A stochastic SEIHR model for COVID-19 data fluctuations

Ruiwu Niu · **Yin-Chi Chan** · **Eric W. M. Wong** · **Michaël Antonie van Wyk** ·
Guanrong Chen

Abstract Although deterministic compartmental models are useful for predicting the general trend of a disease's spread, they are unable to describe the random daily fluctuations in the number of new infections and hospitalizations, which is crucial in determining the necessary healthcare capacity for a specified level of risk. In this paper, we propose a stochastic SEIHR (sSEIHR) model to describe such random fluctuations and provide sufficient conditions for stochastic stability of the disease-free equilibrium, based on the basic reproduction number that we estimated. Our extensive numerical results demonstrate strong threshold behavior near the estimated basic reproduction number, suggesting that the necessary conditions for stochastic stability are close to the sufficient conditions derived. Furthermore, we found that increasing the noise level slightly reduces the final proportion of infected individuals. In addition, we analyze COVID-19 data from various regions worldwide and demonstrate that by changing only a few parameter values, our sSEIHR model can accurately describe both the general trend and the random fluctuations in the number of daily new cases in each region, allowing governments and hospitals to make more accurate caseload predictions using fewer compartments and parameters than other comparable stochastic compartmental models.

Keywords COVID-19 · data fluctuation · SEIHR model · stochastic differential equation · stochastic stability

Ruiwu Niu
College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, PR China

Y.-C. Chan · E. W. M. Wong (✉) · G. Chen
Department of Electrical Engineering, City University of Hong Kong, 83 Tat Chee Ave., Kowloon Tong, Hong Kong
E-mail: eeewong@cityu.edu.hk

M. A. van Wyk
School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2000, South Africa

1 Introduction

Over the past year, the Coronavirus Disease 2019 (COVID-19) pandemic has placed enormous stress on healthcare systems worldwide. The severity of the pandemic, given the fact that it takes a long time to develop effective vaccines, has imposed tremendous pressures and responsibilities onto the healthcare systems in all countries and regions, especially those with limited medical resources such as available staff, equipment and facilities. In particular, differing from most infectious diseases found to date, COVID-19 is especially violent, aggressive and fast-spreading, and even among those deemed "recovered" from the disease, there are many for which adverse effects have lingered for months after the initial symptoms [1]. Meanwhile, failure to meet the demand for hospital resources can lead to resource saturation and a growing backlog of infectious patients requiring hospitalization, in turn increasing the total transmission rate due to unisolated infectious individuals and causing an adverse feedback loop.

1.1 Related studies on COVID-19

Due to an incomplete knowledge of COVID-19 during the early stages, scientists used existing models to forecast the pandemic and made inaccurate predictions. Fortunately, based on historical data, people can still make relatively accurate predictions by applying some kinds of model-free methods [2,3]. However, to generate policy-relevant insights into the nonpharmaceutical interventions, people still need to understand the physical principles of the pandemic. Since the clinical data of COVID-19 has been carefully studied [4], scientists now can make more accurate epidemic predictions using corresponding mathematical models [5,6]. Ref. [7] presents the details of three regional-scale models for pre-

dicting and estimating the trend of the pandemic. We also introduced a compartmental model and made an accurate trend forecasting on Hong Kong's COVID-19 pandemic [8].

To predict the number of cases in a region and the corresponding healthcare demands, epidemiological models are very useful and have been constructed to describe the spread of diseases like COVID-19. In particular, compartmental models, which track the evolution of the number or proportion of the individuals in each disease state (susceptible, infected, removed, etc.) using a system of ordinary or stochastic differential equations, have been developed for nearly a century and have proved effective and useful [9, 10].

1.2 Contributions of this paper

In this paper, we develop a stochastic version of the modified SEIHR compartmental model established in [8], which allows one to accurately model the random fluctuations in the number of daily new cases in the COVID-19 epidemic process. We prove sufficient conditions on the parameters of the new stochastic SEIHR (sSEIHR) model, such that the disease-free equilibrium (DFE) of the system is stochastically stable, based on an estimated basic reproduction number. Our extensive numerical results demonstrate strong threshold behavior near the estimated basic reproduction number, suggesting that the necessary conditions for stochastic stability are actually close to the sufficient conditions established.

The proposed sSEIHR model is then fitted to public COVID-19 data from various regions worldwide, and it is demonstrated numerically that the model can accurately describe both the general trend and the random fluctuation in the number of daily new cases in each region. The ability of our model to accommodate a wide range of geographical regions and the variations in their various COVID-19 outcomes with minimal adjustments to model parameters shows the robustness of our model. This can offer suggestions to governments and hospitals for their making more accurate caseload predictions compared to using a deterministic model.

Recall that it has been demonstrated in [8] that deterministic models using fewer compartments than the SEIHR model, e.g. SIR or SEIR, fail to accurately describe the evolution of COVID-19. Therefore, except potentially by means of more complex compartment interactions, both the SEIHR model developed in [8] and the new sSEIHR model developed here are the smallest compartmental models that can accurately describe the number of active COVID-19 cases in a region of concern. The model's performance suggests that it captures key hidden dynamical features of general epidemics in general environments and deserves independent investigation.

The true significance of this work is the generality of the proposed model with few parameters. In particular, it is surprising to find that the model fits very well to a number of regions/countries worldwide with only a few parameter value changes, under the consideration of the heterogeneity of economic and social features, medical resources, quarantine measures, mitigation strategies and healthcare systems' characteristics in different regions/countries. In addition, *within* each region, changes in non-pharmaceutical intervention (NPI) strength over time can be expressed as a *single* parameter, while all other parameters remaining constant. The robustness and accuracy observed make this model valuable for real applications.

2 Background and related work

Notation For vectors $x = [x_1, \cdots, x_n]^T \in \mathbb{R}^n$, let D_n denote the set $\{x \in \mathbb{R}^n : \sum_i x_i = 1\}$ and D_n^* denote the set $\{x \in [0, 1]^n : \sum_i x_i = 1\}$.

2.1 The SIR model

A classical approach for epidemiological modeling is to use compartmental models [9]. One such compartmental model is the SIR model, in which individuals are classified as susceptible (S), infected (I), or removed (R) (either recovered or deceased). The number of individuals in each state, i.e. compartment, can be described by a system of ordinary differential equations.

