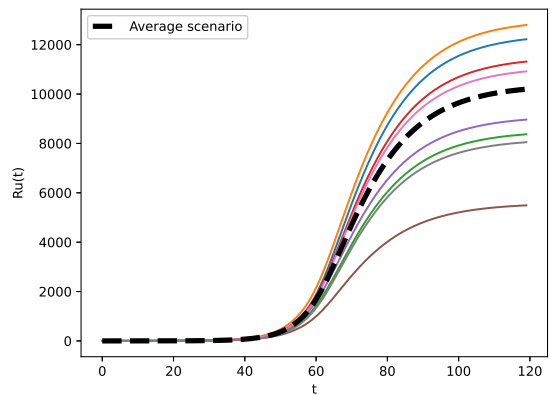
(d) $R_u(t)$.
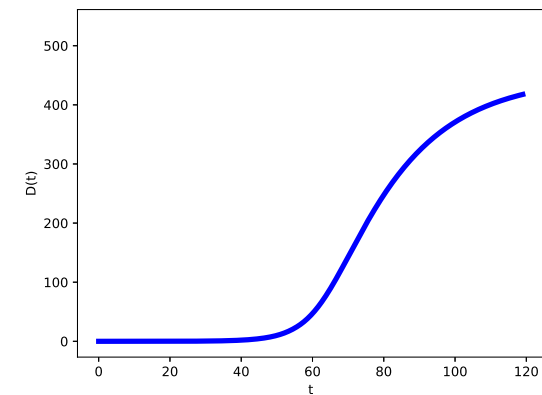
(e) $D(t)$.

(f) $D(t)$.

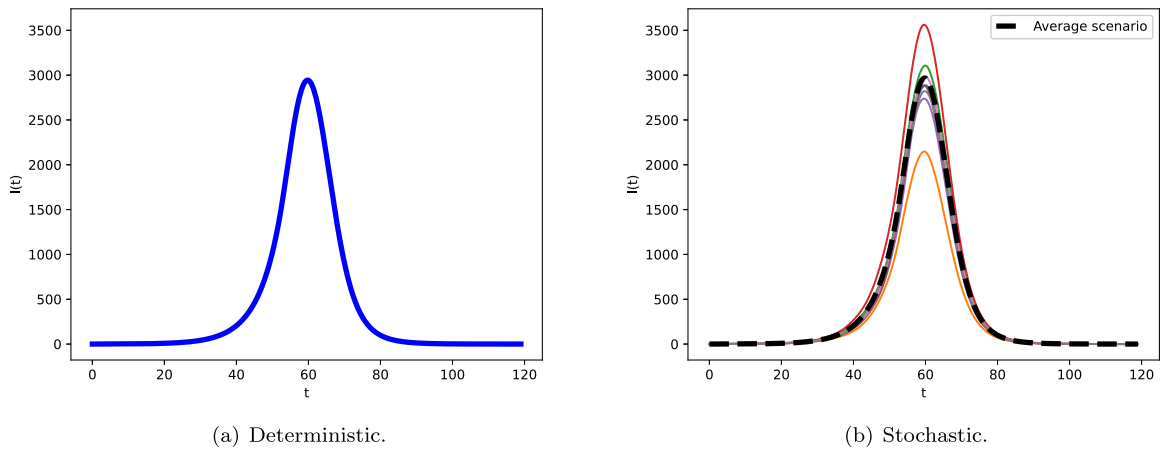**Fig. 3.** Deterministic vs. Stochastic: $v_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $\sigma_{\beta_I} = 0.1$, with $n = 2^{15}$ Monte Carlo simulations (only 8 simulations depicted).

Remark. In order to employ the stochastic SEIR alternative presented above to model the COVID-19 disease evolution, it would require ad hoc calibration for its free parameters, both the model related and the CIR process related ones. However, our goal here is to compare the extension of an existing and already calibrated deterministic model to its stochastic counterpart.

(a) Deterministic.   (b) Stochastic.

Fig. 4. SEIR model: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I^*$ and $\sigma_{\beta_I} = 0.1$, with $n = 2^{15}$ Monte Carlo simulations (only 8 simulations depicted).