Let $\mathcal{S}(t)$, $\mathcal{I}(t)$, and $\mathcal{R}(t)$ denote the numbers of susceptible, infected, and removed individuals, respectively, and let N be the population size, at time $t \geq 0$. The system can then be described by

$$\dot{x} = \begin{bmatrix} \dot{S} \\ \dot{I} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} \mu - (\beta I + \mu) S \\ \beta SI - (\gamma + \mu) I \\ \gamma I - \mu R \end{bmatrix} \tag{1}$$

with

$$x(t) = \begin{bmatrix} S(t) \\ I(t) \\ R(t) \end{bmatrix} = \begin{bmatrix} \mathcal{S}(t)/N \\ \mathcal{I}(t)/N \\ \mathcal{R}(t)/N \end{bmatrix},$$

where $x(t) \in D_3^*$, with $I(0) > 0$ and $\{\beta, \gamma, \mu\} > 0$. For this model, the basic reproduction number $R_0 = \beta/(\gamma + \mu)$ is important, which can be thought of as the number of expected cases directly generated by an infected individual, if all other individuals are susceptible to infection. The epidemic is expected to persist if $R_0 > 1$, but will die out if $R_0 < 1$.

It was proved [11] that, for $x(t) \in D_3^*$, R_0 uniquely determines whether the epidemic through system (1) will die out ($I(\infty) = 0$) or persist ($I(\infty) > 0$), independent of the nonzero initial conditions. This threshold behavior demonstrates the power and usefulness of compartmental modeling in the study of epidemic spreading.

2.2 Other deterministic models

Numerical results shown in [8] demonstrate that the SIR model is inadequate for describing COVID-19 dynamics. Reasons include:

- Like some other diseases, COVID-19 has a long latent period. The SIR model does not take this into account, however. As an improvement, SEIR models have been developed to model the effect of latent periods, where E represents an additional "exposed" compartment [12].
- Ideally, symptomatic individuals are detected and treated in hospitals, wherein they cannot infect the general community. The SIR model does not take this into account, either. As an improvement, the SEIQR or SEIHR model creates an additional compartment for such individuals, with Q and H representing "quarantined" and "hospitalized", respectively [13].
- The COVID-19 pandemic has a large number of asymptomatic infections. The SIR model certainly does not take this into account. Therefore, some other models seek to divide infected individuals into symptomatic and asymptomatic compartments. Alternatively, the E compartment in the SEIR model can be made infectious [8].

To account for the above issues simultaneously, in [14] a nine-compartment model is constructed, and the model is referred to as θ-SEIHRD, in which there are two I, H, and R compartments, respectively. However, while the θ-SEIHRD model is useful for detailed study of COVID-19 dynamics, it may not be desirable to use such a complicated model for more effective parameter fitting, especially if only the total number of active cases is required to be estimated. This problem is resolved in [8], with a model having only five compartments:

$$\dot{x} = \begin{bmatrix} \dot{S} \\ \dot{E} \\ \dot{I} \\ \dot{H} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} -(\eta E + \alpha I) S \\ (\eta E + \alpha I) S - (\beta + \omega_E) E \\ \beta E - (\gamma + \omega_I) I \\ \gamma I - \omega_H H \\ \omega_E E + \omega_I I + \omega_H H \end{bmatrix} \quad (2)$$

with $x(t) \in D_5^*$, $\{\alpha, \eta, \omega_E, \omega_I, \omega_H\} > 0$, and $E(0) + I(0) > 0$, where S, E, I, H, and R denote the proportions of susceptible, exposed, infectious, hospitalized, and removed individuals, respectively. Note that in our model, "infectious" denotes symptomatic infectious individuals, whereas "exposed" includes both non-infectious exposed individuals and asymptomatic infectious individuals. The parameter η allows for asymptomatic transmissions, a key feature of COVID-19 propagation. The meaning of each parameter in (2) is given in Table 1. Further interpretation about the compartments can be found in [8].

The basic reproduction number of system (2) is

$$R_0 = \frac{\eta}{\beta + \delta_E} + \frac{\alpha\beta}{(\beta + \delta_E)(\gamma + \delta_I)}.$$

Table 1 Meaning of parameters in the SEIHR model (2)

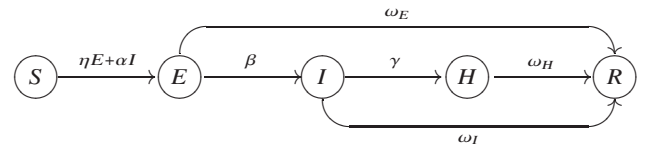| Parameter	Definition
η	Transmission rate of exposed individuals
α	Transmission rate of (symptomatic) infected individuals
β	Reciprocal of the mean latent period, i.e. the rate at which exposed individuals become symptomatic
γ	Rate at which infected individuals are hospitalized
ω_E	Rate of recovery of non-hospitalized exposed individuals
ω_I	Rate of recovery of non-hospitalized infected individuals
ω_H	Rate of recovery of hospitalized individuals

Fig. 1 Graphical depiction of the dSEIHR model.

Hereafter, the term "deterministic SEIHR model" or "dSEIHR model" refers to the one described by system (2). A diagram showing the transitions of the dSEIHR model (2) is given in Fig. 1. Note that (2) differs from the traditional SEIHR model in that the exposed individuals may also be infectious.

To further improve the model ability in prediction, especially over a long time period, a factor $p(t) \in [0, 1]$ was introduced in [8] to denote the strength of various interventions (a smaller $p(t)$ implies a stronger intervention), leading to

$$\dot{x} = \begin{bmatrix} \dot{S} \\ \dot{E} \\ \dot{I} \\ \dot{H} \\ \dot{R} \end{bmatrix} = \begin{bmatrix} -p(t)[\eta E + \alpha I] S/N \\ p(t)[\eta E + \alpha I] S/N - (\beta + \omega_E) E \\ \beta E - (\gamma + \omega_I) I \\ \gamma I - \omega_H H \\ \omega_E E + \omega_I I + \omega_H H \end{bmatrix}, \quad (3)$$

which is called the p-dSEIHR model.