(a) Deterministic.   (b) Stochastic.
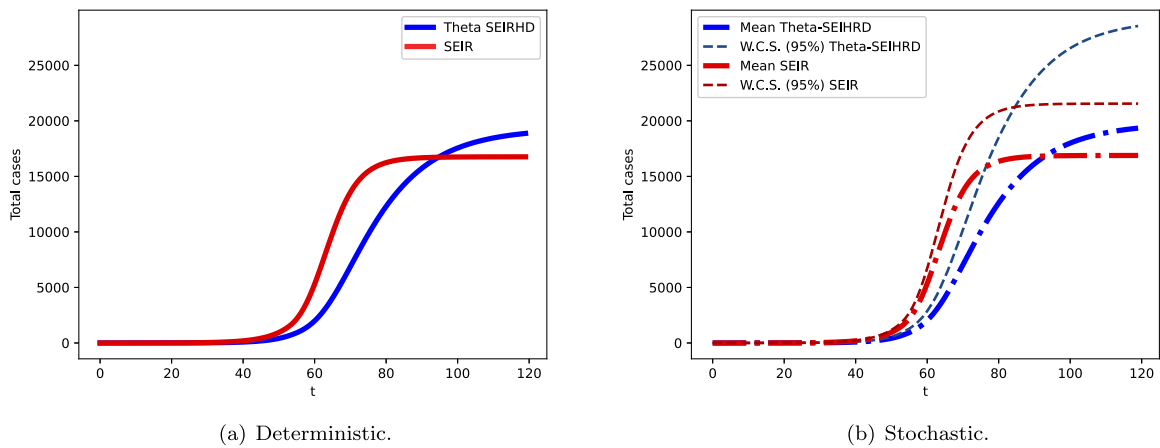
Fig. 5. θ-SEIHRD vs. SEIR. Total active cases. CIR parameter: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I^*$ and $\sigma_{\beta_I} = 0.1$.

4. Numerical and statistical analysis

The experiments have been conducted in a computer system with the following characteristics: CPU Intel Core i7-4720HQ 2.6 GHz, 16 GB RAM memory and GPU GeForce GTX 970M. The numerical codes have been implemented in Python programming language. We consider a uniform time grid, i.e., $\Delta t := t_{i+1} - t_i$, $\forall i$, with time step $\Delta t = \frac{1}{6}$ (around 4 h).

In this section, we perform a numerical and statistical study of the proposed stochastic model. As mentioned, the solution of the system of SDEs in (7) is a set of stochastic processes, meaning that we can "extract" a random variable at each time point. In this way, not only a single value (like for the deterministic case), but also some statistics can be provided for a prescribed time. In this work, we consider the mean, the interquantile interval, $[Q_1, Q_3]$ (with Q_1 and Q_3 being the first and the third quantiles, respectively), and a worst case scenario (WS), applied to both the evolution of the model variables and the possible model outputs.

Here we take advantage of the experiments conducted in [8], referred to the case of China. In Tables 1 and 2 the reported values for the open coefficients are presented, distinguishing between the ones extracted from the literature or by experience, and the ones obtained by a calibration procedure, respectively. The initial data is given by $S(t_0) = N - 1$, $E(t_0) = 1$ and $I(t_0) = I_u(t_0) = H_R(t_0) = H_D(t_0) = R_d(t_0) = R_u(t_0) = D(t_0) = 0$.

4.1. Model variables

We firstly test the evolution of the model variables. Thus, we extract some simulation-based statistics at a couple of time instants, the 8th February (inflection point) and the 29th March (final point). Furthermore, the impact of different levels of uncertainty is also reported by considering several representative values for σ_{β_I}, thus reflecting situations of no

Table 1

Parameters extracted from the experience and/or literature.

Notation	Value	Description
N	1400812636	Total population.
t_0	1-12-2019	Initial date.
T	29-3-2020	Final date.
λ_1	23-1-2020	Date when travel restrictions were imposed in Wuhan.
λ_2	8-2-2020	Inflexion date.
$\underline{\theta}$	14%	Percentage of documented cases at λ_1.
$\overline{\theta}$	65%	Percentage of documented cases at λ_2.
α_H	2.75%	Percentage of infection produced by hospitalized people.
d_E	5.5	Average days in compartment E.
d_I	6.7	Average days in compartment I.
d_{I_u}	$14 - d_I = 7.3$	Average days in compartment I_u.
d_g	6	Maximum reduction of d_I due to the control measures.
C_o	14	The period of convalescence.
$p(t)$	1	Fraction of the infected people hospitalized.

Table 2

Parameters obtained by calibration to the data.

Notation	Value	Description
β_I	0.2887	Disease contact rate of a person in compartment I.
C_E	0.3643	Reduction factor of the disease contact rate β_E w.r.t β_I.
C_u	0.4010	Reduction factor of the disease contact rate $\underline{\beta}_I$ w.r.t β_I.
δ_R	7.0000	Difference between days in compartment H_R and H_D.
δ_ω	0.0206	Difference between $\underline{\omega}$ and $\overline{\omega}$.
$\underline{\omega}$	0.0157	Lower bound of the fatality rate.
κ_1	0.1082	Efficiency of the control measures.

Table 3

Variables of the Stochastic θ-SEIHRD model: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $n = 2^{15}$ Monte Carlo simulations. Columns: Mean, interquartile interval ($[Q_1, Q_3]$) and worst case scenario (WS).

	8th February, 2020 ($t = 69$)						
	$\sigma_{\beta_I} = 0$	$\sigma_{\beta_I} = 0.1$			$\sigma_{\beta_I} = 0.5$		
	Mean	Mean	$[Q_1, Q_3]$	WS (95%)	Mean	$[Q_1, Q_3]$	WS (95%)
$E(t)$	2993	3067	[2506, 3519]	4510	5401	[1049, 5415]	18970
$I(t)$	1340	1376	[1125, 1578]	2017	2419	[476, 2423]	8522
$I_u(t)$	3854	3945	[3249, 4505]	5724	6811	[1434, 6940]	23728
$H_R(t)$	3252	3328	[2732, 3806]	4854	5799	[1182, 5863]	20340
$H_D(t)$	214	219	[181, 250]	318	377	[80, 386]	1311
$R_d(t)$	1846	1888	[1559, 2153]	2726	3231	[701, 3317]	11168
$R_u(t)$	4296	4390	[3656, 4985]	6238	7301	[1738, 7690]	24654
$D(t)$	131	134	[112, 152]	190	222	[53, 235]	747
	29th March, 2020 ($t = 119$)						
	$\sigma_{\beta_I} = 0$	$\sigma_{\beta_I} = 0.1$			$\sigma_{\beta_I} = 0.5$		
	Mean	Mean	$[Q_1, Q_3]$	WS (95%)	Mean	$[Q_1, Q_3]$	WS (95%)
$E(t)$	1	1	[1, 1]	2	2	[0, 3]	10
$I(t)$	0	0	[0, 0]	0	0	[0, 0]	1
$I_u(t)$	173	177	[146, 203]	259	309	[63, 314]	1075
$H_R(t)$	232	237	[194, 272]	348	416	[83, 420]	1458
$H_D(t)$	30	30	[25, 35]	45	53	[11, 54]	186
$R_d(t)$	8460	8662	[7118, 9910]	12624	15087	[3101, 15287]	52651
$R_u(t)$	9969	10198	[8442, 11616]	14681	17386	[3862, 17941]	59616
$D(t)$	417	426	[353, 486]	614	728	[161, 751]	2502

uncertainty ($\sigma_{\beta_I} = 0$), low uncertainty ($\sigma_{\beta_I} = 0.1$) and high uncertainty ($\sigma_{\beta_I} = 0.5$). In all cases, the long-term mean of the perturbation is set to the calibrated value of β_I for the deterministic model, i.e., $\mu_{\beta_I} = \beta_I$. The mean reverting speed parameter is chosen as $\nu_{\beta_I} = 1$, thus representing a regular (not too high, not too low) reversion speed. In Table 3, the obtained results are presented. We can clearly observe the significant impact of the uncertainty in the disease evolution. In the case of higher uncertainty, the number of infections, in average, is almost doubled and the worst case scenario multiplies this value by six. Even when a lower volatility is considered, the increment of cases and deaths in the worst scenario becomes important, up to 50%.

Next to the previous experiment, in Fig. 6 we present the histograms for the model variable $I(t)$ to give an insight of the impact of the uncertainty in the infection evolution. First, we clearly observe that the produced distribution is skewed
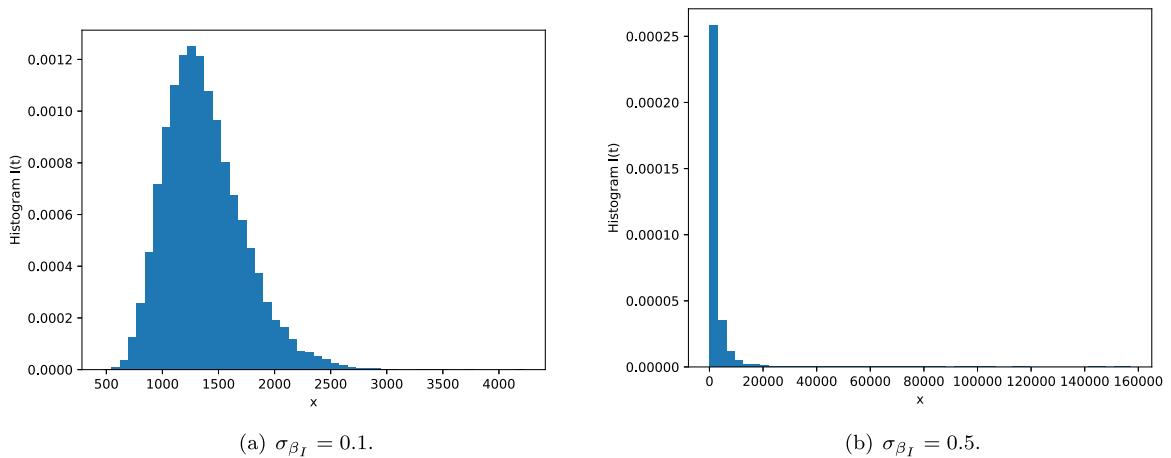
(a) $\sigma_{\beta_I} = 0.1$.   (b) $\sigma_{\beta_I} = 0.5$.