For fitting the p-dSEIHR model (3) to real COVID-19 data, piecewise constant functions were used in [8] to approximate $p(t)$. This matches the discrete nature of governmental and healthcare policy, where occasional policy changes will cause immediate changes in $p(t)$.

2.3 Stochastic epidemic models

While the deterministic models described in Sections 2.1 and 2.2 have proven useful for estimating the cumulative number of infections over time, these do not capture the stochastic nature of disease propagation. This can lead to healthcare dimensioning and resource allocation problems if

the actual number of daily new cases is significantly different from the predicted value, which in turn can lead to resource saturation and a growing backlog of infectious patients requiring hospitalization, increasing the total transmission rate due to unisolated infectious individuals and causing an adverse feedback loop.

To describe random fluctuations, stochastic compartmental models are commonly used, where the system in Section 2.2 is modified to be a drift-diffusion process of the form

$$d\boldsymbol{x} = f(\boldsymbol{x}, t)\, dt + g(\boldsymbol{x}, t)\, dW(t)\,, \tag{4}$$

where f and g are vector-valued functions and $W(t)$ denotes a Wiener process, i.e. that derived by integrating Gaussian white noise.

Examples of stochastic compartmental models in the literature include [15] and [16], which are the stochastic analogs of the deterministic SIR and θ-SEIHRD models described in Section 2.2, respectively. Notably, in [15,16] it is assumed that the noise in these two stochastic models is attributable to a single state transition, i.e. S→I in [15] and S→E in [16]. The same assumption will be adopted for the new sSEIHR model.

3 Stochastic SEIHR model with Gaussian white noise

3.1 Model formulation

While the deterministic SEIHR model (2) addresses the COVID-19 phenomena listed in Section 2.2, it cannot model the random fluctuations in the number of daily new cases. To address this issue, we therefore convert system (2) to a stochastic SEIHR (sSEIHR) model by introducing an additive random component to the I→H transition, thus obtaining the following drift-diffusion process:

$$d\boldsymbol{x} = d\begin{bmatrix} S(t) \\ E(t) \\ I(t) \\ H(t) \\ R(t) \end{bmatrix} = \begin{bmatrix} -[\eta E(t) + \alpha I(t)]\, S(t) \\ [\eta E(t) + \alpha I(t)]\, S(t) - (\beta + \omega_E)\, E(t) \\ \beta E(t) - (\gamma + \omega_I)\, I(t) \\ \gamma I(t) - \omega_H H(t) \\ \omega_E E(t) + \omega_I I(t) + \omega_H H(t) \end{bmatrix} dt$$
$$+ \begin{bmatrix} 0 \\ 0 \\ -\varphi I(t) \\ \varphi I(t) \\ 0 \end{bmatrix} dW(t)\,, \tag{5}$$

with $\boldsymbol{x}(t) \in D_5$ and $E(0) + I(0) > 0$, where $W(t)$ denotes a Wiener process obtained from integrating Gaussian-distributed white noise. Here, in this model, only the two terms I and H are considered having random fluctuations due, for example, to uncertain hospital management.
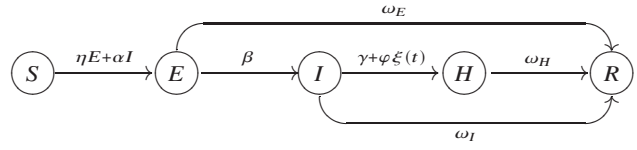

Fig. 2 Graphical depiction of the sSEIHR model (5), where $\xi(t)$ denotes Gaussian white noise.

A graphical depiction of system (5) is shown in Fig. 2, where $\xi(t)$ denotes Gaussian white noise.

3.2 Stability results

It is easy to see that (5) has a disease-free equilibrium DFE = $(1, 0, 0, 0, 0)$. Regarding the system stability at this equilibrium, we have the following main result.

Theorem 1 *If*

$$R_0^* = \frac{\eta}{\beta + \delta_E} + \frac{\alpha\beta}{(\beta + \omega_E)\left(\gamma + \omega_I + \frac{\varphi^2}{2}\right)} < 1\,,$$

then the DFE $(1, 0, 0, 0, 0)$ of system (5) is stochastically stable, in the sense defined in Appendix A.

For a proof of Theorem 1, see Appendix B.

Note that in the absence of noise, i.e. $\varphi = 0$, R_0^* becomes the basic reproduction number of the corresponding deterministic system (2).

3.3 Numerical results

Consider a system with a total population of $N = 10^6$, where $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, $\mathcal{H}$, and $\mathcal{R}$ denote the numbers of susceptible, exposed, infected, hospitalized and removed individuals, respectively, with $S = \mathcal{S}/N$, $E = \mathcal{E}/N$, $I = \mathcal{I}/N$, $H = \mathcal{H}/N$, and $R = \mathcal{R}/N$, which evolve according to the stochastic differential equation (5).

The impact of the noise term $\varphi I(t)\, dW(t)$ on the epidemic process is examined, using the Euler–Maruyama method with initial state $\mathcal{S} = N - 1$, $\mathcal{E} = 1$, and $\mathcal{I} = \mathcal{H} = \mathcal{R} = 0$, with the system parameters $\beta = 0.14$, $\gamma = 0.7$, and $\omega_E = \omega_I = \omega_H = 0.1$.

Figure 3 shows the simulated value of $1 - S(\infty) = 1 - \lim_{t \to \infty} S(t)$ and R_0^*, for various values of $\alpha = \varepsilon$ and φ. It shows that the value of $1 - S(\infty)$ undergoes a sharp transition in the vicinity of the line $R_0^* = 1$, suggesting that the necessary conditions for stochastic stability are close to the sufficient conditions provided in Theorem 1. Furthermore, Figure 3 demonstrates that, when the noise parameter φ increases, the final value $1 - S(\infty)$ slightly decreases.

Figure 4 shows the impact of the noise parameter φ on the epidemic process. Note that for large values of φ, the
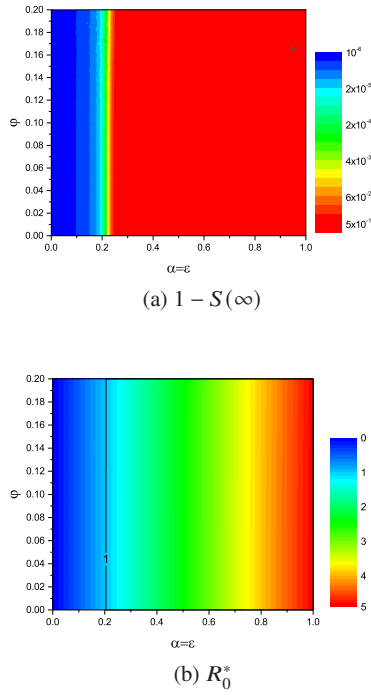
(a) $1 - S(\infty)$

(b) R_0^*

Fig. 3 Comparison between the simulated value of $1 - S(\infty)$ and R_0^* of the sSEIHR model (5) for the initial values and parameter settings defined in Section 3.3.

(a) $\varphi = 0.2$

(b) $\varphi = 0.05$

Fig. 4 Impact of the noise intensity φ on the epidemic process described by the sSEIHR model (5).

probability $\Pr\{\gamma + \varphi\xi(t) < 0\}$ becomes significant, such that patients can return to the state I from state H and even cause $H(t)$ to become negative, as can be seen clearly from Figure 4a. Therefore, smaller values of φ are generally more suitable for modeling the random fluctuations at the hospitalization rate $\mathrm{d}H(t)$ of the system. Furthermore, as the primarily concern is to estimate some rough probabilistic bounds on the number of daily new cases, the simulation with Gaussian white noise is sufficient.

3.4 Real-data analysis

Analogous to our modification of the dSEIHR model to produce the p-dSEIHR model, the sSEIHR model is modified by introducing a scaling factor $p(t)$ to denote the strength of various anti-epidemic interventions:

$$
\mathrm{d}\boldsymbol{x} = \mathrm{d}\begin{bmatrix} S(t) \\ E(t) \\ I(t) \\ H(t) \\ R(t) \end{bmatrix}
$$

$$
= \begin{bmatrix} -p\left[\eta E(t) + \alpha I(t)\right] S(t) \\ p\left[\eta E(t) + \alpha I(t)\right] S(t) - (\beta + \omega_E) E(t) \\ \beta E(t) - (\gamma + \omega_I) I(t) \\ \gamma I(t) - \omega_H H(t) \\ \omega_E E(t) + \omega_I I(t) + \omega_H H(t) \end{bmatrix} \mathrm{d}t
$$

$$
+ \begin{bmatrix} 0 \\ 0 \\ -\varphi I(t) \\ \varphi I(t) \\ 0 \end{bmatrix} \mathrm{d}W(t) .
$$

(6)

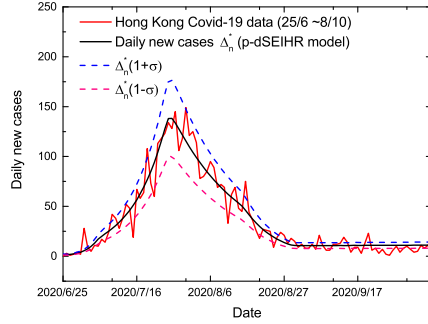This will be referred to as the p-sSEIHR model.

3.4.1 Hong Kong

Now, consider the number of daily new cases $\Delta_n = \mathcal{S}(n) - \mathcal{S}(n+1)$ of COVID–19 in Hong Kong, between 25 June and 8 Oct 2019. This is compared to the expected values Δ_n^* obtained from the p-dSEIHR model (3).

The fitted parameters of the p-dSEIHR model are given in Table 2, where $\omega_E = \omega_I = \omega_H = \omega$ and $p(t) = p_n$ for t in each period t_n shown, with $p(t) = 1$ for t prior to t_1. Note that this dataset corresponds to $\mathcal{H}(t)$ and H is the only observable compartment in the model. Note that the parameters N, β, γ, $\omega_E = \omega_I = \omega_H = \omega$, and t_n ($n = 1, 2, 3, 4$) are fixed parameters, so only η, α, and p_n ($n = 1, 2, 3, 4$) are fitted.
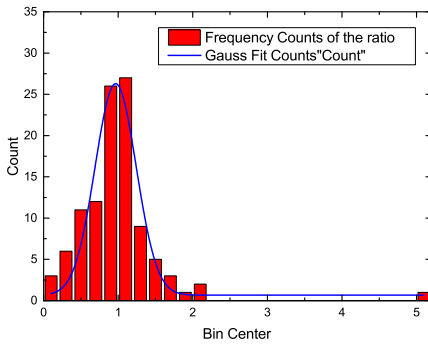
Since the noise term $\varphi I(t)\mathrm{d}W(t)$ in the p-sSEIHR model is proportional to $I(t)$, we are interested in the values $r_n =$

Table 2 Parameter setting for the Hong Kong dataset

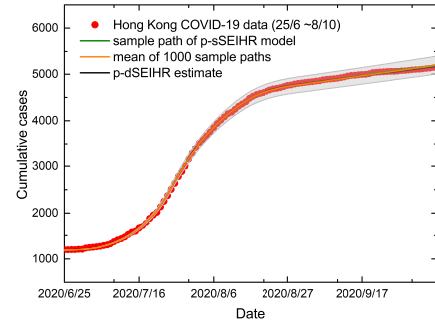| Parameter	Value	Parameter	Value
N	7 500 700	p_1	0.635
φ	0.03	t_1	4/7 ~ 23/7
η	0.48	p_2	0.345
α	0.5	t_2	24/7 ~ 13/8
β	0.14	p_3	0.235
γ	0.7	t_3	14/8 ~ 27/8
ω	0.1	p_4	0.4445
		t_4	28/8 ~ 8/10

(a) Daily new cases

(b) Distribution of $r_n = \Delta_n / \Delta_n^*$

Fig. 5 Ratios r_n and the estimated $(1 \pm \sigma)$-range of fluctuations in the number of daily new COVID-19 cases in Hong Kong, where $t = 0$ refers to 25 June 2020.
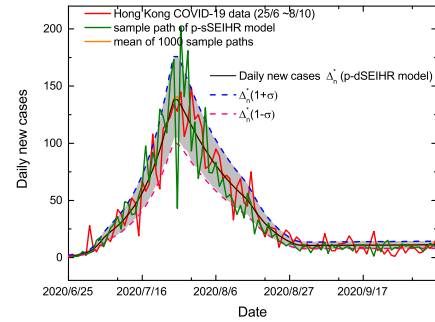

(a) Cumulative cases

(b) Daily new cases

Fig. 6 Actual and predicted numbers of COVID-19 cases in Hong Kong, as estimated using the p-sSEIHR model (5).