Fig. 6. Histogram of $I(t)$. Setting: $\nu_{\beta_I} = 1$ and $\mu_{\beta_I} = \beta_I$, with $n = 2^{15}$ Monte Carlo simulations.

with a fatter right tail. Secondly, the bigger the volatility of process $\tilde{\beta}_I(t)$ (denoted by σ_{β_I}), the fatter the right tail becomes, thus indicating more probability of extreme events.

4.2. Outputs

In [8] the authors proposed a set of possible outputs that can be useful for the authorities to plan the resources allocation (like the number of clinical beds, among other indicators). These outputs are:

- The cumulative number of COVID-19 cases, $c_m(t)$, at time t:

$$c_m(t) = H_R(t) + H_D(t) + R_d(t) + D(t) = c_m(t_0) + \int_{t_0}^{t} \theta(s)\gamma_I(s)I(s)\mathrm{d}s.$$

- The cumulative number of deaths due to the COVID-19 at time t: $d_m(t) = D(t)$.
- The basic reproduction number, R_0, and the effective reproduction number, $R_e(t)$, at time t, where $R_0 = R_e(t_0)$, and

$$R_e(t) = \frac{U_e(t)}{\gamma_E \gamma_I(t)\gamma_{H_R}(t)\gamma_{H_D}(t)\gamma_{I_u}(t)} \frac{S(t)}{N},$$

  with[1]

$$U_e(t) = \left(\left((m_{I_u}\beta_{I_u}(1-\theta)\gamma_{H_R} + m_{H_R}\beta_{H_R}\gamma_{I_u}(\theta - \omega))\gamma_I + m_I\beta_I\gamma_{H_R}\gamma_{I_u}\right)\gamma_E + m_E\beta_E\gamma_I\gamma_{H_R}\gamma_{I_u}\right)\gamma_{H_D}$$
$$+ m_{H_D}\beta_{H_D}\omega\gamma_E\gamma_I\gamma_{H_R}\gamma_{I_u}.$$

- Hospitalized people, $\mathrm{Hos}(t)$, at time t:

$$\mathrm{Hos}(t) = H_D(t) + p(t)\left(H_R(t) + R_d(t) - R_d(t - C_o)\right),$$

  where $p(t)$ is the fraction of people in compartment H_R that are hospitalized and C_o is the period of convalescence.
- Maximum number of hospitalized people in the interval $[t_0, t]$:

$$\mathrm{MHos}(t) = \max_{\tau \in [t_0, t]} \mathrm{Hos}(\tau).$$

- The number of individuals infected by others belonging to compartments E, I_u and $H = H_R + H_D$:
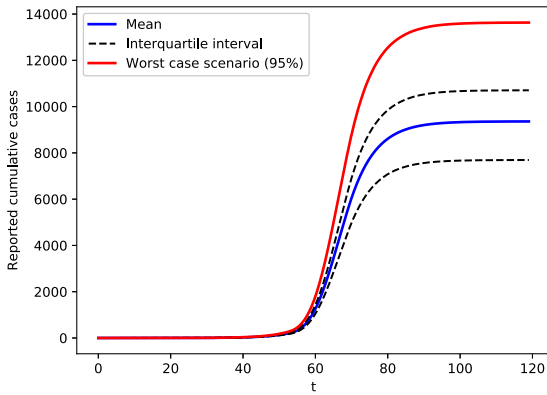
$$\Gamma_E(t) = \int_{t_0}^{t} m_E(s)\beta_E E(s)\frac{S(s)}{N}\mathrm{d}s,$$

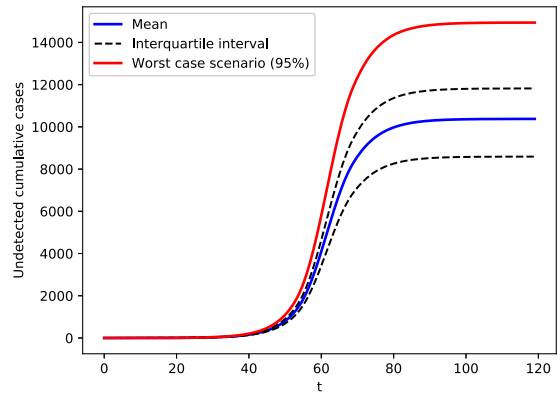$$\Gamma_{I_u}(t) = \int_{t_0}^{t} m_{I_u}(s)\beta_{I_u} I_u(s)\frac{S(s)}{N}\mathrm{d}s,$$

$$\Gamma_H(t) = \int_{t_0}^{t} \left(m_{H_R}(s)\beta_{H_R} H_R(s) + m_{H_D}(s)\beta_{H_D} H_D(s)\right)\frac{S(s)}{N}\mathrm{d}s,$$
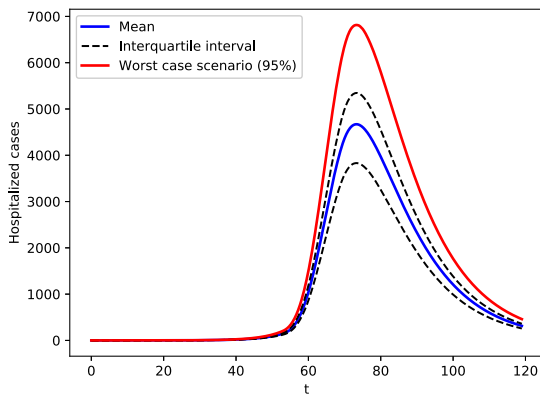
  respectively.

---

[1] The dependence of the coefficients on the time t has been omitted for notational purposes. All the coefficients take their particular values at time t.
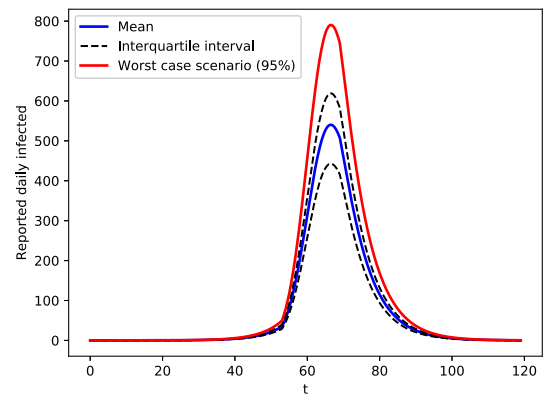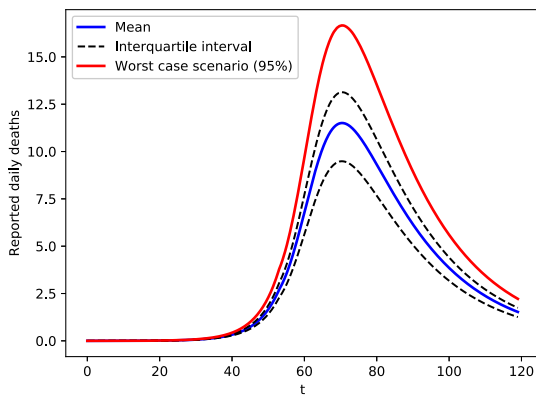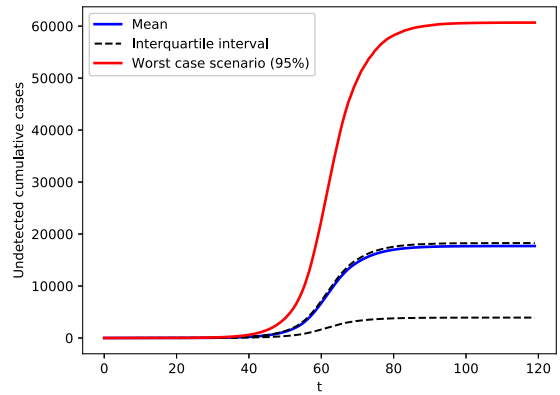
Fig. 7. Epidemic curves: mean, interquartile interval and worst case scenario. Setting: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $\sigma_{\beta_I} = 0.1$, with $n = 2^{15}$ Monte Carlo simulations.