Δ_n / Δ_n^*. As shown in Figure 5b, the values r_n roughly follow a left-truncated Gaussian distribution with mean $\mu = 0.97$ and standard deviation $\sigma = 0.28$. Figure 5a shows Δ_n, Δ_n^*, and $\Delta_n^* (1 \pm \sigma)$ for the Hong Kong dataset.

Next, using the same fitted parameters as that in the p-dSEIHR model, the task is to find a value for φ such that $\rho_n = \Delta_n^{**} / \Delta_n^*$ has an standard deviation of σ, where Δ_n^{**} denotes the mean value of $\mathcal{S}(n) - \mathcal{S}(n+1)$ obtained from 1000 sample paths of the p-sSEIHR model. From the data, the result is found to be $\varphi = 0.03$.

Figure 6 shows the cumulative and daily numbers of COVID-19 cases in Hong Kong between 25 June and 8 Oct 2019, the p-dSEIHR model estimate, and the (2.5,97.5)-percentile range of sample paths from the corresponding p-sSEIHR model. As shown in Figure 6b, this range (colored gray) closely matches the $\Delta_n^* (1 \pm \sigma)$ envelope obtained from the p-dSEIHR model for all time points.

3.4.2 Other regions worldwide

In simulation, the process described in Section 3.4.1 is repeated for four global regions, namely Germany, Spain, South Africa, and the state of New York, using data obtained from [17]. The parameter settings of the fitted p-dSEIHR and p-sSEIHR models are given in Table 3. Again, the values of r_n roughly follow a left-truncated Gaussian distribution, with standard deviations σ ranging from 0.24 to 0.33.

Figure 7 shows the distributions of r_n for each of the four global regions, as well and the mean μ and standard deviation σ for each distribution.

Figure 8 shows the cumulative and daily numbers of COVID-19 cases in each global region during the specified time periods, the corresponding p-dSEIHR estimates, and the (2.5,97.5)-percentile ranges of sample paths from the corresponding p-sSEIHR model. It can be seen that these ranges, colored gray, closely match the $\Delta_n^* (1 \pm \sigma)$ envelope obtained from the p-dSEIHR model for the number of daily new cases (right column) in each region. This result not only demonstrates the usefulness of the p-sSEIHR model, but also reveals that the noise profiles in different regions may have a similar nature.
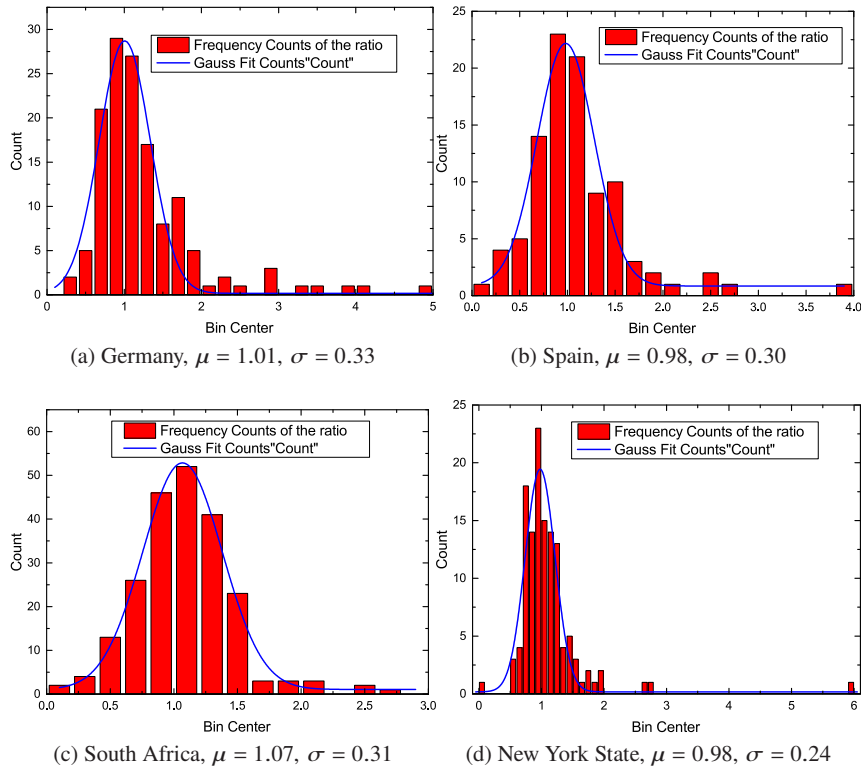
(a) Germany, $\mu = 1.01$, $\sigma = 0.33$

(b) Spain, $\mu = 0.98$, $\sigma = 0.30$

(c) South Africa, $\mu = 1.07$, $\sigma = 0.31$

(d) New York State, $\mu = 0.98$, $\sigma = 0.24$

Fig. 7 Distributions of r_n in four global regions.

Table 3 Parameter settings for four global regions, with $\beta = 0.14$, $\gamma = 0.7$, and $\omega_E = \omega_I = \omega_H = 0.1$ for all regions.

	Germany	Spain	South Africa	New York
N	83 783 945	47 431 256	59 622 350	19 745 289
φ	0.035	0.035	0.03	0.03
η	0.338	0.57	0.47	0.638
α	0.46	0.6	0.54	0.7
p_1	0.476	0.635	0.47	0.87
t_1	27/3 ~ 25/5	13/3 ~ 27/3	22/3 ~ 16/4	13/3 ~ 21/3
p_2	0.64	0.345	0.555	0.35
t_2	26/5 ~ 8/7	28/3 ~ 12/4	17/4 ~ 12/7	22/3 ~ 14/4
p_3	—	0.235	0.378	0.253
t_3	—	13/4 ~ 10/5	13/7 ~ 13/9	15/4 ~ 14/5
p_4	—	0.4445	0.47	0.31
t_4	—	11/5 ~ 31/5	14/9 ~ 12/10	15/5 ~ 6/7

3.5 Discussion

Section 3.2 shows the theoretical significance of the sSEIHR model. By studying the stability of the disease-free equilibrium, an estimated basic reproduction number R_0^* was obtained, which helped understand the critical condition of the COVID-19 outbreak. Section 3.3 confirms the numerical results in accordance with the theoretical results. The numerical results also show the critical value of the noise intensity where the pandemic happened. Most importantly, Section 3.4 shows the accuracy of our sSEIHR model on COVID-19 pandemic predictions in different irrelevant regions.

The key to the success achieved on COVID-19 pandemic forecasting/fitting can be appreciated by as follows. First, our deterministic SEIHR model captures the critical features of the COVID-19 transmission process, such as the disease transmission during the latent period, the hidden infections, and the isolation policy. Then, using only a few parameters, we successfully developed a deterministic model that accurately described the COVID-19 pandemic trend. In order to predict the daily caseload, the corresponding stochastic SEIHR model only needs one more parameter than the deterministic SEIHR model. Surprisingly, the real data analysis shows that the noise intensities in different regions are very close to each other. In the end, we used one parameter to describe the policy changes during the pandemic. Based on the timing of local intervention in each region, we changed the parameter accordingly and obtain an accurate prediction of the trend of the COVID-19 pandemic. All of this proves that our model has excellent performance in COVID-19 pandemic prediction.

4 Concluding remarks

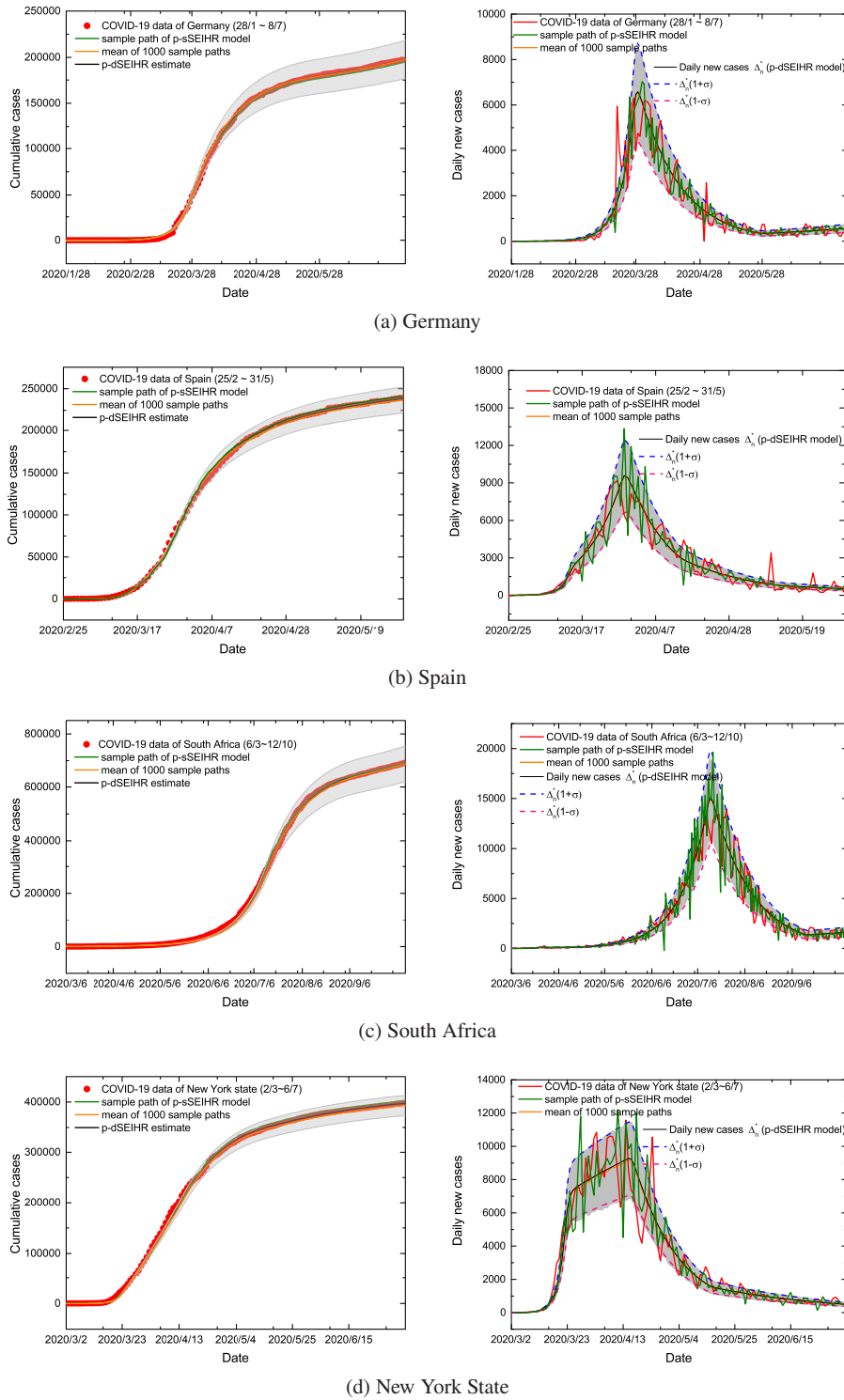In this paper, we developed a model described by a system of stochastic differential equations to describe a stochastic

(a) Germany

(b) Spain

(c) South Africa

(d) New York State

Fig. 8 Real COVID-19 data and p-sSEIHR predicted values for four global regions.

SEIHR (sSEIHR) process. We proved that the disease-free equilibrium is stochastically stable when the estimated basic reproduction number R_0^* is less than one. Extensive numerical results demonstrate strong threshold behavior near the estimated basic reproduction number, suggesting that the necessary conditions for stochastic stability are close to the sufficient conditions given in Theorem 1. We also included a factor $p(t)$, denoting the strength of various anti-epidemic interventions, to establish a p-sSEIHR model, and used this model to fit the daily number of hospitalized cases in Hong Kong and other four global regions worldwide. We find that the five regions have a very similar noise profile, with an estimated σ value from 0.24 to 0.33 and a fitted φ parameter from 0.03 to 0.035 for the p-sSEIHR model.