We therefore perform a similar statistical analysis as before, although now reporting the model outputs. The results are shown in Table 4. Again, it is clear that the uncertainty can significantly affect the disease evolution, justifying why it should be taken into consideration. For example, the people requiring hospitalization may vary from around 4500 with no stochasticity included, to around 30000 including randomness.
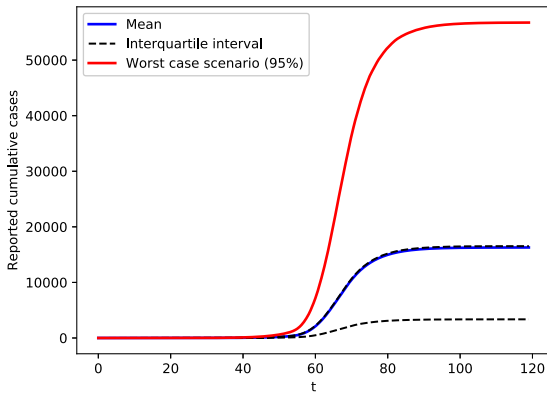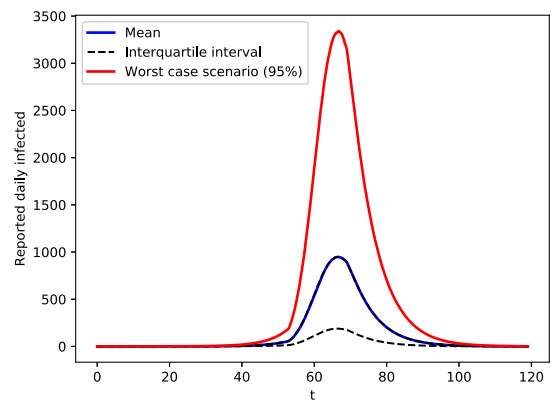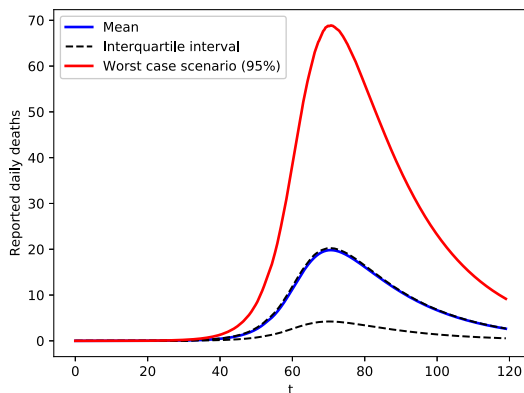
(a) Reported cumulative.
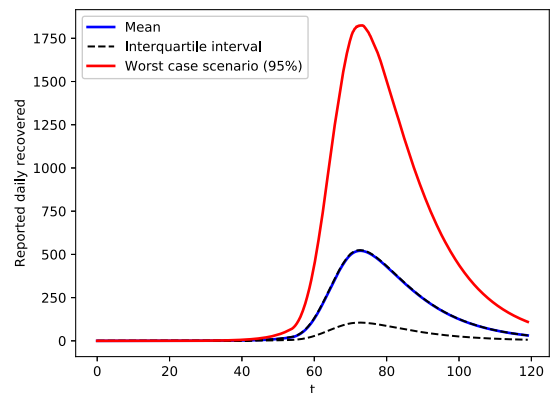
(b) Undetected cumulative.
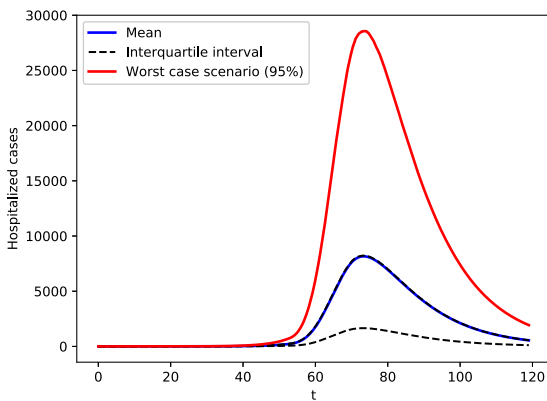
(c) Hospitalized.

(d) Reported daily infected.

(e) Reported daily deaths.

(f) Reported daily recovered.

Fig. 8. Epidemic curves: mean, interquartile interval and worst case scenario. Setting: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $\sigma_{\beta_I} = 0.5$, with $n = 2^{15}$ Monte Carlo simulations.

As previously pointed out, one of the major advantages of the θ-SEIHRD model is that it accounts for important aspects of the COVID-19 pandemic, directly affecting the population or the healthcare systems. Particularly, the evolution of some curves like infected, hospitalized, and deaths is typically reported by the authorities in both cumulative and daily fashion. In Figs. 7 and 8, we show the stochastic model outcomes for these specific curves, considering two uncertainty levels,

Table 4
Outputs of the Stochastic θ-SEIHRD model: $\nu_{\beta_I} = 1$, $\mu_{\beta_I} = \beta_I$ and $n = 2^{15}$ Monte Carlo simulations. Columns: Mean, interquartile interval ($[Q_1, Q_3]$) and worst case scenario (WS).

	8th February, 2020 ($t = 69$)						
	$\sigma_{\beta_I} = 0$	$\sigma_{\beta_I} = 0.1$			$\sigma_{\beta_I} = 0.5$		
	Mean	Mean	$[Q_1, Q_3]$	WS (95%)	Mean	$[Q_1, Q_3]$	WS (95%)
$c_m(t)$	5440	5571	[4586, 6362]	8088	9631	[2026, 9812]	33466
$d_m(t)$	131	134	[112, 152]	190	222	[53, 235]	747
$R_e(t)$	0.3363	0.3364	[0.3151, 0.3562]	0.3891	0.3367	[0.2245, 0.4149]	0.6340
Hos(t)	4040	4134	[3395, 4727]	6026	7197	[1471, 7273]	25262
MHos(t)	4040	4134	[3395, 4727]	6026	7197	[1471, 7273]	25262
$\Gamma_E(t)$	5012	5126	[4255, 5833]	7337	8625	[1979, 9013]	29414
$\Gamma_{I_u}(t)$	4550	4646	[3864, 5285]	6640	7600	[1764, 7976]	25755
$\Gamma_H(t)$	198	202	[168, 230]	288	328	[78, 346]	1106
	29th March, 2020 ($t = 119$)						
	$\sigma_{\beta_I} = 0$	$\sigma_{\beta_I} = 0.1$			$\sigma_{\beta_I} = 0.5$		
	Mean	Mean	$[Q_1, Q_3]$	WS (95%)	Mean	$[Q_1, Q_3]$	WS (95%)
$c_m(t)$	9140	9358	[7691, 10704]	13631	16286	[3358, 16526]	56752
$d_m(t)$	417	426	[353, 486]	614	728	[161, 751]	2502
$R_e(t)$	0.0013	0.0013	[0.0012, 0.0014]	0.0015	0.0013	[0.000, 0.0016]	0.0024
Hos(t)	306	314	[257, 360]	459	549	[111, 555]	1927
MHos(t)	4558	4671	[3832, 5347]	6816	8195	[1662, 8258]	28681
$\Gamma_E(t)$	5259	5379	[4464, 6122]	7705	9070	[2073, 9465]	30925
$\Gamma_{I_u}(t)$	5388	5504	[4570, 6264]	7886	9082	[2080, 9479]	30911
$\Gamma_H(t)$	229	234	[195, 266]	334	384	[89, 402]	1298

$\sigma_{\beta_I} = 0.1$ and $\sigma_{\beta_I} = 0.5$, respectively. Again, we present the mean, the interquartile interval and the worst case scenario. From this experiment, an interesting observation can be extracted. Looking at the different patterns in the results w.r.t. the volatility parameter, we can see that an increasing uncertainty pushes the mean close to the third quartile, Q_3, meaning that the disease evolves, in average, according to the 75% worst case scenario. This fact gives an insight of how important the randomness can be and why it is crucial to include it in the modelling.

5. Discussion and conclusions

We have extended the model developed in [8] by incorporating randomness to some relevant coefficients and we have shown the importance of considering this uncertainty. In this way, besides the information provided by the deterministic model (which allows to obtain a proxy to the average of the main variables), we can take advantage of a more complete modelling approach, which allows not only to compute confidence intervals for these variables in this new random setting but also to obtain the worst case scenario. The information about the model variables in this worst case scenario allows to develop more conservative policies in the actions against the consequences of the COVID-19 as, for example, to plan larger health resources to take care of a larger number of people requiring hospitalization at different levels. The differences between the deterministic and stochastic models in terms of the information contained in the output variables has been clearly illustrated in the previous section.

However, we also understand that research in this line can be extended. A first possible extension comes from making the parameters independent among each other and use different stochastic processes to characterize their dynamics. In the current approach, as in [8], we consider that all parameters depend on β_I. Also, it seems possible to incorporate randomness to the more recent model in [26] that mainly considers a new compartment of persons in quarantine (Q) to model the situation in certain countries like Italy.

CRediT authorship contribution statement

Álvaro Leitao: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft, Writing – review & editing. **Carlos Vázquez:** Conceptualization, Methodology, Validation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgements

References

[1] Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A Math Phys Eng Sci 1927;115(772):700–21. http://dx.doi.org/10.1098/rspa.1927.0118, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0118, URL https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1927.0118.

[2] Hethcote HW. The mathematics of infectious diseases. SIAM Rev 2000;42(4):599–653. http://dx.doi.org/10.1137/S0036144500371907.

[3] Venkatramanan S, Lewis B, Chen J, Higdon D, Vullikanti A, Marathe M. Using data-driven agent-based models for forecasting emerging infectious diseases. Epidemics 2018;22:43–9. http://dx.doi.org/10.1016/j.epidem.2017.02.010, The RAPIDD Ebola Forecasting Challenge, URL http://www.sciencedirect.com/science/article/pii/S1755436517300221.

[4] Squazzoni F, Polhill JG, Edmonds B, Ahrweiler P, Antosz P, Scholz G, et al. Computational models that matter during a global pandemic outbreak: A call to action. J Artif Soc Soc Simul 2020;23(2):10. http://dx.doi.org/10.18564/jasss.4298, URL http://jasss.soc.surrey.ac.uk/23/2/10.html.

[5] Allen LJ. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infect Dis Model 2017;2(2):128–42. http://dx.doi.org/10.1016/j.idm.2017.03.001, URL http://www.sciencedirect.com/science/article/pii/S2468042716300495.