As the new p-sSEIHR model not only predicts the mean number of daily new cases of COVID-19 in a region, but also specifies a possible range of values, it can provide healthcare practitioners with a clearer picture of potential demands for medical resources due to epidemics. Furthermore, the similarity of the noise parameter φ for different regions suggests that the information about a disease learned from one region can be used for other regions regarding management policies, even at a less advanced stage of an epidemic, as long as the transmission parameters η and α in the model can be estimated with a reasonable accuracy.

A Concepts of stochastic stability

Without loss of generality, consider a stochastic system with a zero-vector equilibrium state, i.e. $\boldsymbol{x}_e = [0 \ldots 0]^{\mathrm{T}}$, which is referred to as the *trivial solution*.

Let $|\boldsymbol{x}|$ denote the Euclidean norm of a vector $\boldsymbol{x} = \boldsymbol{x}(t)$ for $t \geq 0$. Assume that there exists a unique solution to the stochastic system, i.e. $\boldsymbol{x}(t)$ exists and is unique, for any initial value $\boldsymbol{x}_0$ and Wiener process $W(t)$.

Recall some definitions and lemmas from the literature, e.g. [18, 19].

Definition 1 The trivial solution of a stochastic system is said to be *stochastically stable*, or stable in probability, if for all $\varepsilon > 0$ and $r > 0$, there exists a $\delta > 0$ such that $|\boldsymbol{x}(t)| > r$ for all $t > 0$ (i.e. $\boldsymbol{x}(t)$ remains within Euclidean distance r of the trivial solution) with probability $1 - \varepsilon$. Otherwise, the system is said to be *stochastically unstable*.

Definition 2 Let $\mathbb{R}_+ = \{x : x \geq 0\}$, and let $C^{2,1}(\mathbb{R}_+ \times \mathbb{R}_+; \mathbb{R})$ denote the set of functions $V(\boldsymbol{x}, t) \in (\mathbb{R}_+ \times \mathbb{R}_+) \to \mathbb{R}$ that are twice differentiable in $\boldsymbol{x} \in \mathbb{R}^d$ and once in $t \in \mathbb{R}_+$. Then, for the stochastic system (4) and some function $V \in C^{2,1}(\mathbb{R}_+ \times \mathbb{R}_+; \mathbb{R})$, define

$$LV(\boldsymbol{x}, t) = \frac{\partial V}{\partial t} + f^{\mathrm{T}}(\nabla_{\boldsymbol{x}} V) + \frac{1}{2}\mathrm{Tr}\left[g^{\mathrm{T}}(\mathbf{H}_{\boldsymbol{x}} V) g\right], \qquad (7)$$

where $\nabla_{\boldsymbol{x}} V$ and $\mathbf{H}_{\boldsymbol{x}} V$ are the gradient and Hessian matrix of V with respect to $\boldsymbol{x}$, respectively.

Lemma 1 *If there exists a positive-definite function $V(\boldsymbol{x}, t) \in C^{2,1}(\mathbb{R}_+ \times [t_0, \infty); \mathbb{R})$ such that $LV(\boldsymbol{x}, t) \leq 0$ for all $(\boldsymbol{x}, t) \in \mathbb{R}_+ \times [t_0, \infty)$, then the trivial solution to system (4) is stochastically stable.*

The method of proving the stability of a stochastic system using $V(\boldsymbol{x}, t)$ and $LV(\boldsymbol{x}, t)$ is known as Lyapunov's second method [18].

B Proof of Theorem 1

Since $\mathrm{d}S(t)$, $\mathrm{d}E(t)$, and $\mathrm{d}I(t)$ do not depend on $H(t)$ or $R(t)$, the compartments H and R can be decomposed from the system, yielding

$$\mathrm{d}\boldsymbol{x} = \mathrm{d}\begin{bmatrix} S \\ E \\ I \end{bmatrix} = \begin{bmatrix} -[\eta E + \alpha I]S \\ [\eta E + \alpha I]S - (\beta + \omega_E)E \\ \beta E - (\gamma + \omega_I)I \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ -\varphi I \end{bmatrix} \mathrm{d}W(t).$$

Substituting $\boldsymbol{u} = [u_1, u_2, u_3]^{\mathrm{T}} = [1 - S, E, I]^{\mathrm{T}}$, one obtains

$$\mathrm{d}\boldsymbol{u} = \begin{bmatrix} -[\eta u_2 + \alpha u_3](1 - u_1) \\ [\eta u_2 + \alpha u_3](1 - u_1) - (\beta + \omega_E)u_2 \\ \beta u_2 - (\gamma + \omega_I)u_3 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ -\varphi u_3 \end{bmatrix} \mathrm{d}W(t) \tag{8}$$

with disease-free equilibrium $\mathrm{DFE}_{\boldsymbol{u}} = (0, 0, 0)$. It suffices to show that $\mathrm{DFE}_{\boldsymbol{u}}$ is stochastically stable.

Linearizing (8) around the point $\mathrm{DFE}_{\boldsymbol{u}}$ (by removing the $u_1 u_2$ and $u_1 u_3$ terms), one obtains

$$\mathrm{d}\boldsymbol{u} = \begin{bmatrix} -(\eta u_2 + \alpha u_3) \\ (\eta u_2 + \alpha u_3) - (\beta + \omega_E)u_2 \\ \beta u_2 - (\gamma + \omega_I)u_3 \end{bmatrix} \mathrm{d}t + \begin{bmatrix} 0 \\ 0 \\ -\varphi u_3 \end{bmatrix} \mathrm{d}W(t). \tag{9}$$

It then suffices to prove the following:

Proposition 1 *The trivial solution to (9) is stochastically stable.*

Proof Lyapunov's second method [18] will be applied. Select a Lyapunov function of the form

$$V(\boldsymbol{u}) = 2u_1 + u_2 + Qu_3^2,$$

where $Q > 0$. Using the differential operator L defined in (7), one has

$$\begin{aligned}
LV(\boldsymbol{u}, t) &= \frac{\partial V}{\partial t} + f^{\mathrm{T}}(\nabla_{\boldsymbol{u}} V) + \frac{1}{2}\mathrm{Tr}\left[g^{\mathrm{T}}(\mathbf{H}_{\boldsymbol{u}} V) g\right] \\
&= 0 + \begin{bmatrix} -(\eta u_2 + \alpha u_3) \\ (\eta u_2 + \alpha u_3) - (\beta + \omega_E)u_2 \\ \beta u_2 - (\gamma + \omega_I)u_3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 2 \\ 1 \\ 2Qu_3 \end{bmatrix} \\
&\quad + \frac{1}{2}\begin{bmatrix} 0 & 0 & -\varphi u_3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2Q \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -\varphi u_3 \end{bmatrix} \\
&= -(\eta u_2 + \alpha u_3) - (\beta + \omega_E)u_2 \\
&\quad + 2Q\beta u_2 u_3 - 2Q(\gamma + \omega_I)u_3^2 + Q\varphi^2 u_3^2 \\
&\leq -(\eta u_2 + \alpha u_3)u_3 - (\beta + \omega_E)u_2 u_3 \\
&\quad + 2Q\beta u_2 u_3 - 2Q(\gamma + \omega_I)u_3^2 + Q\varphi^2 u_3^2 \\
&= [-\eta - \beta - \omega_E + 2Q\beta]u_2 u_3 \\
&\quad - \left[\alpha + 2Q\left(\gamma + \omega_I - \frac{\varphi^2}{2}\right)\right]u_3^2 \\
&\leq [\eta - \beta - \omega_E + 2Q\beta]u_2 u_3 \\
&\quad - \left[\alpha - 2Q\left(\gamma + \omega_I + \frac{\varphi^2}{2}\right)\right]u_3^2
\end{aligned}$$

Next, choose

$$Q = \frac{\alpha}{2\left(\gamma + \omega_I + \frac{\varphi^2}{2}\right)}$$

such that

$$LV(\boldsymbol{u}, t) \leq \left[\eta - (\beta + \omega_E) + \frac{\alpha\beta}{\left(\gamma + \omega_I + \frac{\varphi^2}{2}\right)}\right]u_2 u_3$$

$$= (\beta + \omega_E)\left(1 - R_0^*\right) u_2 u_3\,,$$

where

$$R_0^* = \frac{\eta}{(\beta + \omega_E)} + \frac{\alpha\beta}{(\beta + \omega_E)\left(\gamma + \omega_I + \frac{\varphi^2}{2}\right)}\,.$$

Thus, for $R_0^* < 1$, $V(\boldsymbol{u})$ is positive definite and $LV(\boldsymbol{u}, t) \leq 0$. By Lemma 1, the disease-free equilibrium $(0, 0, 0)$ of system (9) is stochastically stable.   □

Declarations

Conflicts of interest/Competing interests
None

Availability of data and material
Not applicable

Code availability
Not applicable

References

1. Elisabeth Mahase. Covid-19: What do we know about "long covid"? *BMJ*, 370:m2815, July 2020.
2. Seçkin Karasu, Aytaç Altan, Stelios Bekiros, and Wasim Ahmad. A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy*, 212:118750, 2020.
3. Aytaç Altan, Seçkin Karasu, and Enrico Zio. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Applied Soft Computing*, 100:106996, 2021.
4. Liang Tian, Xuefei Li, Fei Qi, Qian-Yuan Tang, Viola Tang, Jiang Liu, Zhiyuan Li, Xingye Cheng, Xuanxuan Li, Yingchen Shi, et al. Harnessing peak transmission around symptom onset for non-pharmaceutical intervention and containment of the COVID-19 pandemic. *Nature communications*, 12(1):1–12, 2021.
5. Choujun Zhan, Chi K Tse, Yuxia Fu, Zhikang Lai, and Haijun Zhang. Modeling and prediction of the 2019 coronavirus disease spreading in china incorporating human migration data. *Plos one*, 15(10):e0241171, 2020.
6. Choujun Zhan, Chi Kong Tse, Zhikang Lai, Xiaoyun Chen, and Mingshen Mo. General model for COVID-19 spreading with consideration of intercity migration, insufficient testing, and active intervention: Modeling study of pandemic progression in japan and the united states. *JMIR public health and surveillance*, 6(3):e18880, 2020.
7. Andrea L Bertozzi, Elisa Franco, George Mohler, Martin B Short, and Daniel Sledge. The challenges of modeling and forecasting the spread of COVID-19. *Proceedings of the National Academy of Sciences*, 117(29):16732–16738, 2020.
8. Ruiwu Niu, Eric W. M. Wong, Yin-Chi Chan, Michaël Antonie van Wyk, and Guanrong Chen. Modeling the COVID-19 pandemic using an SEIHR model with human migration. *IEEE Access*, 8:195503–195514, October 2020.
9. William Ogilvy Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 115(772):700–721, August 1927.
10. Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical Models in Epidemiology*. Springer New York, 2019.
11. Herbert W. Hethcote. Qualitative analyses of communicable disease models. *Mathematical Biosciences*, 28(3-4):335–356, January 1976.
12. K. Dietz. The incidence of infectious diseases under the influence of seasonal fluctuations. In *Mathematical Models in Medicine*, volume 11 of *Lecture Notes in Biomathematics*, pages 1–15. Springer Berlin Heidelberg, Mainz, March 1976.
13. Zhilan Feng and Horst R. Thieme. Recurrent outbreaks of childhood diseases revisited: The impact of isolation. *Mathematical Biosciences*, 128(1-2):93–130, July 1995.
14. B. Ivorra, M.R. Ferrández, M. Vela-Pérez, and A.M. Ramos. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, 88:105303, September 2020.
15. Elisabetta Tornatore, Stefania Maria Buccellato, and Pasquale Vetro. Stability of a stochastic SIR system. *Physica A*, 354:111–126, August 2005.
16. Álvaro Leitao and Carlos Vázquez. A stochastic θ-SEIHRD model: adding randomness to the COVID-19 spread, October 2020.
17. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Retrieved 23 Dec 2020.
18. J. Bertram and P. Sarachik. Stability of circuits with randomly time-varying parameters. *IRE Transactions on Circuit Theory*, 6(5):260–270, May 1959.
19. Xuerong Mao. *Stochastic differential equations and applications*. Woodhead Publishing, Cambridge, 2 edition, 2008.