[6] Bretó C, He D, Ionides EL, King AA. Time series analysis via mechanistic models. Ann Appl Stat 2009;3(1):319–48, URL http://www.jstor.org/stable/30244243.

[7] He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. J R Soc Interface 2010;7(43):271–83. http://dx.doi.org/10.1098/rsif.2009.0151, arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2009.0151, URL https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2009.0151.

[8] Ivorra B, Ferrández MR, Vela-Pérez M, Ramos AM. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. Commun Nonlinear Sci Numer Simul 2020;88:105303. http://dx.doi.org/10.1016/j.cnsns.2020.105303.

[9] Ivorra B, Ngom D, Ramos A. Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries-validation and application to the 2014–2015 ebola virus disease epidemic. Bull Math Biol 2015;77:1668–704. http://dx.doi.org/10.1007/s11538-015-0100-x.

[10] Kucharski A, Russell T, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020;20. http://dx.doi.org/10.1016/S1473-3099(20)30144-4.

[11] Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, et al. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infect Dis Model 2020;5:256–63. http://dx.doi.org/10.1016/j.idm.2020.02.002, URL http://www.sciencedirect.com/science/article/pii/S2468042720300051.

[12] Platen E. Stochastic modelling of the COVID-19 epidemic. 2020, URL https://ssrn.com/abstract=3586208, Available at SSRN: https://ssrn.com/abstract=3586208.

[13] Lipton A, de Prado ML. Mitigation strategies for COVID-19: lessons from the K-SEIR model. 2020, URL https://ssrn.com/abstract=3623544, Available at SSRN: https://ssrn.com/abstract=3623544.

[14] Han X, Li F, Meng X. Dynamics analysis of a nonlinear stochastic SEIR epidemic system with varying population size. Entropy 2018;20(5). http://dx.doi.org/10.3390/e20050376, URL https://www.mdpi.com/1099-4300/20/5/376.

[15] Li F, Meng X, Wang X. Analysis and numerical simulations of a stochastic SEIQR epidemic system with quarantine-adjusted incidence and imperfect vaccination. Comput Math Methods Med 2018. http://dx.doi.org/10.1155/2018/7873902.

[16] Gray A, Greenhalgh D, Hu L, Mao X, Pan J. A stochastic differential equation SIS epidemic model. SIAM J Appl Math 2011;71(3):876–902. http://dx.doi.org/10.1137/10081856X.

[17] Dureau J, Kalogeropoulos K, Baguelin M. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. Biostatistics 2013;14(3):541–55. http://dx.doi.org/10.1093/biostatistics/kxs052, arXiv:https://academic.oup.com/biostatistics/article-pdf/14/3/541/17739168/kxs052.pdf.

[18] Cai S, Cai Y, Mao X. A stochastic differential equation SIS epidemic model with two independent Brownian motions. J Math Anal Appl 2019;474(2):1536–50. http://dx.doi.org/10.1016/j.jmaa.2019.02.039, URL http://www.sciencedirect.com/science/article/pii/S0022247X19301635.

[19] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020;368(6490):489–93. http://dx.doi.org/10.1126/science.abb3221, arXiv:https://science.sciencemag.org/content/368/6490/489.full.pdf, URL https://science.sciencemag.org/content/368/6490/489.

[20] World Health Organization. Coronavirus disease (COVID-2019) situation reports. 2020, URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/.

[21] Cox JC, Ingersoll JE, Ross SA. A theory of the term structure of interest rates. Econometrica 1985;53(2):385–407, URL http://www.jstor.org/stable/1911242.

[22] Han X, Kloeden PE. Random ordinary differential equations and their numerical solution. Singapore: Springer; 2017.

[23] Kloeden PE, Platen E. Numerical methods for stochastic differential equations. Berlin: Springer; 1992.

[24] Oosterlee CW, Grzelak LA. Mathematical modeling and computation in finance. World Scientific (Europe); 2019, http://dx.doi.org/10.1142/q0236, arXiv:https://worldscientific.com/doi/pdf/10.1142/q0236, URL https://worldscientific.com/doi/abs/10.1142/q0236.

[25] Dormand JR, Prince P. A family of embedded runge-kutta formulae. J Comput Appl Math 1980;6(1):19–26. http://dx.doi.org/10.1016/0771-050X(80)90013-3, URL http://www.sciencedirect.com/science/article/pii/0771050X80900133.

[26] Ramos A, Ferrández M, Vela-Pérez M, Kubik A, Ivorra B. A simple but complex enough θ-SIR type model to be used with COVID-19 real data. Application to the case of Italy. Physica D 2021;421:132839. http://dx.doi.org/10.1016/j.physd.2020.132839, URL https://www.sciencedirect.com/science/article/pii/S016727892030840X.